



HAL
open science

Discours d'entreprise et organisation de l'information - Apports de la textométrie dans la construction de référentiels terminologiques adaptables au contexte

Frédéric Erlos

► **To cite this version:**

Frédéric Erlos. Discours d'entreprise et organisation de l'information - Apports de la textométrie dans la construction de référentiels terminologiques adaptables au contexte. Linguistique. Université de la Sorbonne nouvelle - Paris III, 2009. Français. NNT: . tel-00511829

HAL Id: tel-00511829

<https://theses.hal.science/tel-00511829>

Submitted on 30 Aug 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE LA SORBONNE NOUVELLE – PARIS 3

École doctorale 268 – Langage et langues

DISCOURS D'ENTREPRISE ET ORGANISATION DE
L'INFORMATION

APPORTS DE LA TEXTOMÉTRIE DANS LA CONSTRUCTION DE
RÉFÉRENTIELS TERMINOLOGIQUES ADAPTABLES AU CONTEXTE

Tome 1

Thèse de doctorat de Sciences du langage

présentée et soutenue publiquement

par

Frédéric ERLOS

Sous la direction de :
André SALEM, Professeur

Jury :

Loïc DEPECKER
Brigitte GUYOT
Monique SLODZIAN

26 novembre 2008

UNIVERSITÉ DE LA SORBONNE NOUVELLE – PARIS 3

École doctorale 268 – Langage et langues

DISCOURS D'ENTREPRISE ET ORGANISATION DE
L'INFORMATION

APPORTS DE LA TEXTOMÉTRIE DANS LA CONSTRUCTION DE
RÉFÉRENTIELS TERMINOLOGIQUES ADAPTABLES AU CONTEXTE

Tome 1

Thèse de doctorat de Sciences du langage

présentée et soutenue publiquement

par

Frédéric ERLOS

Sous la direction de :
André SALEM, Professeur

Jury :

Loïc DEPECKER
Brigitte GUYOT
Monique SLODZIAN

26 novembre 2008

Remerciements

Qu'il me soit permis d'exprimer ma reconnaissance envers mon directeur de recherche, André Salem, dont les conseils et les chaleureux encouragements m'ont permis de réaliser ce travail. Je tiens aussi à remercier très vivement Loïc Depecker, qui m'a initié aux travaux terminologiques avec générosité et une grande hauteur de vues. Ma gratitude va également à Muriel Amar, Danièle Dégez, Brigitte Guyot, Armelle Le Bars, Bernard Bosredon et Dardo de Vecchi, dont les travaux et les conseils amicaux ont tant contribué à guider ma réflexion. Je remercie également Étienne Brunet, Serge Fleury, Gaston Gross et Maria Zimina-Poirot pour leurs encouragements. Que les membres du département Organisation – qualité – intranet de Crédit Agricole s.a. sachent combien je leur sais gré de leurs encouragements amicaux. Enfin, je tiens à exprimer toute ma reconnaissance à Valérie Assan qui m'a apporté son soutien tout au long de la préparation de cette thèse et sans l'aide de qui ce travail n'aurait pas vu le jour.

SOMMAIRE

Introduction générale	11
PREMIÈRE PARTIE RÉFÉRENTIELS ET DISCOURS D'ENTREPRISE	35
Introduction de la première partie.....	37
Chapitre 1 Contextes : terminologies, intranets et référentiels.....	43
I. Échanges langagiers et problématiques informationnelles dans les entreprises	47
II. Cahier des charges pour un référentiel terminologique adaptable au contexte	92
Chapitre 2 Un référentiel à l'épreuve des discours.....	137
I Corpus et référentiel	140
II. Rapport annuel et rapport d'activité.....	149
III. Le dispositif textuel des rapports d'activité	162
IV. Constitution du corpus.....	186
Chapitre 3 Première approche diachronique.....	201
I Le contrôle des contextes ou la tectonique des dispositifs textuels.....	204
II Série textuelle chronologique et évolution du vocabulaire : faire du neuf avec du vieux ?.....	230
III Accroissements spécifiques et noms propres.....	253
Conclusion de la première partie	261
DEUXIÈME PARTIE « PARLER D'ENTREPRISE » ET COMMUNICATION INSTITUTIONNELLE.....	267
Introduction de la deuxième partie	269
Chapitre 4 Langue commune et banque grand public	275
I. La langue commune, une question de norme	277
II. Langue commune et vocabulaire de la banque	286
Chapitre 5 Un usage « générique » de la terminologie bancaire et financière	301
I. Implantation de la terminologie du domaine.....	304
II. Banalisation du vocabulaire utilisé pour la restitution de l'activité bancaire	314

Chapitre 6	Activité bancaire et phraséologie des rapports d'activité	327
I.	Vocabulaire commun et phraséologie des rapports d'activité.....	329
II.	Deuxième approche diachronique : fluctuation de l'activité bancaire et variétés de langue	347
Chapitre 7	Un vocabulaire peut-il être transparent ?	367
I.	Filtrages thématiques et variations de fréquences dans les rapports d'activité.....	370
II.	Variations de fréquence et variétés de langue.....	377
III.	Les rapports d'activité constituent-ils un genre de discours second ?	388
Conclusion de la deuxième partie		413
TROISIÈME PARTIE DÉNOMINATIONS PROPRES ET RÉFÉRENTIEL TERMINOLOGIQUE		429
Introduction de la troisième partie		431
Chapitre 8	Noms propres en discours	439
I.	Recensement des noms propres du corpus	444
II	Méthode de collecte des dénominations propres et résultats obtenus.....	467
III.	Dénominations propres et composants de l'activité bancaire.....	494
Chapitre 9	L'ancrage d'un référentiel : les toponymes.....	521
I.	Les toponymes dans le corpus de rapports d'activité.....	525
II.	Odonymes : de la fonction de « coordonnées » à l'effet « international ».....	529
III.	Noms de villes : d'une géographie de l'actualité secondaire aux places financières internationales	538
IV.	Pays : la « dimension internationale » et la structuration géographique de l'activité	545
Chapitre 10	contenu des dénominations propres et vocabulaire du corpus.....	587
I.	Le fonctionnement discursif des dénominations propres	591
II.	Formes du contenu d'une dénomination propre.....	601
III.	Identification et caractéristiques des structures de contenu.....	613
Chapitre 11.	Un référentiel en mouvement.....	659
I.	Organisation de la collecte	662
II.	Réalisation de la collecte.....	682
III.	Évaluation de la collecte	715
Conclusion générale.....		765

Corpus, lexiques et dictionnaires utilisés.....	791
bibliographie.....	795
LISTE DES ANNEXES.....	839
INDEX DES NOMS PROPRES.....	843
TABLE DES TABLEAUX.....	849
TABLE DES FIGURES.....	855
TABLE DES MATIÈRES.....	861

INTRODUCTION GÉNÉRALE

« Le têt d'une moule ramassé dans une rue de Paris n'a pas la même importance quand il s'agit de rechercher l'origine et l'histoire de son espèce que celui que l'on a recueilli dans un des terrains, ou secondaires ou tertiaires, qui recouvrent notre globe. Il en est de même pour les mots. Rien de plus imprudent que de spéculer sur un mot isolé : nous voulons dire placé dans l'isolement artificiel du lexique et comme arraché à son milieu naturel. »

J. Gilliéron et J. Mongin, « Scier » dans *la Gaule romane du Sud et de l'Est* (1905).

Depuis son ouverture au grand public à partir du milieu des années 1990, l'Internet n'a cessé de se développer, au point que l'on éprouve aujourd'hui des difficultés à évaluer précisément le volume représenté par les milliards de pages pouvant être consultées sur ce réseau. Les résultats surabondants des moteurs de recherche, l'organisation absconse des pages de certains sites font partie des embûches rencontrées par les internautes au cours de leur navigation. À côté de ce phénomène d'ampleur mondiale se sont développés les intranets, que l'on qualifiera dans un premier temps de réseaux Internet utilisant les technologies du Web (navigateur, liens hypertextes, pages HTML) et mis en place à l'échelle des organisations pour leur usage exclusif. Ces réseaux privés présentent les mêmes défauts qu'Internet, certes à une échelle moindre, mais avec des conséquences parfois plus sensibles.

En effet, les intranets sont passés en une dizaine d'années du stade de médium complémentaire du papier, fréquemment utilisé par les directions de la communication, à celui de ressource électronique incontournable pour la réalisation des activités les plus diverses à l'intérieur des organisations. Cette variété des utilisations s'applique à la diffusion d'informations, comme celles qui sont relatives, par exemple, aux ressources humaines, à la sécurité ou à l'informatique, mais aussi à la constitution de bibliothèques numériques de procédures et de réglementations encadrant la réalisation des activités, ou encore à l'utilisation de programmes de

gestion destinés aussi bien à la saisie en ligne des congés qu'à la production de tableaux de bord destinés au pilotage d'une organisation. Dans les contextes professionnels, on observe que les exigences à l'égard de ce médium et des informations qu'il permet de diffuser se sont renforcées, que ce soit en termes de périodicité de mise à jour, de facilité d'accès ou encore de fiabilité.

On constate ainsi, à l'intérieur d'une même organisation, que l'un des principaux obstacles à la réalisation des échanges entre agents à l'aide de ce médium réside dans la confrontation de manières de dire et de classer qui se trouvent différentes. En effet, alors qu'une conversation de vive voix permet à un néophyte de demander le sens d'une expression qu'il ignore, la communication médiatisée par les écrits numériques mis en ligne sur les sites d'un intranet empêche de tels ajustements. Cet obstacle est redoublé par le fait que la mise en ligne de contenus professionnels a vocation à toucher des populations nouvelles, dont les caractéristiques sont mal connues des émetteurs, parce qu'elles sortent du cadre traditionnel de circulation des écrits au travail.

On peut tenter de contribuer à la résolution de ces difficultés en organisant les sites et les regroupements de contenus qu'ils proposent à partir des façons de parler et de classer propres à leurs différents publics. La mise en place d'un référentiel terminologique adaptable au contexte doit alors rendre compte des variations linguistiques affectant l'usage d'un sociolecte. Cette présentation, sous une forme ordonnée, de la diversité des usages linguistiques à l'intérieur d'une même organisation, est donc destinée à documenter certaines activités d'information apparues récemment, et qui sont liées à la diffusion étendue d'écrits professionnels à l'aide de sites intranets.

L'étude des variations affectant l'usage d'une langue est abordée par différentes disciplines dans le domaine linguistique. La variation diachronique est étudiée par la grammaire historique et la lexicologie ; la variation dans l'espace constitue l'objet de la géographie linguistique et de la dialectologie ; la variation rapportée aux groupes sociaux est le domaine de la sociolinguistique et, plus récemment, la statistique textuelle et la linguistique de corpus étudient les variations affectant le vocabulaire de textes réunis en corpus électroniques en fonction, par

exemple, des énonciateurs, des genres de discours ou de la diachronie. Toutefois, l'articulation entre variation linguistique et organisation de l'information est une problématique qui a surtout été discutée dans le domaine des sciences de l'information. En effet, l'analyse du contenu des documents destinée à produire des classements ou des regroupements thématiques constitue l'un des objets majeurs de cette discipline. Traditionnellement, la réalisation d'une telle description repose sur l'utilisation de langages documentaires.

Ces langages peuvent posséder une structure hiérarchique, comme les classifications, telle que la Classification décimale universelle (CDU) utilisée dans les bibliothèques publiques, ou bien être organisés en fonction d'une logique combinatoire, comme les thésaurus qui rassemblent des termes – ou descripteurs – servant à décrire le contenu des documents. Dans tous les cas, les langages documentaires sont des langages contrôlés et normalisés. Alors que les classifications ont recours à un codage alphanumérique, les thésaurus n'admettent de la langue commune ou spécialisée que les substantifs masculins au singulier. Par ailleurs, une partie des relations unissant les descripteurs ont pour fonction de conjurer différentes formes de variations possibles assimilées à des défauts du langage naturel : existence de synonymes, de mots polysémiques, d'homonymes, mais aussi d'expressions jugées indésirables au profit de termes dits préférentiels. Dans tous les cas, l'utilisation des langages documentaires vise à réduire les effets jugés négatifs des différents lieux de la variation linguistique présents dans un processus d'organisation de l'information : vocabulaires des auteurs de documents, des personnes qui mènent une recherche d'informations, et des bibliothécaires ou documentalistes qui, dans leur conversation ordinaire, n'utilisent pas le langage documentaire dont ils se servent pour décrire, classer et rechercher les documents.

Or, comme on le sait, une part substantielle du travail de ces professionnels de l'information consiste à faire accoucher les utilisateurs de leur besoin d'information et à traduire ce dernier dans le langage ayant servi à décrire les documents susceptibles de contenir une réponse à la question posée. Dans ce mode classique d'organisation de l'information, des intermédiaires humains doivent prendre en charge, entre autres choses, les phénomènes de variation linguistique. En effet, ils traduisent la langue avec laquelle les demandes sont formulées dans la

terminologie particulière utilisée pour décrire le contenu des documents. On devra se demander si cette forme d'organisation est adaptée au contexte de l'intranet. Or, la question se pose en des termes différents dès lors que des systèmes permettent d'interroger directement les documents ou leurs descriptions.

Dans de tels systèmes, qu'il s'agisse d'interroger une base de données structurées à l'aide d'un langage documentaire ou de lancer une recherche plein texte portant sur un ensemble de textes au format électronique, la problématique centrale est celle de l'appariement. Celle-ci consiste en la correspondance exacte entre les mots que l'utilisateur emploie pour formuler sa requête et ceux qui figurent dans l'index d'un moteur de recherche. Cette approche « système », en se concentrant sur l'appariement des chaînes de caractères de la requête et de l'index, permet d'éviter la question des usages linguistiques et des phénomènes de variation propres aux contextes réels d'utilisation. C'est d'ailleurs sur ce point que portent, depuis une vingtaine d'années au moins, les remises en causes et les tentatives d'aménagement¹.

Celles-ci visent à réintroduire l'utilisateur dans la conception et l'évaluation des performances de tels systèmes. En effet, l'expression du besoin d'information et l'appréciation de la valeur d'un résultat obtenu lors d'un processus de recherche d'informations relèvent, en définitive, de l'utilisateur du système. Lorsqu'elles prennent ce constat pour point de départ, certaines modélisations proposent de tenir compte des caractéristiques individuelles des usagers, du cadre où se situent les interactions entre usagers et système, et plus généralement, de tous les aspects contextuels, individuels, cognitifs, systèmes et documentaires relatifs aux situations de recherche

¹ La littérature sur le sujet étant extrêmement volumineuse, on se contente de renvoyer ici à quelques travaux donnés à titre d'exemples. Pour l'aménagement de la conception des systèmes de recherche d'information, on renvoie aux synthèses des travaux de N. Belkin, P. Ingwersen, M. Bates et T. Saracevic, *in* S. Chaudiron et M. Ihadjadene (2002, 2004). Pour l'introduction des traitements automatiques du langage naturel dans de tels systèmes et la prise en compte du vocabulaire des fonds documentaires, on peut citer, entre autres, les travaux de K. Sparck Jones (1971, 1999), D.C. Blair (1990, 2006), T. Strzalkowski (1999), S. Bertrand-Gastaldy et G. Pagola (1991, 1992), M. K. Buckland (1999), M. Amar (2000) et S. Dalbin (2007). Dans le cadre des systèmes traditionnels d'organisation de l'information, la prise en compte des usages linguistiques des usagers est ancienne, même si elle n'est pas dominante. On peut la faire remonter au moins aux remarques de C.A. Cutter (1876 et 1904). Pour la prise en compte de la terminologie du domaine couvert par un fonds documentaire, on renvoie à I. Dahlberg (1992), B. Hjørland (1997, 2002), M. Holzem (1999) et W. Mustafa-Elhadi (1992, 2006).

d'information. Cependant, on constate que, malgré leur grande richesse, ces modèles ne permettent pas d'appréhender les problématiques fines des usages langagiers que l'on envisage d'étudier dans ce travail. En effet, il semble qu'une conception tacite de la langue condamne ces approches à n'envisager les usages qu'à travers les lexiques de la langue commune ou des langues spécialisées. Cette conception, qui est du reste largement partagée, ne permet pas de rendre compte des sélections lexicales opérées par les locuteurs en fonction des situations de communication réelles. La mise en place de référentiels terminologiques adaptables au contexte nécessite donc d'opter pour une approche différente.

Dans la mesure où les caractéristiques réelles des situations de communication jouent un rôle de premier plan dans la méthodologie qui est proposée, il a été nécessaire d'adopter un cadre concret pour servir d'illustration. On a retenu à cette fin le cas d'un groupe bancaire français (le Crédit agricole) et, à l'intérieur de cet ensemble, le réseau intranet de sa maison mère (Crédit agricole s.a.). Ce réseau se compose d'une centaine de sites intranets diffusant des informations à usage local, mais surtout à destination de nombreux publics répartis dans les 350 entités du groupe bancaire. Les référentiels terminologiques adaptables au contexte sont alors destinés à documenter le travail de création, d'organisation, de publication et de recherche de contenus diffusés sur les sites intranets de l'organe central de cette banque.

À l'échelle d'une telle organisation, l'observation et la restitution des usages langagiers nécessitent la mise en place de plusieurs référentiels terminologiques. Chaque site intranet étant susceptible d'être visité par des populations différentes d'intranetes, un même référentiel terminologique a normalement vocation à restituer plusieurs usages linguistiques. Dans la mesure où ce travail vise, outre la présentation des étapes nécessaires à la réalisation d'un tel référentiel, l'exposition de questions méthodologiques, on a choisi de simplifier certains points d'ordre terminographique. C'est pourquoi on a fait correspondre la nomenclature du référentiel terminologique présenté avec la restitution du vocabulaire utilisé dans une situation de communication type. Lorsque l'on adopte pour cadre de référence une situation de communication réelle, les caractéristiques de cette dernière peuvent

faire l'objet d'une description. La mise en ligne d'informations sur un site intranet rassemble une série d'activités qui entrent dans le cadre d'une situation de communication, dans laquelle un destinataire, généralement collectif, s'adresse à un public destinataire, lui-même composé de populations aux caractéristiques différentes. Les productions textuelles utilisées comme ressources pour la construction des référentiels terminologiques sont elles-mêmes issues d'un processus communicationnel incorporant destinataire et destinataires dans un contexte donné. En résumé, la construction d'un référentiel terminologique consiste à documenter une situation de communication en cours à l'aide des traces laissées par un échange ayant eu lieu dans une situation de communication révolue.

Il n'en reste pas moins que la documentation des activités d'information à l'aide des productions issues d'une autre situation de communication nécessite que l'on puisse établir un rapport d'homologie entre les deux sortes d'échanges qui sont en jeu. Ce faisant, il convient également de déterminer les limites d'un tel rapport, chaque situation de communication imprimant ses caractères propres aux productions verbales auxquelles elle sert de cadre. Dès lors qu'elle procède par valeurs approchées, la pertinence de cette tentative d'apprivoisement des usages linguistiques ne peut être que relative. Malgré cette limitation, le caractère pragmatique d'une telle démarche présente l'avantage de reposer sur une explicitation des principales propriétés des échanges. En effet, on peut alors présenter les finalités poursuivies, les interlocuteurs, le propos, les lieux et moments légitimes ainsi que les supports matériels utilisés.

On a retenu pour ce travail une situation assez générale dans laquelle le destinataire s'adresse à un public de néophytes. La réussite de la communication dépend alors dans une large mesure des efforts déployés par le destinataire afin de se faire comprendre de son public. On a indiqué précédemment qu'une telle prise en charge énonciative fait généralement défaut sur les intranets. Dans leur visée, les discours diffusés sur les intranets ne prennent pas en compte la nécessité de combler l'obstacle constitué par un écart cognitif important entre interlocuteurs. C'est pourquoi, l'une des manières de documenter une telle situation consiste d'abord à identifier les types de discours produits au sein des organisations et destinés à évoquer des réalités peu ou mal connues des destinataires. Ce que l'on cherche alors

à capter, c'est la manière dont certains éléments appartenant à un référentiel d'objets sont construits dans et par les discours pour ce type de public.

Sur un intranet, les sites de métiers, c'est-à-dire les sites intranets dédiés à des activités précises dans les organisations, comme les affaires juridiques, la comptabilité ou le marketing, possèdent des publics variés. Ceux-ci se répartissent sur un *continuum* s'échelonnant entre un public d'experts et des intranutes néophytes. En effet, les utilisateurs peuvent être de nouveaux embauchés qui ont tout à découvrir, mais aussi à des agents exerçant leur métier de façon cloisonnée alors qu'ils contribuent à la réalisation d'un même processus (établissement d'un dossier, élaboration d'un nouveau produit). Il peut également s'agir de salariés exerçant la même profession sur des sites distants ou à l'intérieur d'entités différentes appartenant au même groupe, ces distances géographiques suffisant généralement à introduire des différences dans les usages linguistiques. La prise en compte d'un public de néophytes permet donc d'embrasser des cas de figure assez divers. Cependant, l'usage d'un intranet étant réservé à une population caractérisée par l'appartenance de ses membres à un même organisme, l'éventail des thématiques à prendre en compte peut être resserré.

Dans un tel cadre, émetteurs et récepteurs sont censés partager ou souhaiter partager un référentiel commun, nécessaire à la réalisation de leurs activités. Ce référentiel d'entreprise est alors défini comme l'ensemble des êtres et des objets dont un locuteur peut parler dans une langue donnée, dès lors que ces référents entretiennent un rapport direct ou indirect avec la réalisation d'activités professionnelles dans le cadre d'une organisation. Ce référentiel est constitué de différents cercles relatifs au métier, au secteur d'activité ainsi qu'aux orientations stratégiques de l'organisation ou encore aux relations que celle-ci entretient avec son environnement. Par ailleurs, il est susceptible d'être verbalisé de différentes manières selon le caractère plus ou moins formel des échanges et selon le niveau de connaissance des locuteurs.

La verbalisation d'un tel référentiel constitue la part principale du sociolecte que l'on cherche à caractériser. En effet, si le premier niveau de variation linguistique se situe à l'échelle individuelle avec l'idiolecte propre à chaque agent, les échanges qui ont lieu quotidiennement dans une grande entreprise montrent que

l'utilisation personnelle d'une langue incorpore nombre d'éléments dont la connaissance est partagée par tout un groupe. D'un autre côté, l'existence des langues spécialisées montre que la langue commune, idiome partagé par une communauté très étendue de locuteurs, ne suffit pas à la réalisation des échanges, dès lors que ceux-ci sont insérés dans des formes spécialisées de l'activité humaine. On admet donc l'existence d'un sociolecte propre aux agents réalisant leurs activités au sein d'un même organisme. Un tel sociolecte, qui devra être caractérisé plus finement, incorpore non seulement les éléments de la langue commune nécessaires aux échanges, mais aussi les unités terminologiques en rapport avec les activités exercées, ainsi que les expressions propres au groupe de locuteurs étudié.

Par ailleurs, la source utilisée pour la construction d'un référentiel terminologique adaptable au contexte doit être contemporaine des activités de partage d'informations qu'il s'agit de documenter : organisation des pages d'un site et recherche d'information. La mise en place d'un tel référentiel nécessite donc une mise à jour « au fil de l'eau », de manière à capter, d'une part, les modifications du référentiel (apparition et disparition d'objets) et, d'autre part, les variations susceptibles d'affecter les usages linguistiques dans la durée. Enfin, compte tenu des formes d'activité à documenter, il est souhaitable que le support de la ressource utilisée existe au minimum sous une forme électronique diffusable sur un intranet. L'importance accordée à la situation de communication dans la construction des référentiels terminologiques adaptables au contexte confère à ce type d'outils un statut particulier qu'il convient de préciser.

En effet, il apparaît tout d'abord que ces outils sont plus proches des terminologies structurées que des thésaurus ou des classifications. Ils visent dans un premier temps à restituer les éléments caractéristiques d'un sociolecte, et ce n'est que dans un deuxième temps qu'ils peuvent être utilisés pour structurer les pages d'un site ou caractériser les thèmes présents dans les contenus publiés. Ils sont destinés à documenter des activités d'information, non à se substituer aux classifications ou aux langages documentaires existants. À ce titre, ils fournissent des unités candidates, pouvant enrichir d'autres référentiels, tels qu'une terminologie, un dictionnaire utilisé pour le traitement du langage naturel, ou encore,

un langage documentaire. Dans le contexte de l'intranet, ils constituent une ressource autonome qui vient renforcer les différents procédés destinés à adapter l'organisation d'un site à ses différents publics.

Dans la mesure où ces référentiels terminologiques obéissent à une finalité spécifique, il est également nécessaire que l'on compare leurs principales caractéristiques avec celles des référentiels terminologiques plus classiques destinés, par exemple, à la traduction ou à la normalisation des usages linguistiques à l'intérieur d'une discipline, d'un domaine ou d'un secteur d'activité. Ce type d'outils devant s'adapter à un contexte précis et permettre la documentation de certaines activités d'information, leur mise en place a nécessité que l'on s'écarte des conceptions terminologiques traditionnelles, telles qu'elles ont été exposées dans les travaux d'E. Wüster¹. En cela, la démarche que l'on a retenue s'inscrit dans le prolongement des développements récents de la terminologie textuelle².

Tout d'abord, on notera que de tels référentiels sont attachés à des contextes opérationnels délimités de façon très précise, puisqu'ils prennent place à l'intérieur d'organisations particulières, telles qu'une institution, une entreprise ou une association. Dans ce cadre, ils visent en outre la documentation d'usages linguistiques caractéristiques de situations de communication particulières. Plutôt que la prescription d'un usage uniforme, la constitution de tels référentiels s'inscrit dans un programme de description des usages linguistiques à l'intérieur d'une entreprise. Cependant, on se concentrera dans ce travail sur les différents aspects à prendre en compte pour la mise en place d'un référentiel destiné à documenter une situation de communication assez répandue, sans aborder les problèmes spécifiques

¹ Ingénieur électricien de formation, E. Wüster (1898-1977) s'est intéressé aux travaux sur la langue technique dans l'Allemagne des années 1930. Sa thèse (*Internationale Sprachnormung in der Technik*, 1931), a porté sur la normalisation internationale de la langue dans les domaines techniques. « Développé et publié en 1931, ce livre suscita la création du comité 37 « Terminologie » au sein de l'ISA et l'élaboration des principes terminologiques. » in H. Felber (1987 : 24). Ses travaux sont à l'origine de l'École de Vienne (*Vienna Theory of Terminology*), et ont eu une influence considérable dans le domaine de la terminologie.

² La terminologie textuelle propose de partir des textes et non plus des concepts comme dans la terminologie wüsterienne, afin de fonder les travaux terminologiques. Elle pose que le terme n'est pas une simple étiquette associée à un concept, mais résulte d'une construction. Celle-ci repose sur une analyse du fonctionnement discursif des termes et doit être évaluée par rapport à son utilité dans l'application visée. Une telle terminologie se veut plus descriptive que normative, et s'appuie sur des analyses réalisées à partir de corpus de textes. Sur ce sujet, on renvoie aux travaux de R. Kocourek (1991) et de M. Slodzian (1999, 2000).

attachés à une cartographie des différents usages. De même, on n'abordera pas les problématiques du multilinguisme, même si les options retenues ménagent des possibilités d'extension dans cette direction.

Les référentiels terminologiques adaptables au contexte visant avant tout la restitution des usages langagiers caractéristiques d'une situation de communication, leur élaboration repose sur des ressources textuelles plutôt que sur un ensemble préexistant de notions. Dans la mesure où l'on pose comme premières les traces discursives laissées par certains échanges, la nomenclature de ces outils a vocation à intégrer des unités non spécifiquement terminologiques, dès lors qu'elles sont utilisées afin d'évoquer des activités spécialisées. Ces référentiels sont donc ouverts à certains mots pouvant relever a priori de la langue commune, ainsi qu'aux noms propres, ces dernières unités jouant un rôle important dans le programme de cartographie mentionné précédemment. Les ressources textuelles retenues étant relativement homogènes en raison de leur appartenance à des situations d'énonciation similaires, la construction d'un référentiel à partir d'une telle source ne peut prétendre représenter la terminologie en usage dans une organisation. Cette limitation assumée, qui peut être levée dès lors que le nombre adéquat de corpus a été exploité, autorise à n'intégrer dans ces référentiels que des unités caractéristiques et, par conséquent, à ne pas tendre vers l'exhaustivité, dont les contours sont souvent arbitraires. De là, un référentiel de cette sorte, pris isolément, ne peut pas non plus offrir une vision systématique de la terminologie collectée. En revanche, il doit restituer une sélection particulière opérée dans l'usage d'un sociolecte, à partir de laquelle les agents réalisant certaines activités d'information pourront extrapoler.

Afin de restituer finement les usages langagiers propres à telle situation de communication, on propose que la construction des référentiels terminologiques s'appuie sur des corpus électroniques organisés sous la forme de séries textuelles chronologiques. Ces séries doivent par ailleurs être homogènes (du point de vue de l'émetteur et du genre de discours) et ouvertes (il s'agit de corpus de suivi). Cette organisation est destinée, d'une part, à faciliter le contrôle des caractéristiques de la situation d'énonciation à l'origine des discours rassemblés en corpus, et d'autre part, à suivre l'évolution des usages linguistiques. En effet, si la collection de textes est hétérogène, il devient difficile de faire précisément le départ entre le récurrent et

l'accidentel. Le suivi des usages linguistiques impose une mise à jour permanente, qui dépendra néanmoins de la périodicité de parution des ressources textuelles utilisées. Enfin, les textes réunis doivent avoir pour propriété d'être produits dans le contexte de l'entreprise prise en exemple et pour la réalisation de ses activités.

D'autre part, relations entre les unités présentes dans les référentiels terminologiques adaptables au contexte sont motivées par les relations sémantiques établies dans et par les textes utilisés comme sources. Lorsque les traces syntaxiques sont elliptiques, l'établissement des relations est explicité à l'aide d'une interprétation des éléments fournis par les contextes d'apparition des unités. L'établissement de telles relations, ainsi que le jugement de lexicalisation porté sur un syntagme, reposent sur les attestations présentes dans les discours avec un certain niveau de récurrence, plutôt que sur un avis « expert ». Les critères de récurrence peuvent néanmoins varier en fonction de la nature des relations ainsi que des types d'unités concernées. Les unités collectées pour ces référentiels sont liées entre elles à l'aide des principales relations formelles utilisées pour la construction de terminologies, c'est-à-dire les relations hiérarchiques et associatives ainsi que la relation de synonymie.

Compte tenu, d'une part, du fait que ces référentiels ont avant tout une fonction descriptive et qu'ils sont orientés vers des problématiques d'accès à l'information, et d'autre part, du fait qu'ils sont destinés à suivre les évolutions affectant les usages d'un sociolecte, les données qu'ils incorporent doivent être limitées à ce qui est nécessaire et suffisant pour ces utilisations. Elles peuvent ainsi se réduire à la nomenclature et aux principales relations formelles établies entre unités, sans que celles-ci soient accompagnées de leur définition. Par ailleurs, les informations relatives aux sources utilisées et à la datation peuvent être portées collectivement par le référentiel, dans la mesure où sa construction repose sur une source homogène qui fera elle-même l'objet d'une documentation précise. Le stockage et l'utilisation des données des référentiels doivent pouvoir s'intégrer facilement aux autres outils utilisés couramment pour la réalisation des activités d'information. Ainsi, la restitution des données doit correspondre à deux sortes d'utilisations principales. Un export formaté doit pouvoir être facilement obtenu, afin de réutiliser les données dans des dictionnaires spéciaux, tels que ceux qui sont

utilisés pour réaliser des expansions de requêtes sur les moteurs de recherche. D'autre part, le contenu des référentiels terminologiques doit être accessible depuis un intranet, soit sous la forme d'une navigation arborescente, soit par une recherche réalisée à l'aide de mots-clés.

La présentation de ce cahier des charges soulève certaines questions qu'il convient d'aborder maintenant. En effet, l'établissement de la nomenclature d'un référentiel terminologique adaptable au contexte doit résoudre trois difficultés principales. Tout d'abord, une telle construction suppose que certains objets appartenant au référentiel étudié soient suffisamment partagés par les locuteurs pour qu'ils puissent être évoqués dans de nombreuses situations de communication. De façon plus précise, et compte tenu de la finalité des référentiels terminologiques, l'évocation de ces objets doit être partagée entre les discours utilisés comme sources du référentiel terminologique et la situation de communication que l'on cherche à documenter. En second lieu, le repérage de ces objets suppose que les mots employés pour les désigner varient peu d'un discours à l'autre. Il s'agit par conséquent d'identifier des unités relativement imperméables aux variations linguistiques induites par la diversité des situations de communication. Enfin, ces points de repères doivent être caractéristiques du vocabulaire étudié ou permettre de capter la présence d'autres unités possédant cette propriété.

Il semble alors que les noms propres, à condition de ne pas les réduire à une classe spécifique de substantifs et d'accepter qu'ils puissent dénommer autre chose que des personnes et des lieux, possèdent des propriétés intéressantes sous le rapport des questions qui viennent d'être évoquées. En effet, leur pragmatique les signale comme une sorte d'unités particulières permettant d'isoler des entités uniques au sein d'ensembles d'objets déjà catégorisés. Ainsi, parmi l'ensemble indéfini que constitue un référentiel, les noms propres permettent d'isoler la mention de certains objets, à la manière des cartouches qui distinguent les noms de rois et de reines dans l'écriture hiéroglyphique. Cette mise en exergue, manifestée dans les écrits par la présence d'une majuscule, répond en outre à un besoin d'ordre communicationnel. En effet, on a attribué un nom propre à ces objets parce qu'ils sont sujets à de fréquentes prédications et qu'une autre expression référentielle plus efficace n'a pas

pu être trouvée. Enfin, les échanges linguistiques relatifs à des objets individuels indiquent l'existence d'un groupe de locuteurs plus ou moins étendu. L'usage d'un nom propre supposant ordinairement que les interlocuteurs connaissent le référent qu'il sert à désigner, un groupe de locuteurs peut être caractérisé par le fait de partager une telle connaissance.

Cependant, ces propriétés pragmatiques ne garantissent pas que les mêmes noms propres appartiennent à la fois au vocabulaire des ressources textuelles destinées à confectionner les référentiels terminologiques et à celui des pages et des contenus qui doivent être organisés. Néanmoins, on a vu que la présence de ces unités dans les discours est destinée à évoquer des réalités partagées par les interlocuteurs, ou bien des objets propres au référentiel d'une organisation précise. On peut donc faire l'hypothèse que ces unités circulent d'une situation de communication à l'autre dans le cadre de la réalisation ou de l'évocation des activités au sein d'une même organisation.

Par ailleurs, la pragmatique des noms propres régule de diverses manières leur forme linguistique : leur morphologie les rattache à un idiome ; ils héritent des formes de nomination qu'une culture donnée a adoptées relativement à certaines classes d'objets ; enfin, ils garantissent un lien stable entre un référent et une dénomination indépendamment de la situation d'énonciation. Il s'agira d'apprécier la portée de cette stabilité dans la mesure où l'on sait également que cette forme de dénomination connaît des variantes orthographiques. Il sera également nécessaire d'apprécier l'imperméabilité de ces dénominations aux phénomènes de variation linguistique. Sur ce dernier point, on peut faire l'hypothèse que la sélection de séries textuelles homogènes est de nature à limiter l'ampleur des variations à observer, tout en autorisant certains écarts autour d'un noyau commun. Dans tous les cas, la présence des noms propres dans le vocabulaire d'un corpus devra faire l'objet d'une double évaluation. En effet, il est probable que leur présence puisse varier selon les textes et les genres de discours utilisés pour la confection de différents référentiels terminologiques. D'autre part, la présence de ces unités n'indique pas quelles sortes de relations sont établies entre elles et les autres composantes d'un vocabulaire.

Alors qu'il semble possible d'identifier dans les discours certains éléments caractéristiques d'un référentiel à l'aide des noms propres qui les dénomment de

manière stable, il faut néanmoins convenir qu'un sociolecte ne saurait être réduit à cette classe d'unités. On propose donc d'utiliser ces repères comme « unités-pilotes », afin de capter d'autres éléments présents dans le vocabulaire des textes du corpus utilisé comme source pour la construction d'un référentiel terminologique. Par conséquent, lorsque la présence des noms propres dans un corpus est non seulement avérée mais aussi significative, il est nécessaire de savoir si ces unités sont reliées au reste du vocabulaire. Cette vérification s'impose, car les noms propres peuvent être utilisés en dehors du texte principal en étant isolé à l'intérieur de dispositifs textuels particuliers, tels que des listes d'objets divers, des organigrammes ou des tableaux résumant les caractéristiques de certaines entités.

Si l'ensemble des noms propres, ou seulement certaines classes, sont associés au reste du vocabulaire, il faut ensuite essayer de connaître la nature et la récurrence des relations sémantiques entretenues avec les autres unités présentes dans les textes. Là encore, il s'agit de vérifier que la réalisation de cette étape ne se heurte pas à de nouveaux écueils. Le premier pourrait surgir de la diversité de ces relations et rendre vaine toute tentative de caractérisation. Le second proviendrait, au contraire, d'une trop grande pauvreté des relations, qui hypothèquerait la productivité de la collecte et les chances de capter d'autres éléments caractéristiques du vocabulaire du corpus. Il faudra donc parvenir à préciser le fonctionnement discursif des classes de dénominations propres présentes dans le corpus avant d'engager la collecte du vocabulaire à l'aide d'« unités pilotes ».

En outre, l'examen du fonctionnement des noms propres dans les discours suppose que l'on adopte une position claire en ce qui concerne les propriétés sémantiques de ce type d'unités. Sur cette question, les points de vue d'un nom propre vide de sens¹ et d'un nom propre figurant parmi les noms « les plus significatifs² » ont été défendus dès la deuxième moitié du XIX^{ème} siècle. Entre ces

¹ « Des observations qui précèdent, il est facile de conclure que lorsque les noms fournissent quelque information sur les objets, c'est-à-dire, lorsqu'ils ont proprement une signification, cette signification n'est pas dans ce qu'ils dénotent, mais dans ce qu'ils connotent. Les seuls noms qui ne connotent rien sont les noms propres ; et ceux-ci n'ont, à strictement parler, aucune signification. » John Stuart Mill, *Système de logique déductive et inductive*, Tome 1, (1889 : 35).

² « La différence avec les noms communs est une différence tout intellectuelle. Si l'on classait les noms d'après la quantité d'idées qu'ils éveillent, les noms propres devraient être en tête, car ils sont les plus significatifs de tous, étant les plus individuels. » M. Bréal, *Essai de sémantique* (1897 : 198).

deux positions extrêmes s'échelonnent diverses conceptions, dans lesquelles se négocient de façons variées les réponses à cette interrogation : comment une unité linguistique, soit vide de sens, soit chargée d'un trop plein de significations, parvient-elle à désigner un être ou un objet uniques, indépendamment des changements qu'ils subissent et des situations où ils sont évoqués ?

Depuis les années 1970, cette question a fait l'objet de nombreux débats dans le domaine de la linguistique. Ces travaux ont mis au jour, entre autres, l'existence d'un *continuum* entre noms communs et noms propres, dont l'antonomase constitue la figure emblématique. Plus récemment encore, c'est l'étude du fonctionnement discursif des noms propres qui a fait l'objet d'investigations nouvelles. Bien que ce travail s'inscrive dans le prolongement de ce dernier courant, il s'écarte des tentatives qui auraient pour conséquence de réduire les propriétés du nom propre à celles du nom commun. On pense, en effet, qu'il y a bien une spécificité du nom propre, dont témoignent, entre autres, les caractéristiques pragmatiques présentées précédemment. Il sera donc nécessaire, à propos de la sémantique des noms propres, d'adopter une position qui tienne compte de cette spécificité, tout en permettant d'aborder le fonctionnement discursif de ces unités. Pour cela, on a été amené à reprendre la conception d'un sens instructionnel – reposant sur le lien dénomiatif –, couplé à un contenu discursif, c'est-à-dire élaboré dans et par les discours.

C'est sur une telle assise que l'on sera conduit à formuler une double hypothèse. Tout d'abord, les noms propres utilisés dans une série textuelle homogène permettent de capter les éléments caractéristiques d'un vocabulaire. En second lieu, ces éléments font partie des ingrédients dont est constitué le contenu des dénominations propres. Cependant, on verra que la construction d'un tel objet discursif suppose que les noms propres entretiennent avec les autres composantes du vocabulaire des relations dotées d'une certaine organisation, c'est-à-dire dont il est possible de dégager les principaux traits récurrents. Si l'on peut mettre au jour ce que l'on propose de nommer la structure-type du contenu d'une classe de dénominations propres, l'identification d'une telle construction discursive devrait permettre d'affiner la connaissance du fonctionnement des noms propres dans les textes d'un corpus. En rendant possible la sélection des classes de dénominations

propres présentant le plus grand intérêt en fonction des buts poursuivis, de telles observations pourront être utilisées afin d'orienter la collecte avec précision.

Ainsi, les classes de dénominations propres utilisées pour la collecte seront celles dont les référents font l'objet d'opérations de catégorisation ou sont regroupés au sein d'ensembles définis empiriquement. En effet, ces deux sortes de relations possèdent un caractère exemplaire dès lors qu'il s'agit de guider des activités d'information, car elles restituent la manière dont certains objets sont catégorisés et classés lorsqu'ils sont évoqués dans le cadre d'une situation de communication particulière. Ces catégorisations et classements pourront alors être considérés comme des manières de dire et de classer adaptées à un certain public d'intranautes. C'est également à partir de cette « structure discursive » que l'on cherchera à savoir si les variations affectant l'association de certaines unités aux noms propres permettent de saisir des phénomènes de variation diachronique.

Les solutions qui seront proposées pour répondre aux différentes questions soulevées par la mise en place et la réalisation d'une telle collecte devront ensuite être évaluées à travers les résultats qui auront été obtenus. La nomenclature des noms propres du corpus peut être évaluée sous le rapport de l'exhaustivité ou de l'enregistrement de variantes utiles pour la recherche d'informations. Mais un tel résultat doit aussi être apprécié sous l'angle de la sélection d'éléments caractéristiques du vocabulaire du corpus. Plusieurs plans doivent alors être distingués. Comme on l'a vu précédemment, l'usage des noms propres sélectionne des individus au sein de classes d'objets. Ce faisant, leur usage opère aussi un choix parmi les classes embrassant les différents éléments dont se compose un référentiel d'objets. On peut alors se demander si ces choix correspondent à ceux que l'on observe pour les différentes catégories d'objets plus particulièrement évoqués dans les textes du corpus. Il s'agit alors de mesurer une première limite possible de la collecte. On peut également chercher à savoir si les unités associées aux dénominations propres utilisées comme unités-pilotes appartiennent aux mêmes strates du sociolecte étudié que celles qui caractérisent le vocabulaire du corpus pris dans son ensemble. Enfin, la collecte peut être confrontée à un modèle théorique des activités d'une organisation, de manière à évaluer, par contraste, la richesse des formes de regroupements d'objets qui ont été mises au jour.

Comme les précédents développements l'auront déjà suggéré, les textes occupent une place essentielle dans l'approche que l'on propose pour la réalisation de référentiels terminologiques. Parce qu'ils sont les traces laissées par des échanges langagiers situés, ils constituent la matière première de référentiels adaptables au contexte. De ce point de vue, les référentiels terminologiques ont pour fonction de rendre explicites les circulations intertextuelles qui échappent aux locuteurs pris dans la réalisation de leurs activités professionnelles. Ce n'est alors plus l'unité terminologique qui est première, mais les discours qui la véhiculent d'une situation de communication à l'autre. L'observation des variations qui affectent l'usage d'un sociolecte rend par conséquent nécessaire la mise en place d'un dispositif spécifique, capable de faciliter l'exploitation de ressources textuelles dans un cadre terminographique.

Tout d'abord, on propose que le choix des textes rassemblés dans le corpus source d'un référentiel terminologique s'appuie sur les caractéristiques d'une situation de communication précise, plutôt que sur le seul critère de l'appartenance à la littérature représentative d'un domaine. La constitution de ressources homogènes rend possible le contrôle des évolutions d'un vocabulaire en limitant le nombre des caractéristiques de la situation d'énonciation à prendre en compte. Encore faut-il que l'étude des relations établies entre ces caractéristiques et les mouvements d'un vocabulaire puissent être documentés en intégrant les contraintes de délais (réalisation et mise à jour) et de fiabilité (reproductibilité), pesant généralement sur la réalisation des projets terminographiques.

C'est pourquoi, il a été jugé nécessaire de faciliter ce travail d'exploitation de ressources textuelles en l'outillant à l'aide de programmes informatiques adaptés. Depuis une cinquantaine d'années, la statistique lexicale, la linguistique quantitative et la statistique textuelle, ou textométrie, ont cherché à outiller les opérations essentielles représentées par la segmentation des textes en mots et l'exploitation des

données ainsi constituées. On dispose aujourd'hui de programmes très aboutis, tels que les logiciels SATO, Hyperbase, Lexico, Weblex ou encore Xaira¹.

Dans la mesure où la prise en compte des phénomènes de variation linguistique intègre une dimension diachronique, les corpus utilisés sont organisés sous la forme de séries textuelles chronologiques. Ces ensembles de textes homogènes comportent couramment des centaines de milliers d'occurrences. Les logiciels de textométrie doivent alors permettre de comparer des ensembles de vocabulaire importants, mais aussi de livrer les contextes précis dans lesquels les occurrences d'une même forme apparaissent. Les comparaisons du premier type portent sur différents ensembles de textes. Il s'agit, par exemple, de repérer la présence d'usages linguistiques différents en comparant entre eux des stocks distributionnels du vocabulaire². On utilise pour cela aussi bien des tests statistiques (pour la comparaison de fréquences attestées) que des évaluations qualitatives (étude d'implantation d'une terminologie, repérage de marques métalinguistiques et linguistiques attestant une prise en charge énonciative). On peut également réaliser des observations sur les variations diachroniques affectant l'usage d'un sociolecte à partir des textes d'une même série.

Par ailleurs, il s'avère également nécessaire d'opérer des découpages à un niveau infra-textuel, soit à des niveaux correspondant au découpage original des textes, tels que les chapitres ou les rubriques délimités par des intertitres, soit à des niveaux plus standards, comme le paragraphe ou la phrase, ou encore, à des paliers infra-phrastiques, comme le segment de phrase ou le syntagme. On doit ainsi pouvoir disposer des principales sortes de contextes linguistiques dans lesquels apparaissent les mots présents dans un corpus.

¹ Ces programmes font l'objet d'une présentation et peuvent être utilisés en ligne ou téléchargés aux adresses suivantes :

<http://www.ling.uqam.ca/ato/sato/>

<http://ancilla.unice.fr/~brunet/pub/hyperbase.html>

<http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW>

<http://weblex.ens-lsh.fr/wlx/>

<http://www.xaira.org/>

Dans ce travail on a plus particulièrement utilisé le logiciel Lexico 3.

² Un stock distributionnel du vocabulaire d'un fragment de texte correspond au « *vocabulaire de ce fragment assorti de comptages de fréquence pour chacune des formes entrant dans sa composition.* » in L. Lebart et A. Salem (1994 : 318).

Lorsqu'elle est retenue, une telle approche propose également l'utilisation du dictionnaire des formes graphiques d'un corpus de textes comme point de départ, afin d'identifier et de recenser certaines catégories d'unités. Bien que des programmes d'extraction automatique d'unités terminologiques ou d'entités nommées soient disponibles pour ce type de traitement, on a préféré opter pour une solution permettant de traiter entièrement la chaîne de collecte que l'on propose. Par conséquent, on présente une méthode dans laquelle le repérage, la délimitation et l'extraction des dénominations propres sont assistés par des techniques textométriques. Ces dernières correspondent, selon les besoins, au filtrage des formes graphiques en fonction de l'usage de la majuscule, au tri par ordre de fréquence, mais aussi aux concordances regroupées en fonction des partitions du corpus, ou encore, à la recherche de chaînes incluses dans les segments répétés présents dans les textes. L'établissement de la nomenclature des dénominations propres du corpus permet ensuite la répartition de ces unités en grandes classes. C'est à l'aide de représentants de chaque classe que l'on cherche ensuite à construire les structures types de contenu.

On a donc pris le parti d'organiser ce travail en fonction des principales étapes de la méthode que l'on propose de mettre en œuvre pour la construction de référentiels terminologiques adaptables au contexte. Cela explique que les différents points théoriques qu'il a été nécessaire d'aborder se répartissent sur les trois parties traitant chacune d'une étape fondamentale : la constitution d'un corpus source, la caractérisation du vocabulaire étudié du point de vue des variétés de langue employées dans les textes du corpus et, enfin, la collecte des dénominations propres, puis celle des unités caractéristiques du vocabulaire du corpus qu'il aura été possible de capter par leur intermédiaire.

Dans la première partie, on présente tout d'abord le cadre dans lequel est apparu le besoin de référentiels terminologiques adaptables au contexte. L'intranet de Crédit agricole s.a. permet d'illustrer la manière dont les activités liées à la diffusion d'informations professionnelles à l'aide de ce médium doivent être documentées du point de vue des usages linguistiques. Un cahier des charges destiné à la mise en place des référentiels terminologiques adaptables au contexte est alors

proposé (chapitre 1). La première étape de la réalisation de ce type d'outil consiste à définir les critères retenus pour le choix du corpus de textes qui sera utilisé comme source principale. Dans la mesure où l'on a souhaité illustrer la méthode en cherchant à documenter une situation de communication assez générique dans le contexte des intranets – il s'agit du cas où un public néophyte prend connaissance d'informations relatives à un secteur d'activité du groupe bancaire –, il a été nécessaire d'identifier parmi les discours produits par l'organisation, ceux qui s'adressent à un public de ce type. On a ainsi constitué un corpus de rapports d'activité couvrant une période allant de 1995 à 2003. Les spécificités du genre et ses dispositifs textuels caractéristiques sont présentés, ainsi que la constitution d'un corpus textuel homogène au format électronique (chapitre 2). On a également considéré que la dizaine d'années couvertes par le corpus était susceptible d'être le théâtre de changements pouvant affecter tant la situation d'énonciation que les éléments du référentiel véhiculé par ces textes. C'est pourquoi on a exposé les moyens de contrôler les évolutions des dispositifs textuels et du vocabulaire du corpus. On met ainsi en évidence l'utilité de tels contrôles dès lors que l'on fait le choix d'un corpus ouvert (chapitre 3).

La seconde partie est consacrée à la caractérisation du vocabulaire du corpus. En effet, cette étape est nécessaire si l'on veut être en mesure d'évaluer le caractère représentatif de la collecte que l'on projette de réaliser à l'aide des noms propres. C'est pourquoi on a cherché à disposer d'une vue d'ensemble concernant les propriétés du vocabulaire de référence. Pour cela, on a eu recours à un certain nombre de tests textométriques destinés non seulement à identifier la présence de telle ou telle variété de langue, mais surtout à préciser la manière dont chacune est utilisée dans les textes du corpus. L'un des procédés auxquels on a eu recours consiste à repérer la façon dont les principaux composants de l'activité bancaire sont évoqués.

Ainsi, on d'abord cherché à caractériser l'usage qui est fait de la langue commune dans les rapports d'activité (chapitre 4), avant de mettre au jour les propriétés des unités terminologiques présentes dans ces textes (chapitre 5). Par ailleurs, les rapports d'activité paraissant contenir également une phraséologie propre au groupe bancaire, on a donc étudié un certain nombre d'unités

représentatives de cet usage (chapitre 6). Enfin, on a cherché à déterminer dans quelle mesure ce genre de discours, destiné à renseigner un public de néophytes sur une réalité particulière, relève d'une forme de vulgarisation. En effet, on peut penser que la finalité des rapports d'activité a contraint leur émetteur à utiliser certains procédés de prise en charge énonciative destinés à combler les lacunes présumées du lectorat. Ces procédés, qui peuvent être ramenés à une explicitation des termes utiles à l'exposition du propos mais qui restent mal connus du public, constituent une source qu'il est utile de repérer, lorsque l'on cherche à identifier des utilisations concomitantes de variétés de langue différentes. Ce faisant, on a également cherché à préciser l'usage qui est fait des noms propres dans les différentes strates du vocabulaire des rapports d'activité (chapitre 7).

Enfin, la troisième partie est consacrée à la réalisation de la collecte à proprement parler. Dans la mesure où celle-ci débute avec la collecte des noms propres, il s'est avéré nécessaire de définir l'extension d'une telle notion. Par ailleurs, on a opté pour une collecte extensive en ajoutant aux noms propres certains mots dotés d'une majuscule initiale en dehors des débuts de phrase. Ce parti pris, autorisant la prise en compte de certains usages « abusifs » de la majuscule, mais néanmoins dignes d'intérêt d'un point de vue cognitif, a nécessité que l'on précise les techniques textométriques qui semblent les plus adéquates pour la collecte. L'identification des dénominations et de leurs variantes fait ensuite l'objet d'évaluations quantitatives et qualitatives (chapitre 8). Cette première étape, en mettant en lumière l'existence de différentes classes de dénominations propres inégalement représentées dans le corpus et dotées d'une hiérarchisation spécifique des fréquences, suggère que l'étude du fonctionnement discursif des dénominations propres prenne en compte le rôle que leur fait jouer l'énonciateur. On étudiera ainsi la manière dont les noms de lieux sont utilisés dans les rapports d'activité. On s'interrogera ensuite sur ce que cet usage révèle de la géographie réelle et imaginaire véhiculée par ces discours (chapitre 9). Cependant, le fonctionnement discursif des noms propres ne se limite pas à cette sorte d'usage, surtout lorsque les réalités évoquées ne sont pas censées appartenir au référentiel partagé par les interlocuteurs. Dans le but de préparer la deuxième étape de la collecte, on propose une méthode permettant d'évaluer la capacité d'une classe de dénominations propres à mettre en

évidence des unités nouvelles caractéristiques du vocabulaire du corpus. Les structures-types de contenu, en offrant une vision synthétique des sortes de relations tissées par les dénominations propres avec d'autres unités, permettent de choisir les « unités-pilotes » et les sortes de relations qui ont des chances d'être les plus productives (chapitre 10). La deuxième étape de la collecte est alors présentée à partir d'un échantillon d'unités-pilotes appartenant aux classes des noms de produits et de personnes morales. Les fréquences associées aux unités étant très différentes dans les deux ensembles, on présente deux approches textométriques distinctes. Cette deuxième collecte fait à son tour l'objet d'une évaluation à la fois quantitative et qualitative. Les modalités d'intégration des données collectées dans un référentiel terminologique sont abordées, ainsi que le stockage et la restitution des données. Les observations réunies lors de cette étape permettent enfin de mettre en évidence certains mécanismes relatifs au traitement discursif des informations associées aux noms propres dans les rapports d'activité, ainsi que les évolutions affectant l'évocation d'un référentiel en continuel renouvellement (chapitre 11).

PREMIÈRE PARTIE

RÉFÉRENTIELS ET DISCOURS D'ENTREPRISE

INTRODUCTION DE LA PREMIÈRE PARTIE

Les échanges langagiers se développent à l'intérieur des organisations en empruntant des formes nouvelles, telle que la diffusion d'informations à l'aide de sites intranets¹, apparue au milieu des années 1990. Cette diffusion prend place dans un ensemble de pratiques que l'on peut qualifier d'activités d'information. Celles-ci visent, dans ce cadre, à produire, classer, diffuser des informations liées à la réalisation des activités d'une organisation, et destinées à un public s'étendant au-delà du cadre habituel de circulation des écrits au travail. Ces activités comprennent également les tâches liées à la recherche des informations diffusées sur un tel médium. Ainsi, les destinateurs, habitués à diffuser leurs écrits auprès de publics restreints et identifiés, se trouvent-ils généralement démunis, dès lors qu'il leur incombe de toucher un public nouveau, aux caractéristiques mal identifiées, et partant, difficiles à prendre en compte. De même, les destinataires, que la réalisation de leurs tâches rend dépendants d'informations produites et mises à disposition par d'autres, sont-ils confrontés à des navigations hypertextuelles qui les obligent à entrer dans des manières de dire et de classer qui ne leur sont pas toujours très familières.

L'accomplissement de ces tâches par chaque agent soulève donc des problèmes nouveaux, qui ne font pas l'objet d'une attention suffisante dans le cadre de l'organisation du travail. En effet, en dehors des échanges se déroulant au sein de groupes de taille réduite, les activités d'information sont centrées sur la gestion du stock d'informations propre à chaque agent, ou bien prises en charge par des agents spécialisés, comme les documentalistes, pour les informations provenant de l'extérieur, les archivistes, pour les écrits produits par l'organisation et qui sont devenus obsolètes, ou encore, les chargés de communication, qui diffusent les informations de l'entreprise vers des publics variés aussi bien internes qu'externes. Les échanges d'informations écrites liées à la réalisation des activités et effectués par sites intranets interposés se sont développés depuis une dizaine d'années, sans

¹ Ensemble de pages HTML (*hypertext markup language*) et de liens hypertextes consultables sur un réseau privé (intranet) par l'intermédiaire d'un navigateur et donnant accès à des informations utilisées dans le cadre de la réalisation des activités d'une organisation. Cette définition provisoire sera discutée dans le premier chapitre.

que dans les organisations on ait pris toute la mesure de cette nouvelle forme de communication écrite opérationnelle à diffusion élargie.

Les difficultés qui n'ont pas manqué d'apparaître lors de l'accomplissement de ces échanges, et dont les plus répandues concernent l'utilisation de moteurs de recherche donnant des résultats décevants et peu exploitables, ou la navigation dans des arborescences de sites inextricables, ont néanmoins trouvé des pistes de solutions. La plupart sont apparues concomitamment sur Internet¹, dès lors que les émetteurs d'information ont été soucieux d'améliorer la diffusion de leurs contenus à destination publicitaire, commerciale ou institutionnelle. Le moteur de recherche² permet de rechercher simultanément sur plusieurs sites d'un intranet ; le référencement³ permet, entre autres, de répertorier en un seul endroit les sites disponibles ; les tests d'utilisabilité⁴, hérités des contrôles d'ergonomie appliqués aux interfaces homme-machine, ont été étendus aux attentes des visiteurs d'un site en termes de contenu informationnel ; enfin, la cybermétrie⁵ livre à l'analyste des statistiques reposant sur l'enregistrement des parcours empruntés et des questions posées par les visiteurs d'un site. Toutes ces approches positionnent l'utilisateur au centre de la solution devant conduire à l'amélioration des échanges. Cependant, toutes ne sollicitent pas de la même manière. La participation des internautes s'inscrit sur un *continuum*, allant de la responsabilité presque complète du bon

¹ « Réseau informatique mondial constitué d'un ensemble de réseaux nationaux, régionaux et privés, qui sont reliés par le protocole de communication TCP-IP » in *Le Grand dictionnaire terminologique de Office québécois de la langue française* (<http://www.granddictionnaire.com>).

² « Programme qui indexe le contenu de différentes ressources Internet, plus particulièrement de sites Web, et qui permet, à l'aide d'un navigateur Web, de rechercher de l'information selon différents paramètres, en se servant de mots-clés, ou par des requêtes en texte libre, et d'avoir accès à l'information ainsi trouvée. (...) Google, Alta Vista et Excite sont des exemples de moteurs de recherche. » *Ibid.*

³ « Enregistrement d'un site Web dans les moteurs et répertoires de recherche, afin de le faire connaître aux internautes, d'en accroître la visibilité et d'en augmenter ainsi le nombre de visiteurs. » *Ibid.*

⁴ « Qualité d'un matériel ou d'un logiciel qui est facile et agréable à utiliser et à comprendre, même par quelqu'un qui a peu de connaissances en informatique. (...) Bien que souvent utilisé avec le sens de « convivialité », le terme utilisabilité a un sens plus large qui correspond à la « capacité d'un système à permettre à ses utilisateurs de faire efficacement ce pourquoi ils l'utilisent ». Afin que le travail soit fait, le système « utilisable » doit non seulement être facile à utiliser, mais aussi fiable et efficace. » *Ibidem.* Le terme anglo-saxon est « usability ».

⁵ La cybermétrie, également nommée « web analytics » consiste en l'étude du comportement des visiteurs d'un site, réalisée à partir de données statistiques collectées au cours des navigations hypertextuelles, et dans le but d'améliorer celles-ci.

accomplissement de l'échange, avec les moteurs de recherche (à indexation de moteur égale, il faut savoir utiliser ce dernier au mieux, identifier les bonnes ressources d'information, dépouiller les résultats, etc.¹), jusqu'à l'enregistrement, à leur insu, de leur navigation sur les pages d'un site (cybermétrie), en passant par des sollicitations ponctuelles, au cours de tests d'utilisabilité. Bien que toutes ces approches soient susceptibles, à des degrés divers, d'apporter des éléments de solution, elles paraissent plus ou moins bien adaptées aux exigences liées aux sites intranets.

En effet, contrairement à Internet, l'intranet – que l'on qualifiera provisoirement de Web² privé réservé à une organisation –, en raison de la nature et de l'utilisation des informations qu'il sert à diffuser, jouit d'une situation paradoxale. Intégré au déroulement des activités d'une organisation, il ne bénéficie pas toujours de la visibilité nécessaire pour lui garantir les moyens adaptés à la bonne réalisation des tâches de diffusion d'information. Si les informations indexées par les moteurs de recherche s'y trouvent en quantité bien moindre que sur le Web, les exigences relatives à la qualité des résultats y sont bien plus grandes. Elles portent sur la facilité d'accès, la fraîcheur de l'information, sa fiabilité ainsi que sur la garantie de son caractère opérationnel. Par ailleurs, l'organisation des sites intranets doit fournir un accès balisé et permanent à des informations ayant les mêmes qualités que celles qui viennent d'être citées pour les résultats de recherche, car elles sont destinées à des publics professionnels. Or, les tests d'utilisabilité et la cybermétrie ne permettent que des ajustements ponctuels ou a posteriori. Les premiers sont, en effet, fréquemment liés à la phase de construction des sites, mais ils sont plus difficiles à établir après leur ouverture. Quant à la cybermétrie, elle produit des comptages a posteriori dont l'interprétation n'est pas toujours aisée et qui, de ce fait, permettent difficilement de transformer les observations ainsi obtenues en enseignements susceptibles d'être utilisés afin d'adapter l'organisation des sites. De ce point de vue, les sites d'un intranet semblent occuper une position

¹ C'est la raison d'être d'ouvrages tels que le *Guide de recherche sur Internet – Outils et méthodes* de B. Foenix-Riou (2002).

² « Système basé sur l'utilisation de l'hypertexte, qui permet la recherche d'information dans Internet, l'accès à cette information et sa visualisation. » *Grand dictionnaire terminologique* de l'Office québécois de la langue française.

intermédiaire entre les sites Internet, dont ils héritent les technologies mais aussi les traitements de l'information peu formalisés, voire inexistantes, et les systèmes classiques d'organisation de l'information, mis en place de longue date, et qui disposent de moyens spécifiques destinés à garantir une diffusion et un accès effectifs à l'information dans des contextes professionnels.

On cherchera donc à identifier les raisons pour lesquelles les intranets n'ont généralement pas hérité des caractéristiques des systèmes d'organisation de l'information, dans la mesure où ils partagent des objectifs similaires. Ce faisant, on mettra en lumière les principales caractéristiques de ces moyens utilisés pour organiser et diffuser l'information dans les organisations. La dimension langagière des échanges paraissant y occuper une position centrale, et les réponses apportées par les systèmes d'organisation de l'information ne semblant pas adaptées, on proposera une approche reposant sur la mise en place de ressources linguistiques dédiées, les référentiels terminologiques adaptables au contexte. Ces ressources auront pour fonction principale de guider les activités d'information, en proposant à l'agent qui organise et publie des informations sur un site, une image linguistique de la manière dont chacun de ses publics verbalise, catégorise et classe certaines sortes d'objets relatifs à l'exercice de son activité.

La mise en place de tels outils, qui permettront également de constituer des dictionnaires destinés à améliorer la qualité de la recherche d'informations, nécessite que l'on adopte une conception des échanges langagiers, médiatisés par les sites intranets, qui opère une rupture avec les vues de sens commun prévalant en ce domaine. Dans ce but, on recherchera un modèle de la communication capable d'embrasser les différentes caractéristiques des situations dans lesquelles ont lieu des échanges verbaux. Ce préalable est nécessaire, dans la mesure où les ressources terminologiques proposées visent à produire des échantillons situés de l'usage d'un sociolecte. Leur construction nécessitera que l'on puisse également caractériser un tel sociolecte, de manière à permettre l'identification et l'exploitation de discours produits par l'organisation et représentatifs de l'usage linguistique attesté dans des situations de communication aux propriétés comparables. Les contraintes, mais aussi les possibilités liées au traitement de situations de communication types au sein d'une organisation précise, permettent d'aborder la constitution de telles ressources

terminologiques, moins en termes d'exhaustivité que d'exemplarité. Il sera alors nécessaire d'adopter une sorte de fil conducteur capable de permettre une exploitation homogène, mais aussi des comparaisons entre les résultats obtenus à partir de corpus aux propriétés différentes. On fera alors l'hypothèse qu'un tel fil conducteur peut résider dans une certaine classe d'unités linguistiques, présentes dans différentes sortes de discours et susceptibles d'établir un lien entre l'étude des variations affectant les usages linguistiques et les objets propres au référentiel manipulé dans les situations d'organisation de l'information. Enfin, dans la mesure où la mise en place de tels référentiels terminologiques ne prend sens que dans des contextes situés, on a adopté comme exemple d'organisation pour cette étude le groupe bancaire Crédit agricole au début des années 2000.

Après avoir présenté le contexte des échanges langagiers médiatisés par les intranets dans les organisations, on s'attachera à définir le cadre théorique nécessaire au développement des référentiels terminologiques adaptables au contexte. On envisagera également les problématiques liées à la réalisation d'un projet terminographique particulier (chapitre 1). On présentera ensuite le type de ressource textuelle retenue pour la constitution d'un corpus, dont les propriétés seront discutées au regard de l'activité d'information que l'on aura choisi de documenter. Il s'agira alors de présenter un genre de discours, l'échantillon retenu pour illustrer la manière dont un référentiel peut être verbalisé, ainsi que les étapes de la constitution d'un corpus de textes au format électronique (chapitre 2). Enfin, le suivi des usages linguistiques nécessitant l'adoption de corpus textuels ouverts, on proposera un moyen de tester l'évolution du vocabulaire de ces derniers, ainsi que les changements ayant pu affecter leurs conditions de production. Cela permettra d'identifier les moyens de contrôler l'évolution des usages que l'exploitation d'un échantillon de discours d'entreprise permet de capter (chapitre 3).

CHAPITRE 1

CONTEXTES : TERMINOLOGIES, INTRANETS ET RÉFÉRENTIELS

Le pluriel associé au titre de ce premier chapitre suggère que l'on va présenter plusieurs sortes de contextes. En réalité, il vise plutôt à introduire l'articulation que l'on cherchera à développer entre deux contextes de nature différente et que la linguistique oppose habituellement. Le contexte du premier type correspond aux « *éléments de la réalité non linguistique associés à la production d'un énoncé ou d'un élément d'énoncé*¹. » Il s'agit donc de la réalité servant de cadre aux échanges linguistiques. Ces derniers se matérialisent sous la forme de discours, servant à leur tour de contexte aux unités plus petites dont ils sont composés : « *Ensemble du texte qui entoure un élément de la langue (mot, phrase, fragment d'énoncé) et dont dépend son sens, sa valeur*². » Les référentiels terminologiques adaptables au contexte ont vocation à tenir une position intermédiaire entre ces deux réalités. D'une part, leur utilisation n'a de sens que dans un contexte d'usage particulier où ils devront participer à la réalisation d'activités d'information, elles-mêmes intégrées à des degrés variables à l'organisation du travail dans une entreprise. D'autre part, en tant que formalisation et restitution des usages langagiers observés dans un tel cadre, les référentiels terminologiques devront être composés d'unités et de relations sémantiques attestées dans les discours.

Les activités d'information pour lesquelles les référentiels terminologiques adaptables au contexte sont susceptibles d'être utilisés seront limitées ici aux tâches d'organisation, de diffusion et de recherche d'informations sur les réseaux électroniques des entreprises du secteur tertiaire, en France, dans la première décennie du XXI^e siècle. Ces tâches particulières devront être repalquées dans les évolutions récentes de l'organisation du travail, de même que les intranets, qui matérialisent l'introduction des nouvelles technologies de l'information et de la communication (NTIC) dans les entreprises. C'est à partir du double constat, d'une part, des dysfonctionnements qui affectent la réalisation des activités d'information et, d'autre part, du caractère inadapté des réponses apportées par les systèmes

¹ *Grand Robert de la langue française*, version électronique 2.0, 2005.

² *Ibid.*

d'organisation de l'information, que l'on sera conduit à proposer la mise en place d'outils aux caractéristiques particulières, les référentiels terminologiques adaptables au contexte.

La construction de tels outils nécessite un modèle capable de rendre compte des différentes composantes présentes dans tout contexte d'énonciation. S'il est nécessaire de préciser les caractéristiques principales du cadre réel de production des énoncés, ces informations comportent un intérêt en tant qu'elles sont présentes dans les contenus des textes exploités pour la construction des référentiels. C'est donc une perspective pragmatique englobante qui a été retenue¹. Pour l'analyse de discours et le courant pragmatique, un énoncé et sa situation d'énonciation agissent l'un sur l'autre : le premier tient son sens autant de la situation d'énonciation que de la structure linguistique de la phrase, la seconde intègre dans sa configuration l'effet de la parole. Ce double mouvement, auquel s'ajoutent les relations entre énoncés – ou intertextualité bakhtinienne –, se matérialise dans les discours mêmes sous la forme de traces : *« Si l'on admet que l'influence exercée par l'énoncé est avant tout une influence prétendue, la construction imaginaire d'une sorte d'environnement idéal, et si l'on admet d'autre part que la situation déterminant le sens est en grande partie celle projetée par l'énoncé lui-même, on est amené à établir d'étroites relations entre les deux pragmatiques [le double mouvement dont il vient d'être question, présenté par O. Ducrot comme « pragmatique 1 et 2 »] : elles concernent l'une et l'autre la construction du monde par la parole². »*

On assignera pour objectif aux référentiels terminologiques adaptables au contexte, de rendre compte de la manière dont le monde est construit par la parole dans des situations de communication précises. L'option pragmatique s'insère donc dans un courant plus large, qui adopte pour sources des discours situés et dont la lexicologie sociale de G. Matoré³, et plus récemment, la socioterminologie de F.

¹ « D'une manière tout à fait générale, on définira la pragmatique comme l'étude de l'usage du langage, par opposition à l'étude du système linguistique, qui concerne à proprement parler la linguistique. Si l'on parle de l'usage du langage, c'est que cet usage n'est neutre, dans ses effets, ni sur le processus de communication, ni sur le système linguistique lui-même. » in J. Moeschler et A. Reboul (1994 : 17).

² O. Ducrot (1995 : 134).

³ G. Matoré (1953).

Gaudin¹, jalonnent le développement depuis une cinquantaine d'années. On fait alors l'hypothèse que la restitution des variations d'un sociolecte est de nature à guider certaines activités d'information, en présentant des images linguistiques des publics destinataires d'un site intranet. L'établissement de telles ressources doit s'appuyer sur une méthodologie qui constitue en même temps une forme de cahier des charges. Celui-ci devra prendre en compte non seulement le contexte particulier d'élaboration des référentiels terminologiques, mais aussi garantir le caractère opérationnel des techniques utilisées. Après avoir présenté les principales caractéristiques du contexte dans lequel les référentiels terminologiques doivent prendre place, on proposera un modèle devant servir de référence pour la constitution de telles ressources. Chacune de ces deux étapes proposera une articulation différente des composantes de la problématique retenue : terminologies, intranets et référentiels.

¹ F. Gaudin (2003).

I. ÉCHANGES LANGAGIERS ET PROBLÉMATIQUES INFORMATIONNELLES DANS LES ENTREPRISES

A. Évolutions récentes de l'organisation du travail

1. Le modèle taylorien d'organisation du travail

Les activités liées au langage, et, au sein de celles-ci, les activités d'information, ont pris une part croissante dans l'organisation du travail. La structure complexe d'un groupe bancaire français, le Crédit agricole, permet de fournir un exemple du caractère complexe que peut revêtir le fonctionnement des intranets dans de telles organisations. Dans ce nouveau contexte, les questions relatives à l'organisation de l'information appellent des réponses novatrices qui ne peuvent plus s'appuyer sur les seules propositions des systèmes d'organisation de l'information traditionnels.

Les activités liées à la gestion de l'information dans les organisations, et plus particulièrement dans les entreprises, ont évolué avec l'organisation du travail. Il semble donc nécessaire de commencer par présenter les caractéristiques essentielles de cette évolution. On s'appuiera pour cela sur les descriptions fournies par les sociologies du travail et de l'entreprise. Celles-ci ont dégagé les grands modèles d'organisation du travail, qui ont été adoptés jusqu'à nos jours dans les sociétés industrielles. À l'aide des travaux de R. Sainsaulieu, on opposera aux tendances actuelles le modèle taylorien, historiquement plus ancien. On empruntera ensuite aux travaux de P. Zarifian un modèle décrivant une organisation du travail émergente, en admettant que les réflexions de cet auteur, bien que tirées d'investigations majoritairement conduites dans des entreprises industrielles, puissent être extrapolables à l'entreprise bureaucratique qui nous intéresse plus particulièrement ici¹. C'est à partir de cet éclairage que les problématiques liées à l'usage de l'information seront ensuite abordées de manière plus spécifique.

¹ « L'entreprise bureaucratique, présente dans les milieux tertiaires publics et privés (administrations d'État, entreprises publiques de services, banques, assurances...), se caractérise

La question de la communication dans l'entreprise semble se poser actuellement en des termes différents de ceux promus par le modèle taylorien d'organisation du travail et les différentes adaptations qu'il a connues jusqu'à nos jours. Ce modèle, qui est encore répandu dans les économies industrielles, vise, dès son origine, à mettre en place une organisation du travail qui « *repose sur une analyse scientifique des temps, des gestes, des pauses, de telle façon qu'en échange d'une somme d'argent horaire, ou à la pièce, considérée comme stimulante, on arrive à une productivité maximale de l'ouvrier. Plus la tâche est complexe, plus il faudra que l'analyse des facteurs techniques, psychologiques et économiques de la réussite soit poussée loin. Au bout du compte, on est en mesure de décrire un mode opératoire considéré comme le meilleur : le one best way qui sera traduit en gammes ou en séquences d'opérations à réaliser successivement, en un temps minuté pour chacune, de telle façon que le cycle opératoire soit parfaitement prévisible*¹. »

Dans cette organisation, les tâches sont décomposées, isolées en fonction de leur nature (la fabrication, le contrôle, le stockage, etc.), et spécialisées, avec pour conséquence que ce n'est pas le travailleur qui organise son travail : « *Il faut s'en remettre à une série d'experts en études des temps, des complexités techniques, des rédactions de règlements de travail, etc. Bref, à côté du travailleur, d'autres fonctions exigeant d'autres compétences sont confiées aux bureaux des méthodes pour préparer le travail ; aux services du contrôle pour apprécier les résultats ; aux agents de maîtrise pour surveiller la bonne application des règles*². » Cette organisation du travail dispense le travailleur de produire un discours sur sa propre activité, non pas qu'il n'y ait pas de *Paroles au travail*³, mais il n'est pas admis par la Direction que ce type de discours puisse légitimement influencer sur les pratiques. Comme le souligne P. Zarifian : « *La grande force du taylorisme aura été d'établir*

surtout par une logique économique de statut quo où la performance est atteinte par l'effet d'un double mécanisme d'adaptation sociale [reconnaissance d'une nouvelle professionnalité pour les agents en contact avec le public et bonnes conditions de travail et développement de projets personnels pour les autres] dans un contexte de secteurs protégés et d'organisation rationnelle. » R. Sainsaulieu (1997 : 274).

¹ R. Sainsaulieu (1997 : 47).

² *Ibid.* p. 48.

³ Pour reprendre le titre de l'ouvrage collectif dirigé par J. Boutet (1995)

un discours sur le travail et la technique, un domaine discursif propre, et, avec lui, un champ d'expérience partagé, notamment, par les agents des méthodes, les ingénieurs, mais aussi par les formateurs, les psychologues, les sociologues, etc., et dont les ouvriers auront été exclus¹. » Cette forme d'organisation a ainsi entretenu de manière structurelle les clivages linguistiques, entre le langage de l'atelier ou du bureau et celui de la direction, mais aussi, entre jargons de métiers spécialisés au sein de la même entreprise, activités du secteur tertiaire comprises : celui de l'informaticien, celui du comptable, celui du commercial, etc.

2. L'émergence d'un modèle de la coopération

a. Caractéristiques générales de l'évolution récente de l'organisation du travail

Autant le modèle taylorien semble convenir pour rendre compte de l'organisation du travail dans l'économie industrielle de l'après Seconde Guerre mondiale, autant il semble peu adapté pour expliquer des comportements apparus dans le contexte de crise des années 1980. Ces évolutions se jouent à deux niveaux : celui de l'entreprise et celui du salarié. Pour la première, R. Sainsaulieu brosse le tableau suivant en ce qui concerne l'organisation du travail : « (...) *la production repose sur l'appel croissant aux capacités d'autonomie personnelle et collective des milieux opérationnels, aussi bien dans les secteurs industriels que dans les services et les administrations. Cette évolution, qui souligne la fin probable du taylorisme comme mode organisationnel dominant, s'inscrit dans la complexification, la flexibilité et la commercialisation de la production².* » Le même auteur montre en parallèle l'évolution qui caractérise le travail des salariés : « *La situation actuelle des entreprises : complexité des techniques et des organisations, flexibilité de la production, gestion des ressources humaines, en développant considérablement l'autonomie dans le travail et les modalités dans les structures, conduit les individus à vivre la position d'acteur de façon beaucoup plus subjective³.* » Ces évolutions conjointes des entreprises et du travail des salariés sont à la source d'un nouveau

¹ P. Zarifian (1996: 35).

² R. Sainsaulieu (1997 : 423).

³ *Ibid.* p. 301-302. Renaud Sainsaulieu renvoie sur ce point à l'ouvrage éclairant de B. Ollivier *L'acteur et le sujet – Vers un nouvel acteur économique*, Paris, Desclée de Brouwer, 1995, 476 p.

modèle d'organisation du travail profondément différent du modèle taylorien, dans lequel priment les seules rationalités économiques et techniques : « *La régulation de légitimation des objectifs de l'entreprise ne peut plus alors résulter d'affirmations idéologiques ou technocratiques ou de seuls rapports de puissance dans un monde où l'appel à la subjectivité devient source de construction de l'avenir. Antérieurement fondée sur l'inculcation de cultures dominantes, ou sur la pratique de négociation entre acteurs sociaux, la justification de compromis doit, à présent, résulter de séquences de débat où s'argumentent et se confrontent les orientations personnelles comme moyen d'inventer les choix collectifs. Pour reprendre la conceptualisation marxiste de la construction du sens de l'histoire, c'est explicitement la superstructure culturelle, à condition d'y voir avec J. Habermas une véritable dynamique d'échanges et d'agir communicationnel entre acteurs en procès de subjectivation, qui oriente la construction des objectifs et du sens de l'histoire, et non pas d'abord les rapports de forces issus des infrastructures matérielles et de leurs effets d'aliénation*¹. »

Ces réflexions peuvent être éclairées à l'aide d'un modèle plus proche des acteurs, permettant ainsi d'aborder à l'échelle des pratiques, l'articulation entre l'évolution du travail et le cadre dans lequel cette évolution s'inscrit.

b. Paradigme de la coopération et nouvelles évidences culturelles

Pour P. Zarifian², les tâches classiques de la production étant de plus en plus automatisées, l'activité dans le travail est amenée à se positionner non plus par rapport aux tâches, mais par rapport aux événements qui viennent perturber les modes opératoires en place. Il peut s'agir de dysfonctionnements internes mais aussi de problèmes inédits posés par l'environnement. Dès lors, il est demandé aux salariés d'anticiper, de savoir réagir, de suggérer des adaptations, ce qui implique des initiatives, la mobilisation de réseaux d'acteurs ; bref, le travail doit sortir de son mode d'organisation habituel et tenir compte d'une interdépendance accrue. Alors qu'il semblerait que dans le travail on ait de plus en plus recours à la

¹ *Ibid.* p. 302.

² P. Zarifian (1996 : 31 et suiv.).

communication, comprise au sens de coopération¹, cette évolution se heurte au modèle de division du travail issue du modèle taylorien. Il existe ainsi, dans l'organisation actuelle du travail, une tension qui résulte de la coexistence des exigences associées aux deux modèles.

Le paradigme de la coopération sous-tend, selon P. Zarifian, la diffusion de nouvelles « évidences » culturelles, de nouveaux lieux communs de la production moderne, sans que pour autant il soit encore possible d'en conclure que le nouveau système a remplacé l'ancien. Néanmoins, à partir d'expériences de terrain, P. Zarifian parvient à préciser les contours du modèle émergent :

« 1/ dans l'univers productif se déploie l'activité de sujets singuliers et autonomes, centres d'initiative, auteurs et responsables de leurs actes, et, en tant que tels, capables d'assurer les formes modernes de la production ;

2/ ces sujets ne peuvent agir qu'au sein d'un travail coopératif, en assurant leur participation à une activité commune, évaluée en tant que telle quant à ses performances ;

3/ cette participation s'appuie sur une communication transversale authentique, c'est-à-dire orientée vers la réalisation d'accords partiels quant à la définition des problèmes, l'élucidation des objectifs, le choix des moyens, la menée des plans d'action. C'est de la solidité de ces accords et de l'initiative qui en résulte pour les concrétiser que découle l'engagement personnel de chacun dans l'activité commune ;

4/ cette participation appelle un principe d'engagement et de respect mutuel entre chaque personne et le collectif où elle travaille. Cet engagement et respect sont de

¹ P. Zarifian (1999 : 16) : « La version « forte » de la coopération pousse à des remises en cause plus radicales. Dans cette version, il s'agit bel et bien de travailler ensemble, et pas seulement de coordonner des travaux séparés. Et travailler ensemble, dans les conditions modernes de la production, c'est communiquer, au sens de construire et de développer un espace d'intersubjectivité, de compréhension réciproque, c'est établir des accords solides sur :

- la nature des problèmes à traiter et des savoirs à développer ;
- l'identité (au moins partielle) des objectifs, et donc le processus commun de définition de ces objectifs ;
- le sens donné aux actions, et donc les valeurs qui les fondent ;
- les implications subjectives « croisées » des individus qui agissent ensemble, et donc la convergence des mobiles de ces individus (qui est beaucoup plus qu'une simple convergence des actes). »

nature morale, au sens pratique du terme, car ils conditionnent, beaucoup plus que le ferait un contrat formel, la solidité de la coopération, sa pérennité dans le temps ;

5/ cette participation suppose que chaque individu puisse, à égalité avec les autres, se former, s'informer, s'exprimer, argumenter, proposer. Les différences de niveau de responsabilité ne doivent pas influencer sur cet accès ;

6/ cette activité, à la fois individuelle et collective, se valide, in fine, dans la manière dont un client ou un usager accueillera les fonctionnalités qui lui sont offertes. Ce rapport de service à un client ou un usager est donc constitutif de l'activité elle-même et peut être intégré en amont de cette activité, en faisant, de l'usager, un partenaire de l'organisation ;

7/ cette activité reste soumise au rapport économique capitaliste. Néanmoins, pour pouvoir assurer, en connaissance de cause, ses actions, toute personne doit pouvoir accéder à l'intelligence des conditions économiques de son activité, juger par elle-même des contraintes et des possibilités d'action, pouvoir en discuter et les négocier¹. »

Cette description des caractéristiques concrètes du nouveau modèle permet d'articuler entre elles ses deux composantes principales : le salarié et l'ensemble organisé dans lequel il exerce son activité. Afin de penser cette articulation dans un cadre qui dépasse celui de la communication entre les membres d'une même équipe de travail, P. Zarifian tente de concilier les apports théoriques de l'« agir communicationnel » habermassien, qui n'a pourtant pas été conçu pour se développer dans un milieu régi par les rapports économiques², et la « communauté »

¹ P. Zarifian (1996 : 168-169).

² *Ibid.* p. 117.

Dans la *Théorie de l'agir communicationnel*, J. Habermas distingue deux grands types d'interactions sociales, selon qu'elles sont orientées vers l'intercompréhension ou le succès. « *Dans le cas où les acteurs s'orientent exclusivement vers le succès, autrement dit lorsqu'ils se concentrent exclusivement sur les conséquences de leur action, ils essaient de parvenir aux fins qui motivent leur action en influant, pour ce faire – par des moyens extérieurs, usant de la « carotte et du bâton », de menaces ou de promesses séduisantes –, sur les termes qui définissent la situation et, en l'occurrence, sur les décisions ou les motifs de son vis-à-vis. La coordination des actions conduites par des sujets qui procèdent de cette façon, c'est-à-dire stratégiquement, dépend de la manière dont s'enchaînent les calculs égocentriques des profits.* » À cette forme d'activité, assimilable à l'activité rationnelle en finalité de Weber et pouvant s'appliquer aux entreprises capitalistes, J. Habermas oppose « l'activité communicationnelle ». Celle-ci « *se produit lorsque les acteurs essaient d'accorder leurs projets d'action de l'intérieur et de ne tendre vers leurs buts respectifs qu'à la seule*

weberienne, entendue comme le lieu d'expression d'une activité « communautaire en entente¹ », et, semble-t-il, mieux à même de rendre compte des formes d'activité concrètes. Ainsi, lorsqu'il précise son propos, P. Zarifian cherche à se positionner par rapport à ces modèles de référence, tantôt sur le mode de l'interrogation, tantôt sur celui de la négation : « (...) *quelle peut être la forme d'une organisation qui ne soit pas une institution et qui autorise cependant la production d'effets de socialisation, avec une rationalisation des expectations, qui permette d'être une organisation productive efficiente ?* » Quelques pages plus loin, il précise ainsi les conditions de développement de l'activité communautaire en entente : « *Une telle communauté ne se développe : - ni par la fiction ou la réalité d'un accord préalable formellement organisé, qui « instituerait » et initialiserait l'existence communautaire ; ni par l'effet du fonctionnement d'un appareil idéologique ; ni par l'imposition de normes et de réglementations à partir d'un « centre ». Néanmoins, cette communauté doit comporter des règles observées par ses participants, permettre de compter sur des expectations solides, réaliser une entente, se donner des formes organisées².* » En admettant l'existence d'une telle organisation communautaire, on est amené à se demander ce qui sous-tend la communication entre ses participants. Pour P. Zarifian, « *L'opérateur de cette entente est la **culture**, non pas, bien sûr au sens culturaliste du terme, ni au sens d'un simple arrière plan de la communication, d'un monde vécu, mais au sens de la transformation des individus en sujets d'une même communauté, en ce que ces sujets s'attendent à*

condition qu'une entente sur la situation et les conditions escomptées existe ou puisse être ménagée. » J. Habermas (1999 : 148).

¹ M. Weber (1913) : Pour la notion d'« activité communautaire » : « *Nous parlerons d'activité communautaire* » [Gesellschaftshandeln] là où une activité humaine se rapporte de façon subjectivement **significative** au comportement d'autrui. » p. 347.

Pour la notion d'« entente » : « *Nous entendons par « entente » [Einverständnis] le fait qu'une activité qui s'oriente d'après les expectations que suscite le comportement d'autrui possède une chance « valant » empiriquement de voir ses expectations se réaliser, pour la raison qu'il existe objectivement une probabilité selon laquelle les autres considéreront pratiquement eux aussi ces expectations comme significativement « valables » pour leur propre comportement, malgré l'absence de tout accord préalable. Les motifs pour lesquels on peut compter sur ce comportement d'autrui sont conceptuellement indifférents. Nous appellerons « activité en entente » [Einverständnis] l'activité communautaire qui s'oriente (en tant qu'elle s'oriente) dans son développement d'après cette sorte de chances d' « entente ». » p. 371. Le « comme si » weberien correspond à l'attitude qui consiste, dans une activité en entente, à faire comme s'il y avait eu un accord préalable.*

² P. Zarifian (1996 : 160 et 172).

partager, sur des thèmes donnés, les mêmes comportements, et donc sont interpellés de la même manière¹. » Les observations relatives aux échanges d'information que l'on abordera dans la section suivante suggèrent que le modèle proposé, pour éclairant qu'il soit, rend lisible une réalité en devenir plus qu'il ne fait état d'une situation acquise.

c. Un horizon pour les échanges langagiers dans les organisations

En effet, si les univers de production industrielle et bureaucratique existent déjà, et si, par ailleurs, un nombre croissant d'expériences semblent converger vers le nouveau modèle, P. Zarifian constate toutefois qu'il manque encore un élément fondamental pour le développement de la communauté en entente dans le monde du travail : *« Ce qui manque, c'est le « comme si » de Max Weber. Un double « comme si » : comme si l'activité avait été déjà déterminée par un règlement convenu, comme si la communauté de cette activité existait déjà. « Comme si », cela veut dire que les participants déjà engagés, d'une manière ou d'une autre, dans le nouveau modèle d'organisation productive, déjà engagés dans l'agir communicationnel qui lui correspond, fassent « comme si » leur communauté existait, « comme si » les règles utilisées localement avaient valeur globale, « comme si » on pouvait désormais compter sur des attentes régulières quant à de nouveaux comportements. Faire « comme si » est une forme particulière de socialisation, qui apparaît importante dans les périodes de basculement². »*

À la lumière de ces analyses, il apparaît que la forme et la nature des problématiques relatives à l'usage de l'information, ainsi que les échanges langagiers qui les sous-tendent, doivent être appréhendés comme les résultats d'une confrontation en devenir entre deux modèles d'organisation du travail. Le modèle taylorien aura tendance à pousser en direction des salariés et des agents le discours de la Direction, alors que le modèle de la coopération cherchera plutôt des

¹ *Ibid.* p. 153-154.

² *Ibid.* p. 173. Pour le « comme si » : *« Il y a des complexes d'activité communautaire qui, tout en restant étrangers à un règlement convenu de façon rationnelle par finalité, 1) se développent pourtant, quant à leur effet, comme s'il y avait eu un accord préalable et 2) qui se caractérisent par le fait que cet effet spécifique se trouve conditionné par la nature de la relation significative de l'activité des individus isolés. »* M. Weber (1913 : 366).

passerelles entre sites discursifs, de manière à permettre le développement d'un nouveau mode d'organisation du travail.

Les relations étroites existant entre formes d'activité sociale et productions langagières ont été soulignées par Max Weber et par Jürgen Habermas. Pour le premier, la communication linguistique est présentée comme une forme d'activité communautaire en entente¹. Pour J. Habermas, l'agir communicationnel a pour cadre une situation d'action qui est aussi une situation de parole² : « *Les processus d'intercompréhension visent une entente qui dépend de l'adhésion, rationnellement motivée, au contenu d'une expression. Cette entente ne peut pas être imposée à l'autre partie pas plus qu'elle ne peut être extorquée au partenaire par une quelconque manipulation ; ce qui résulte manifestement d'une pression extérieure ne peut pas être pris en ligne de compte en tant qu'entente. Celle-ci repose constamment sur des convictions communes. On peut analyser l'avènement des convictions sur le modèle qui thématise la prise de position faisant suite à une offre d'acte de langage. L'acte de langage de l'un ne réussit que si l'autre accepte l'offre qui y est incluse en prenant position affirmativement par rapport à une exigence de validité qui, en principe, peut être critiquée*³. » Rapportées aux interactions langagières dans les organisations, les conditions de réalisation de l'agir communicationnel se posent comme un modèle possible qu'il peut s'agir d'atteindre, mais qui n'est jamais ni donné, ni acquis. Elles permettront de mesurer l'incomplétude qui caractérise les échanges d'information⁴ au sein de communautés qui sont souvent encore à l'état naissant.

¹ « Dans le cas-limite idéaltypique et « rationnel par finalité », on présente une communauté linguistique comme une multiplicité d'actes singuliers de l'activité communautaire qui s'orientent d'après l'espoir de parvenir chez autrui à la « compréhension » du sens visé. Du fait que ce phénomène se reproduit en masse d'une manière ou d'une autre chez une multitude d'individus grâce à l'utilisation dans un sens significativement analogue de certains symboles extérieurement analogues, « comme si » ceux qui parlent orientaient leur comportement d'après des règles grammaticales convenues à cet effet, nous nous trouvons en présence d'un deuxième exemple qui répond à la caractéristique indiquée au début de ce paragraphe [développement d'une activité communautaire « comme s'il » y avait eu un accord préalable], étant donné que ce phénomène est déterminé par la relation significative des actes des individus singuliers qui se parlent. » M. Weber (1913 : 367).

² J. Habermas (1987 : 296 et suiv.).

³ J. Habermas (1999 : 149).

⁴ On a présupposé que les échanges d'information se font par l'intermédiaire du langage. Les deux réalités ne se recoupent pas complètement dans leurs manifestations. Cependant, pour grossière

B. Structures organisationnelles et échanges d'information sur les intranets

1. Structures d'un groupe bancaire : le Crédit agricole

Même si l'on a déjà restreint les limites d'application des précédentes remarques aux entreprises bureaucratiques, il convient maintenant de préciser le cadre des échanges d'information auxquels on va s'intéresser. En effet, il faut convenir à ce stade que l'utilisation de l'information écrite est dépendante, en plus du type d'organisation du travail, de caractéristiques telles que le secteur d'activité, mais surtout du caractère plus ou moins lettrées des cultures du travail selon les pays¹. Par ailleurs, parmi toutes les formes de communication écrite utilisées dans les organisations², on s'intéressera plus particulièrement à celles qui sont présentes sur ce que l'on appelle les intranets. On tentera d'articuler les usages et leur cadre à partir du contexte fourni par une banque française, le Crédit agricole³. Enfin, les différents paliers auxquels cette articulation peut être observée seront présentés à partir d'une lecture libre de la grille proposée par B. Guyot pour la compréhension des dynamiques informationnelles traversant une organisation⁴.

Le contexte particulier qui servira de support à ce travail est un groupe bancaire français, Crédit Agricole S.A., au début des années 2000. En 2004, cette banque comptait 62 000 salariés, dont 50 000 en France, parmi lesquels 40% étaient des cadres. Les entreprises dénombrées dans les résultats consolidés du groupe bancaire étaient au nombre de 450, dont 43 caisses régionales de Crédit agricole. En

qu'elle soit, cette assimilation est globalement valable pour les problématiques liées à l'information écrite que l'on aborde dans la section suivante.

¹ A. Borzeix et B. Fraenkel (2005 [2001] : 120).

² Outre l'ouvrage qui vient d'être cité, on renvoie pour un panorama de l'analyse de discours appliquée à ces écrits à J. Boutet, B. Gardin et M. Lacoste (1995 : 12-31) ainsi qu'à J. Boutet (1995) et F. Bailly *et alii* (2002).

³ Pour une présentation de l'histoire du Crédit Agricole, on renvoie aux ouvrages d'A. Gueslin (1978, 1984 et 1985), ainsi qu'à J.-P. Henry et M. Régulier (1986), et pour la période récente, G. Jacoud (1999) et A. Choinel (1996 et 2002). Pour les présentations réalisées dans le cadre de l'institution, on peut citer également G. Decourselle (1988) et C. Bosséno (1994). Par ailleurs, les rapports annuels constituent une source d'information actualisée en permanence. On renverra plus particulièrement au document réalisé pour préparer la cotation du groupe bancaire « Document de référence » référencé R.01-453 par la Commission des Opérations de Bourse (absorbée depuis par l'Autorité des Marchés Financiers – AMF).

⁴ B. Guyot (2006).

2004, le groupe Crédit Agricole comptait plus de 130 000 salariés¹ répartis dans une soixantaine de pays. Enfin, le groupe bancaire avait à cette époque vingt et un millions de clients. Le graphique ci-dessus reprend les principales composantes de ce groupe.

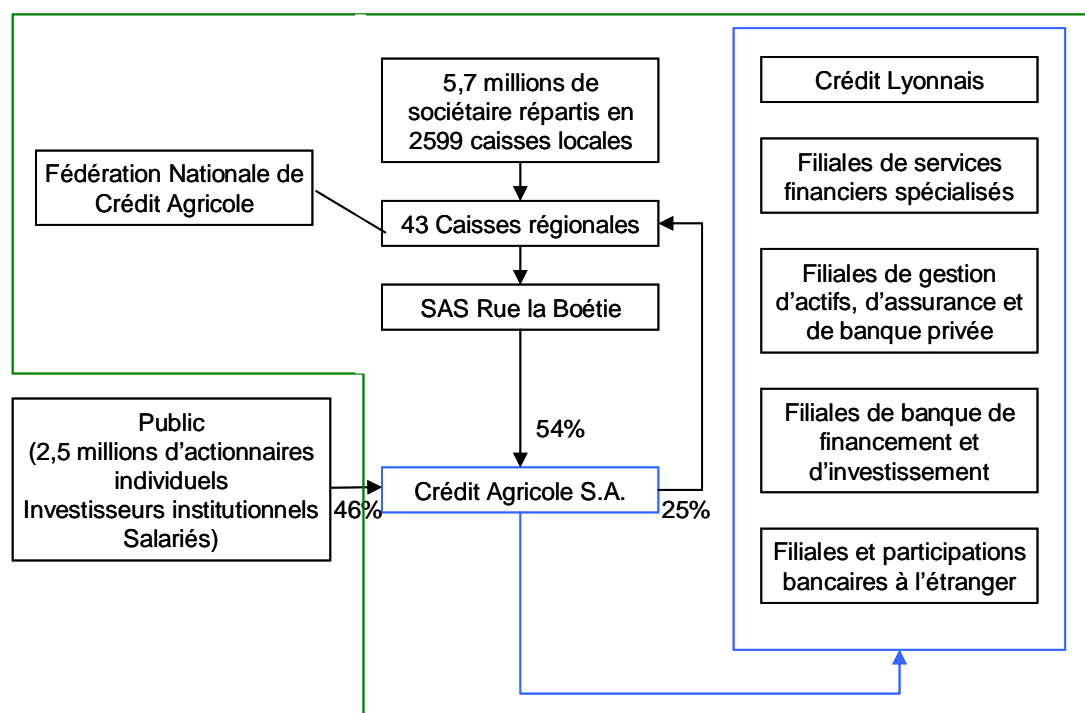


Figure 1.
Organigramme simplifié du groupe Crédit Agricole en 2004

Lecture de la figure 1 : le cadre vert dessine les contours du groupe Crédit Agricole. Le cadre bleu rassemble les composantes qui forment le groupe Crédit Agricole S.A. Les liens fléchés indiquent des relations capitalistiques. Ainsi, les caisses locales, dont le capital est détenu par les sociétaires, détiennent l'essentiel du capital des Caisses régionales, qui elles-mêmes détiennent l'essentiel du capital de la holding de contrôle SAS Rue la Boétie, qui elle-même détient 54% du capital de Crédit Agricole S.A., qui elle-même détient l'essentiel du capital du Crédit Lyonnais et des filiales spécialisées, ainsi qu'un montant variable du capital de certaines banques étrangères et 25% du capital des Caisses régionales de Crédit Agricole. Le public détient pour sa part 46% du capital de Crédit

¹ Les effectifs de ces Caisses régionales (auxquels sont ajoutés ceux de la FNCA et de SCT) s'élevaient à 67 000 personnes au 31/12/2000 et ne sont pas pris en compte pour le groupe Crédit Agricole S.A., mais pour les effectifs du groupe Crédit Agricole qui s'élèvent donc en 2004 à plus de 130 000 personnes, compte tenu des modifications intervenues entre 2000 et 2004. Document de référence de la Caisse Nationale de Crédit Agricole visé par la COB et portant le n° R01-453.

Agricole S.A. La Fédération Nationale de Crédit Agricole est une association qui n'entretient pas de relations capitalistiques avec les Caisses régionales, dont elle constitue une « instance d'information de dialogue et d'expression » (*Rapport d'activité 2004*, p. 126).

Chacune des composantes mentionnées dans les cartouches de la figure présentée ci-dessus possède naturellement sa propre organisation. On présente, à titre d'illustration celui qui est probablement le plus complexe, l'organigramme de direction de l'organe central du groupe, Crédit Agricole S.A.¹

¹ Il s'agit d'une version abrégée de l'organigramme de direction du groupe Crédit Agricole S.A. au 30 avril 2004 (*Rapport Annuel 2003*, p. 12 et 13). Les titres ainsi que les noms des directeurs ont été retirés, de même que les noms des filiales lorsqu'elles n'étaient pas placées au premier niveau de hiérarchie, comme Calyon et le Crédit Lyonnais.

Tableau 1.
Organigramme de direction simplifié du groupe Crédit Agricole S.A. en 2003

Direction générale		Formation
Finances	Budget, contrôle de gestion et Participations	Système d'information
	Finances et Comptabilité	Fonctionnement et logistique
	Fiscal	Informatique d'exploitation bancaire
	Relations investisseurs	Réseaux et systèmes d'information
Risques		Immobilier, logistique et achats
	Risques crédit	Sécurité
	Risques filiales hors Calyon	Filiale
	Risques marchés	Services financiers spécialisés
	Risques opérationnels	Fonctions centrales
	Risques Caisses régionales	Filiales
Inspection générale		Caisses régionales
	Inspection et contrôle interne CL	Relations avec les CR
	Inspection et contrôle interne Calyon	Marketing stratégique
	Inspection autres filiales	Agriculture, agroalimentaire, développement durable
	Inspection Caisses régionales	Entreprises, collectivités locales et professionnels
	Sécurité financière	Particuliers
	Administration, support et méthodes	Distribution
	Contrôle interne CA	Marque et communication clientèle
Secrétariat général		Filiale
	Affaires juridiques	Assurances et Banque privée
	Etudes économiques	Assurance de personnes
	Stratégie et développement	Assurance dommages
	Déontologie et compliance	Filiales
	Secrétariat central	Banque privée France et international
	Communication	
Ressources humaines		Crédit Lyonnais
	Gestion des carrières et recrutement	Gestion d'actifs, titres et services aux institutionnels
	Politiques RH	Banque de financement et d'investissement - Calyon
	Relations sociales	Banque de détail internationale et capital investissement

Lecture du tableau 1 : ce tableau présente une version abrégée de l'organigramme de direction du groupe Crédit Agricole S.A. au 30 avril 2004 (*Rapport Annuel 2003*, p. 12 et 13). Il est constitué de deux niveaux, le premier rassemblant des « super directions » chapeautant des directions ordinaires, rangées sur le deuxième niveau (décalé vers la droite). Afin de rendre la présentation plus lisible, les noms des directeurs, ainsi que leurs titres, ont été retirés, de même que les noms des filiales lorsque ceux-ci n'étaient pas placés au premier niveau de hiérarchie, comme Calyon et le Crédit Lyonnais.

Cet organigramme met en évidence le fait que la direction de Crédit Agricole S.A. correspond à la direction du groupe éponyme. Il ne s'agit donc pas seulement de la direction d'une entreprise, répartie ensuite entre plusieurs sous directions (les risques, l'inspection générale, etc.), mais de la direction d'une holding dans la mesure où les filiales, acteurs économiques à part entière, sont vues d'une certaine manière comme des directions. Or, ce que ne montre pas ce schéma, c'est que ces entreprises sont elles-mêmes organisées en directions, alors que les directions de Crédit Agricole S.A. *stricto sensu* sont ensuite subdivisées en départements et

services. Cette asymétrie est liée au statut particulier de Crédit Agricole S.A. dans le groupe bancaire.

En effet, Crédit Agricole S.A. joue un quadruple rôle au sein de ce dispositif. En tant que banque centrale du groupe Crédit Agricole, elle garantit son unité financière, gère sa trésorerie et assure la péréquation nationale des ressources d'épargne collectées par les Caisses régionales. En contrepartie de ce dernier point, elle accorde aux Caisses régionales des avances destinées à financer leurs prêts à moyen et long terme. En tant que tête de réseau, cette entreprise assure la cohérence du développement commercial du groupe Crédit Agricole. Avec ses filiales, dont elle est responsable, elle conçoit les produits commercialisés par le groupe en France et à l'étranger. Par ailleurs, elle constitue l'organe central du réseau constitué par les Caisses régionales, les caisses locales et les établissements de crédit créés par ces caisses. À ce titre, elle représente les établissements de crédit qui lui sont affiliés auprès de la Banque de France, du Comité des établissements de crédit et des entreprises d'investissement (CECEI) et de la Commission bancaire. Elle veille en outre à l'application des dispositions législatives et réglementaires propres à ses établissements affiliés et exerce un contrôle administratif, technique et financier sur leur organisation et leur gestion. Dans le cadre de ces attributions, elle peut prendre les sanctions prévues par les textes législatifs et réglementaires propres au groupe Crédit Agricole¹. Enfin, en tant que société mère et société cotée en bourse, Crédit Agricole S.A. cherche à développer une organisation de type industrielle avec ses filiales, promeut le développement du groupe bancaire à l'international et prend en charge sa communication financière vis-à-vis du public et des marchés. Cette position centrale au sein du groupe bancaire conditionne dans une très large mesure le volume, la nature et la forme des échanges d'information.

2. Échanges d'information et intranet

a. L'intranet

¹ Document de référence COB R. 01-453, p. 13.

Au sein d'un tel ensemble, les échanges d'information sont massifs et très hétérogènes. Aussi convient-il de délimiter ceux dont il sera question plus bas, ainsi que les systèmes informatiques qui leur servent de support. On restreindra donc le champ aux informations diffusées sur les sites intranets des entités du groupe bancaire, et plus spécifiquement sur ceux de Crédit Agricole S.A. Cela suppose que l'on commence par préciser la notion d'intranet.

Le terme « intranet » est apparu en France au milieu des années 1990. On en trouve l'une des premières définitions officielles dans un guide conçu sous l'égide de l'AFNOR : « *Type de réseau utilisant les mêmes technologies qu'Internet (protocoles et applications TCP/IP), mais uniquement pour communiquer à l'intérieur d'une organisation sur son réseau local ou sur un grand réseau privé*¹. » On peut constater une certaine stabilité dans la caractérisation de la notion, en comparant cette définition avec celles fournies par les dictionnaires de langue. Ceux-ci ne l'ont prise en compte que plus tardivement, au début du vingt-et-unième siècle. Ainsi, selon *Le Petit Robert*, « intranet » serait construit sur le modèle d'« Internet » : « 1996 ; de intra-, sur le modèle de Internet. Réseau informatique interne, utilisant les techniques d'Internet, accessible aux seuls membres d'un même groupe. « Le développement des intranets, réseaux internes d'entreprise fonctionnant avec les protocoles d'Internet » (*Le Monde*, 1996). »². »

La définition proposée par le *Merriam Webster Dictionary* est très proche : « *Etymology : intra- + net. A network operating especially like the World Wide Web but having access restricted to a limited group of authorized users (as employees of a company)*³. »

Le *Grand dictionnaire terminologique de l'Office Québécois de la langue française* donne, pour sa part, la définition suivante : « *Réseau informatique privé, à l'intérieur d'une organisation, qui utilise les protocoles de communication et les technologies du réseau Internet*⁴. » On retiendra de ces définitions qu'elles

¹ M. Amar (1997).

² Version électronique du *Nouveau Petit Robert*, version 2.1 (2001).

³ *Merriam-Webster's Unabridged Dictionary*, version 3.0 (2003).

⁴ [Grand dictionnaire terminologique de l'Office québécois de la langue française, 2003 ; consulté le 26 août 2004, mot clé « intranet ».

Adresse : http://w3.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp]

rattachent l'intranet à l'existence d'Internet au sens de World Wide Web (réseau de réseaux ou toile d'araignée mondiale bénéficiant de protocoles standard pour la diffusion, l'affichage et la création de contenus consultables via une interface graphique, un « navigateur » universel, de type Netscape, Firefox, Safari ou Internet Explorer), mais qu'elles l'en distinguent en indiquant que l'usage de ce réseau est restreint à une organisation ou un groupe d'utilisateurs autorisés.

Même si le rattachement aux technologies du Web (protocoles TCP/IP, HTTP et langage HTML¹) constitue une caractéristique toujours saillante de l'intranet, on cherchera plutôt ici à caractériser l'usage privatif d'un tel réseau. À l'Internet, réseau public, s'oppose l'intranet, réseau privé, fonctionnant à l'intérieur d'une organisation, accessible aux seuls membres d'un groupe, aux individus autorisés, et donc excluant par définition le reste du monde. On voit ainsi que ce n'est pas tellement la dimension planétaire qui est importante, sauf si elle fonctionne comme métaphore de ce qui est partageable par tous, car les multinationales sont dotées d'intranets reliés entre eux par des lignes spécialisées et échangeant des informations cryptées à travers le monde entier. En revanche, la notion d'usage privé paraît essentielle. Encore faut-il, là aussi, préciser ce que l'on entend par usage privé. L'Internet commercial permet d'utiliser des services ou de consulter des informations attachées à l'individu qui en fait usage ; mais dans ce cas, cet individu doit prouver, par une série d'authentifications, qu'il est le seul à pouvoir utiliser légalement ces services. Or, le caractère privé de l'intranet dépasse la simple sphère individuelle pour renvoyer à l'existence d'un groupe. On peut dire en ce sens qu'un intranet suppose, et donc établit ne serait-ce que de façon virtuelle, l'existence d'un groupe particulier. Mais celui-ci est défini uniquement par la faculté que possèdent ses membres d'accéder aux informations et aux services accessibles sur un tel réseau.

¹ TCP/IP : Transmission Control Protocol / Internet Protocol. Ce protocole de communication de type client-serveur est indépendant du logiciel et du matériel utilisés pour faire communiquer entre eux plusieurs ordinateurs.

HTTP : Hypertext Transmission Protocol. Ce protocole assure le transfert des données contenues dans les pages fabriquées avec le langage HTML.

HTML : Hypertext mark-up language. Langage permettant la construction de pages (mise en page, liens hypertextes, balises pour méta-données), destinées à être diffusées sur Internet à l'aide du protocole HTTP.

Afin de présenter les usages que l'on fait de l'intranet dans les organisations, il est possible de reprendre une description, certes un peu ancienne et dont le ton prosélyte paraît daté, mais qui pour l'essentiel demeure valable : « *Les services de l'Intranet visent à rapprocher, dans une configuration structurée et une cohérence commune, les applications les plus fréquemment utilisées dans la vie de travail quotidienne. L'échange de courrier électronique, la consultation d'annuaires, le partage de documents, la validation de comptes rendus, la gestion de projets, l'accès à des bases de données, l'exploitation de formulaires informatisés, la formation à distance, les forums de discussion comptent au nombre de ces usages répandus*¹. » Ce tableau, qui date de la fin des années 1990, doit être complété par la principale innovation apparue depuis, et qui réside dans la mise en place de portails intranets. Ceux-ci ont commencé à être largement diffusés au début des années 2000, en même temps qu'un nouveau langage pour l'organisation des données, le XML². Leur mise en place a généralement correspondu au souci de rationaliser l'accès à l'information par l'intermédiaire de sites intranets devenus trop foisonnants et plus ou moins bien organisés. Ils consistent en un « *intranet qui donne au personnel d'une entreprise et aux partenaires de cette entreprise accès, d'une part, à l'ensemble des données et des informations qui appartiennent à l'entreprise en question, que ces données soient structurées (ex. : bases de données internes) ou non (ex. : documents de traitement de texte), et, d'autre part, à une série de sites Web ou de portails verticaux qui se rapportent à la sphère d'activité de l'entreprise. Note(s) : Un portail d'entreprise possède habituellement un moteur de recherche qui lui est propre, pour la recherche dans les documents internes, et offre la possibilité de personnaliser la page d'accueil pour une personne ou un groupe de personnes. De cette manière, les membres du personnel ont accès quotidiennement à l'information actualisée dont ils ont besoin pour leur travail, celle-ci étant concentrée et mise à jour sur le portail*³. » Le mouvement continu « d'intranetisation

¹ M. Germain (1998).

² XML : *eXtended Markup Language* ; issu du SGML (*Standard Generalized Markup Language*) apparu au milieu des années 1980, le XML est un langage informatique permettant de structurer les données diffusées à l'aide des protocoles de type Internet, sans les associer directement à leur mise en forme, comme c'est le cas avec le langage HTML.

³ *Grand dictionnaire terminologique* de l'Office québécois de la langue française, 2002.

des entreprises¹ » s'est traduit plus récemment encore par la dimension d'infrastructure acquise par l'intranet, lequel est de plus en plus intégré au système d'information général de l'entreprise.

b. Diffusion de l'information électronique au sein du Crédit Agricole

Rapporté au cadre d'un groupe bancaire, et si l'on s'en tient aux seuls sites intranets présentant des informations sous la forme de pages HTML, de pièces jointes et de liens hypertextes, l'intranet du Crédit Agricole correspond d'abord à une juxtaposition de plusieurs centaines de sites répartis dans l'ensemble des entités dont le groupe se compose. Chaque entreprise possède généralement plusieurs sites, destinés à des publics divers. Ces derniers peuvent correspondre aussi bien à quelques dizaines de salariés qu'à tout le personnel travaillant dans une entité donnée seulement, ou encore, à plusieurs centaines de personnes exerçant le même métier, mais dans des entités différentes, voire à l'ensemble des 134 000 salariés du groupe. Cela ne signifie pas que tout site soit potentiellement visible par tout intranaute : les différents réseaux sont plus ou moins interconnectés, les contenus sont plus ou moins profilés et réservés à certaines populations qui ne les consultent que si elles ont les droits nécessaires pour le faire.

Si l'on se rapporte à l'organigramme fourni dans le tableau 1, chacune des soixante lignes, correspondant aux dénominations de « super-directions » ou de directions ordinaires, est susceptible d'être associée à un site intranet, si ce n'est à plusieurs, selon l'importance des départements ou des services que la direction recèle. Or, si l'on rapproche l'organigramme simplifié du groupe bancaire des fonctions de son organe principal, Crédit Agricole S.A., il devient possible de caractériser plusieurs sortes d'informations présentes sur les différents sites de cette entité. Les fonctions régaliennes d'organe central induisent la diffusion de normes internes relatives à l'exercice du métier bancaire. Ces normes, pouvant provenir de réglementations européenne ou nationale font l'objet d'une interprétation propre au Crédit agricole. Les normes et leurs interprétations débouchent sur des procédures internes qui doivent également être tenues à jour en permanence. La fonction de

¹ Pour le néologisme, on renvoie à l'article de F. Alin dans M. Germain et C. Malaisson (2004 : 21).

banque centrale induit, quant à elle, des échanges relatifs aux positions des Caisses régionales, indépendamment des flux financiers générés par le mécanisme de redistribution interne de l'épargne. La tête de réseau ne se contente pas de diffuser les orientations de la politique marketing devant être suivies par les réseaux de distribution. Elle doit également diffuser les informations relatives à l'accompagnement de la commercialisation des produits. Cela va de la fiche technique des produits destinée aux gestionnaires de compte dans les agences, à la documentation commerciale destinée à la clientèle, en passant par les bagages de formation des commerciaux, les formulaires types et les argumentaires de vente. La société cotée diffuse, pour sa part, les informations « officielles » relatives au pilotage du groupe bancaire et à son activité sur différents marchés, nationaux et internationaux. Enfin, la société mère diffuse des recommandations, des standards techniques, des retours d'expérience susceptibles de devenir des modèles. Ces informations descendantes n'ont pas toutes la même valeur contraignante, mais la plupart d'entre elles sont de nature à intéresser les différentes entités du groupe.

Les sites de cet intranet sont également utilisés pour diffuser des informations plus ciblées en direction de certains métiers relatifs à la fiscalité, aux risques, aux contrôles, à la gestion du personnel ou à la logistique. Enfin, le public peut dans certains cas se réduire au seul personnel impliqué dans de vastes projets réglementaires, ou concernant le développement des activités internationales du groupe. Dans ce type de projets, les visiteurs peuvent être amenés à contribuer eux-mêmes aux contenus présentés par les sites¹. Cependant, les technologies du Web 2.0² ne sont encore que très peu utilisées dans les intranets du groupe bancaire à cette époque. La caractéristique dominante des sites informatifs est encore celle d'une asymétrie entre les informations diffusées et leur provenance, celle-ci

¹ Ces échanges, lorsqu'ils ont lieu dans un cadre collaboratif, se traduisent par des partages de fichiers, de calendriers, qui sont la plupart du temps invisibles à partir des sites intranets. Cela tient, d'une part, au fait que ce sont souvent des technologies différentes qui sont utilisées, tout comme la messagerie fait partie de la « panoplie » intranet sans être généralement intégrée directement aux sites informatifs, et d'autre part, à la nature des échanges dans le cadre des projets qui génèrent des documents de travail dont seule la version définitive est généralement présentée sur les sites.

² Il n'y a pas actuellement de consensus sur la définition du Web 2.0. On peut dire de façon assez générale qu'il s'agit de technologies et de sites Internet qui encouragent la participation des visiteurs. Entrent dans cette définition des sites tels que *YouTube*, *MySpace*, *Wikipedia*, les *blogs*, les forums en ligne, les sondages associés aux pages d'un site, etc.

empruntant généralement des circuits parallèles à celui offert par les sites, afin de remonter jusqu'à l'entité centrale du groupe bancaire. Il peut alors s'agir de données collectées dans diverses applications informatiques, de rapports plus ou moins confidentiels expédiés en version papier ou par messagerie électronique. À l'autre bout de la chaîne, la réutilisation de l'information dans les entités vers lesquelles elle est diffusée reste assez mystérieuse, même si elle fait l'objet de contrôles, principalement lors de dysfonctionnements avérés. Ces derniers points peuvent être précisés plus finement lorsque l'on s'intéresse aux activités d'information.

C. Activités d'information

1. Un agent producteur, utilisateur et gestionnaire d'informations

L'activité d'information prend une place considérable dans les tâches incombant aux agents engagés dans l'évolution de l'organisation du travail qui a été brièvement décrite plus haut. Cette activité peut être définie laconiquement comme « *chercher, classer, écrire et lire*¹ » ou bien être décrite sans que cette liste de tâches l'épuise. Son observation même révèle que les agents sont loin de posséder une représentation claire de cette forme d'activité: « *L'étude de l'activité d'information s'avère délicate, tant elle est devenue inconsciente et routinière. Cependant, nous sommes toujours surpris de constater combien, une fois incité à parler du déroulement d'une journée de travail en y associant les informations mobilisées et les démarches entreprises pour les obtenir, l'interviewé s'étonne de la multiplicité des tâches, de la variété des modalités d'information qu'entraîne son activité principale ainsi que de la diversité des démarches et des savoir-faire qu'il met en œuvre*². ». Cette activité place l'agent successivement dans les positions de producteur-transmetteur d'informations, d'utilisateur et de gestionnaire qui vont être brièvement présentées.

¹ B. Guyot (2006 : 151).

² B. Guyot (2006 : 130). L'auteur appuie ses remarques sur des entretiens conduits avec une centaine de personnes travaillant dans le secteur tertiaire.

Les écrits de travail constituent un champ relativement nouveau pour lequel une approche pluridisciplinaire semble fructueuse, comme en témoignent les études du groupe *Langage et Travail*¹. Ces derniers ont permis de mettre au jour une grande diversité d'écrits pour lesquels un classement consensuel reste, semble-t-il, à construire. On se contentera ici de donner quelques exemples ayant pour point commun de se détacher des notes personnelles et des échanges relativement peu formalisés, comme ceux qui caractérisent les messages électroniques entre collègues, mais qui se rencontrent peu sur les sites intranets. Outre les documentations de référence produites au nom de l'organe central du groupe bancaire, les comptes rendus, rapports, notes de synthèse et supports de présentation des réunions de travail ou de conférences, sont fréquemment produits de manière anonyme. Cet anonymat ne concerne souvent que l'association du nom d'un auteur précis au texte produit, mais cela ne signifie pas que, dans le contexte de travail, le rédacteur ne soit pas facilement identifiable². Ce rapport distendu se rencontre également fréquemment dans des écrits primaires ou secondaires³ publiés directement sur les pages des sites d'un intranet : alors que le gestionnaire de contenu ne permet au rédacteur de produire et de publier son texte que s'il a été authentifié par le système (à l'aide d'un identifiant et d'un mot de passe), la page publiée pourra très bien être associée au nom d'une autre personne, désignée pour répondre aux questions des visiteurs du site. Selon la portée de l'écrit, il semble que celui-ci soit généralement produit au nom d'un auteur collectif allant de l'entreprise,

¹ On renvoie à l'ouvrage déjà cité, coordonné par A. Borzeix et B. Fraenkel, et en particulier aux chapitres de ce dernier auteur « La résistible ascension de l'écrit au travail » et « Enquêter sur les écrits dans les organisations », respectivement (2005 : 113-142 et 231-261).

² De ce point de vue, la signature d'un responsable hiérarchique peut n'identifier que de manière anecdotique une personne pour désigner de façon plus substantielle le collectif anonyme correspondant au personnel d'une direction ou d'une entreprise, comme la marque de fabrique associée à un produit quelconque.

³ On emprunte cette distinction au découpage classique dans les sciences et techniques de l'information entre document primaire et secondaire. Le premier présente « *une information à caractère original, c'est-à-dire lue ou vue par le lecteur dans le même état où l'auteur l'a écrite ou conçue.* » Un document secondaire présente « *des informations résumées (analytiques) et paratextuelles (sythétiques) de documents primaires.* » S. Cacaly et alii (2004). Sur les intranets on rencontre fréquemment des documents primaires sous la forme de fichiers joints, auxquels un texte introductif ou une synthèse sont joints en guise d'introduction. L'écrit secondaire a souvent pour vertu principale d'éclairer le visiteur sur le contenu des fichiers afin de limiter une ouverture systématique pour leur lecture qui peut s'avérer longue, si ce n'est rédhibitoire.

pour des documents réglementaires, contractuels ou relevant de la communication institutionnelle, à l'unité pour des comptes rendus de réunion. Cet effacement de l'auteur est comme la marque du passage d'un écrit individuel au document résultant d'un co-travail et ayant les caractéristiques formelles admises par l'entreprise : *« Tout passage à une publication suppose des règles, la construction d'un consensus sur les méthodes d'accès, d'enrichissement et la circulation documentaires. Toute collectivité partage un code commun, un jargon propre au métier et dès qu'il y a un élargissement de la diffusion, la formalisation des codes d'écriture et de lecture se renforce¹. »* Cette rédaction collective dissimule le rédacteur un peu à la manière dont les écrits diffusés dans l'entreprise masquent le lecteur : le temps consacré à manipuler l'information est fréquemment perçu par le salarié lui-même et par sa hiérarchie comme une parenthèse insérée dans le cours de l'activité, sans lui être tout à fait assimilée. Ce statut aux contours mal définis concerne également l'utilisation de l'information.

L'utilisateur de l'information diffusée sur un intranet se trouve fréquemment dans une situation paradoxale, où l'alimentation en information est abondante, sans qu'elle le dispense de devoir entreprendre de nombreuses recherches. La première situation a pu être caractérisée par certains de saturation cognitive : *« Les personnes concernées, observées en situation de travail, ont l'impression d'être noyées sous une masse d'informations croissante, et de ne plus être en mesure de se centrer sur leur « vrai travail ». Elles l'imputent à l'inflation d'informations liée aux TIC, dans le contexte de changements importants des processus de production tertiaire². »* La messagerie électronique est souvent utilisée comme moyen pour pousser l'information vers les lecteurs. Aux informations importantes du jour s'adjoignent ainsi des alertes thématiques, signalant que telle information est parue sur tel site, mais aussi des revues de presse générales et spécialisées, etc. Avant même qu'il ait entrepris de consulter l'intranet pour s'informer, l'utilisateur est quotidiennement

¹ B. Guyot (2006 : 134).

² P. J. Benghozi (2002 : 36).

abreuvé d'informations sans qu'il soit toujours possible de refuser l'abonnement à ces services.

Lorsque les conditions sont réunies pour que l'utilisateur prenne conscience de son besoin d'information, c'est-à-dire lorsqu'il est parvenu à une première représentation explicite de ce besoin¹, l'utilisation de l'intranet va le transformer en usager d'un système d'information² particulier. Cette situation est susceptible de connaître d'importantes variations, selon la nature de l'information recherchée (opérationnelle ou accessoire, relative au métier ou périphérique, etc.), la plus ou moins grande expertise de l'intranaute, ainsi que sa localisation dans le groupe Crédit Agricole. En effet, la réponse à sa question peut se trouver sur l'intranet local, sur celui d'une communauté ou d'un métier, ou encore sur l'intranet national. Elle peut être accessible directement sur les pages d'un site, dans les fichiers joints de type bureautique qui lui sont associés, ou encore dans une base de données dont l'accès est restreint aux utilisateurs habilités. L'identification d'une ressource susceptible de contenir la réponse nécessite enfin que l'utilisateur puisse se repérer et naviguer dans l'arborescence qui organise l'information dans le site intranet. Or, on a vu que le rédacteur et le diffuseur de l'information n'ont pas nécessairement pris en compte les repères nécessaires au visiteur. Ce dernier peut être aidé dans cette tâche par le raccourci que fournit le moteur de recherche. Cependant, la qualité de cet outil va dépendre dans une large mesure de sa capacité à déjouer les pièges courants de la langue écrite (par exemple, l'orthographe des mots, les variantes graphiques et linguistiques, le choix des sigles ou de leurs développés, etc.), et à exploiter les informations associées aux documents (leur provenance, leur date de mise à jour, leur catégorisation, etc.). Dans les deux cas – navigation dans l'arborescence et moteur de recherche –, l'utilisateur dépend de la manière dont sa visite a été préparée par les producteurs et les diffuseurs de contenus pour l'intranet.

¹ Y.-F. Le Coadic (2007)

² Y.-F. Le Coadic (2001).

En définitive, l'activité d'information gravite autour d'un « système d'information personnel¹ » que l'agent construit pour lui-même, afin d'organiser ses production et les ressources informationnelles qu'il consomme dans le cadre de son activité : « *Cette gestion matérielle est loin d'être secondaire même si on la sous-estime. Pièces de dossiers égarées, coupures de presse qui s'entassent, fichier introuvable, font partie de la partie « volumineuse » de l'information, et ces tâches de classement prennent du temps alors même que l'on voudrait bien pouvoir s'en passer. Il faut pouvoir gérer l'accumulation de dossiers ou de fichiers informatiques, garder ou jeter ce qui constitue une documentation personnelle. Car, avant d'être prise en charge par les archivistes, cette masse d'information est gérée par les acteurs eux-mêmes qui se dotent d'un certain nombre de règles pour le faire².* » Chaque producteur dispose ainsi sur son propre ordinateur, ou dans un répertoire sur un serveur partagé, d'une partie de l'information circulant sur l'intranet. Le stockage de ces documents soulève à son tour des questions relatives à la mise à jour des informations qui, une fois stockées localement, n'évoluent plus dans le sens voulu par leur émetteur ; le classement de référence propre à chacun ne suit pas nécessairement celui de l'intranet, ce qui oblige à intégrer au minimum un double système de repères ; les adresses (signets) de ressources pointant sur des pages précises, outre qu'elles sont régulièrement « cassées » en raison des réaménagements de l'organisation des sites, détournent le visiteur d'informations nouvelles qui pourraient lui être utiles, etc.

Il existe donc une tension entre le système personnel d'information et le système commun partagé par les agents appartenant à une même organisation et dont l'intranet constitue une composante importante. Une conséquence visible de cette tension se matérialise par le fait que les informations publiées sur les sites le sont en fonction des seuls repères de l'émetteur. Ce dernier projette sur le site une organisation qui lui paraît naturelle parce qu'elle a été intériorisée comme telle, du fait de son immersion dans une organisation du travail spécifique qui façonne les activités d'information. Tant que le système de référence de l'émetteur demeure

¹ B. Guyot (2006 : 145).

² *Ibid.* p. 143.

tacite, il bloque la prise en charge des visiteurs qui réclament au contraire une exposition explicitée des informations qui leur sont destinées. Ces multiples facettes de l'activité d'information ont donc pour point commun d'être peu lisibles et difficiles à cerner pour les agents eux-mêmes. Cette caractéristique ne facilite pas une prise en charge explicite¹ de la diffusion des documents sur les intranets.

2. Activités d'information, intranet et langage

a. Un espace de publication décloisonné pour des informations compartimentées ?

Au paradoxe, évoqué plus haut, de l'agent qui souffre à la fois d'être trop informé et de n'avoir jamais fini de rechercher les informations qui lui font défaut, correspond le statut paradoxal de l'intranet lui-même. En effet, on constate que l'intranet développe surtout une forme de communication médiatisée : pages HTML de sites consultables en permanence, fichiers associés en pièces jointes, rangés dans des bases de données de type GED (gestion électronique des documents). Cela revient à dire que la majorité des échanges réalisés à travers les intranets se font par écrits numériques interposés et diffusés par différents dispositifs mis en place sur les réseaux informatiques. Ce sont ces dispositifs qui tiennent lieu d'interface principale lorsque, en dehors des échanges routiniers pouvant lier des individus travaillant habituellement ensemble, les comptables cherchent à parler aux informaticiens, les commerciaux aux experts marketing du siège, etc. Tout se passe comme si « l'entreprise intranet² » postulait un intranaute amphibie capable de passer d'un secteur à l'autre et possédant des capacités cognitives telles qu'elles lui permettent de s'affranchir des barrières invisibles dressées entre les référentiels et leurs jargons. Or, les observations qui viennent d'être faites relativement à l'activité d'information dans les entreprises viennent contredire cette représentation de l'utilisateur.

On a vu que la fabrication des documents diffusés sur les sites d'un intranet (actualités-brèves, agenda institutionnel, barèmes-taux bancaires, comptes rendus,

¹ On trouve une présentation éclairante de cette dialectique du tacite à l'implicite dans J. Girin in A. Borzeix et B. Fraenkel (2005 : 167-185).

² On reprend le titre de l'ouvrage *L'entreprise intranet : guide de conduite de projet* F. Alin et alii (2002).

documentation sur la concurrence, sur les produits, dossiers techniques ou thématiques, enquêtes, guides, résumés-synthèses à usage divers, lettres circulaires, lettres d'information aperiodiques, listes de contacts, listes de liens externes, notes internes, offres d'emploi, organigrammes, periodiques internes, présentations ou supports de reunion, supports de formation, tableaux de bord, informations pratiques, consignes de securite, regles de deontologie, etc.), se fait dans le respect d'un format precis plus ou moins integre par le redacteur. Ce format prevoit generalement des regles de diffusion liees a un usage habituel : mention des destinataires d'un compte rendu, mise en forme pour les supports de presentation orale, etc. Cependant, la mise en ligne de ces documents sur un site intranet les rend disponibles pour un public qui ne correspond plus seulement aux destinataires initiaux. On constate que ce qui relève souvent d'une deuxieme diffusion est rarement pris en charge, ou du moins, celle-ci beneficie rarement de l'eclairage necessaire a une realisation efficace.

De meme, c'est un lieu commun que d'evoquer les difficultes rencontrees par les intranauts qui se sont aventures sur des sites censés contenir l'information recherchee ; cette perplexite est rarement estompée par la lecture d'un resultat de recherche : le calcul de pertinence du moteur paraît toujours ironiquement decalé par rapport aux attentes de l'utilisateur.

Enfin, on a vu plus haut que le systeme d'information personnel des agents est plus centré sur un « bricolage » propre a chacun que tourné vers la mise en oeuvre d'une rationalite commune devant presider a l'organisation de l'information dans l'entreprise.

D'une maniere generale on doit constater que l'information diffusée sur les intranets beneficie rarement de la prise en charge necessaire qui lui assurerait le passage d'un cercle restreint de lecteurs a un public étendu, correspondant a la visibilite et donc a l'utilite potentielle du site dans l'entreprise. Cette question peut-etre abordée sous plusieurs angles qui paraissent tous complémentaires, et qui vont de l'etude des besoins des usagers a l'optimisation des outils, en passant par la formation des utilisateurs a la production de contenus pour l'intranet ou a la recherche d'information. Dans le cadre de cette etude, on a choisi d'aborder la question sous ses aspects linguistiques. En effet, il semble que la modification des

fonctionnements constatés relativement à la diffusion de l'information passe dans tous les cas par une rupture avec les usages de la langue qui en font tacitement soit un instrument transparent, soit un obstacle insaisissable. Il convient donc de présenter plus finement certains lieux propices à cette rupture dans le processus de publication d'un contenu sur un site intranet.

b. L'artisanat de la publication numérique

Afin d'aborder plus spécifiquement les étapes d'un processus assez générique de publication des informations sur un intranet, on peut distinguer trois sortes de tâches : celles qui sont relatives à la construction de l'arborescence d'un site, celles qui touchent à la création et à la publication de contenus et, enfin, celles qui concernent la catégorisation des documents ou des contenus diffusés sur le site.

La naissance des intranets a souvent reposé sur le travail de défricheurs, car même lorsque des cabinets extérieurs à l'organisation avaient pris en charge la mise en place du site, celui-ci avait été ensuite administré par un webmestre. Dans la deuxième moitié des années 1990, de nouveaux métiers sont donc apparus dans les organisations¹. Parmi les compétences requises, on compte principalement le fait de savoir utiliser les outils relevant des NTIC. Le webmestre est également chargé de diffuser autour de lui les nouvelles connaissances relatives à ces technologies. Mais la mise en place du site se fait fréquemment sans la participation du premier cercle des utilisateurs. Les contributeurs susceptibles de l'alimenter avec leurs publications habituelles ont tendance à se décharger sur le webmestre de toutes les tâches relatives au maniement des nouveaux outils, comme la simple mise en ligne d'un fichier. Mais rapidement, ce sont non seulement les outils qui leur échappent, mais aussi l'organisation de l'information sur le site. Le webmestre semble naviguer avec aisance, mais en solitaire, sur un espace virtuel qui reste étranger aux contributeurs.

¹ Les compétences des webmestres sont diverses du fait de leur recrutement : les compétences techniques initialement requises sont devenues moins prégnantes avec l'apparition des gestionnaires de contenus (outils destinés à faciliter les tâches d'organisation et de contribution sur les sites). On rencontre donc fréquemment des agents possédant au moins deux compétences : une certaine familiarité avec l'utilisation des nouvelles technologies et des compétences métier qui font d'eux des « êtres hybrides » comme comptable-webmestre, déontologue-webmestre, etc. Cette question a été abordée, entre autres, dans V. Houzelle (2005).

Le développement des intranets « métiers » – c'est-à-dire les sites pris en main par des agents non spécialisés dans la communication numérique et à destination d'un corps de métier particulier –, ainsi que l'apparition des logiciels de gestion de contenu, ont favorisé l'attribution aux contributeurs-rédacteurs d'espaces de publication, sur lesquels ils peuvent détenir une large autonomie (création de contenus, mise en ligne de pièces jointes, ajout de nouvelles rubriques, etc.). Cependant, même si cette forme d'organisation se fait en concertation, elle n'a pas nécessairement produit une amélioration de la navigation pour l'intranaute. En effet, en l'absence d'autre référence, l'organisation qui tire principalement sa cohérence d'être l'oeuvre d'un seul individu, mais qui, pour la même raison, a du mal à convenir aux contributeurs du site, a souvent été remplacée par une juxtaposition de territoires administrés par des contributeurs différents. La cohérence d'un site pâtit également de ce mode d'organisation, qui donne rapidement l'impression d'un bourgeonnement anarchique.

Avec le recul que confère une dizaine d'années d'expérimentations et de fonctionnements opérationnels, l'organisation du contenu des sites intranets apparaît comme une tâche complexe, du fait qu'elle requiert la participation de toutes les parties prenantes à la vie d'un site. Les webmestres, les rédacteurs, mais surtout, ceux qui consultent, doivent pouvoir disposer de parcours dans l'arborescence qui leur soient familiers. Au-delà de l'ergonomie et de l'utilisabilité¹, la constitution d'une arborescence pose la question du classement des informations et des points de vue nécessairement divers qui sont amenés à s'exprimer sur un même sujet.

En outre, même lorsque les connaissances propres à un métier sont prises en compte, il reste à s'assurer qu'elles sont partagées par les intranauts qui visitent le site. Aux questions de classification s'ajoutent donc celles des dénominations, des vocabulaires utilisés pour organiser l'information et la rechercher. Dans la perspective d'une extension des périmètres initiaux de diffusion sous l'effet du développement d'une forme de travail de type coopératif, le vocabulaire du métier doit être rendu accessible pour celui qui cherche à s'informer, sans avoir encore la

¹ Outre la définition qui a été donnée plus haut, on revoie sur cette question au mémoire de DESS de l'INTD (CNAM) de J. Paillaré (2005).

maîtrise du domaine d'activité couvert par le site. L'élaboration d'une arborescence suppose donc que soient intégrées deux préoccupations nouvelles dans le cadre de l'activité d'information : d'une part, l'explicitation du profil d'un lecteur modèle « étendu »¹, et d'autre part, le souci d'assurer une forme de vulgarisation des productions publiées.

Les difficultés soulevées par l'organisation des informations sur un site revêtent des formes voisines de celles qui sont relatives à la création et la publication d'un contenu informatif. On illustrera ce point à l'aide d'un exemple que l'on rencontre couramment, celui de la publication d'un document utilisé au préalable comme support d'une présentation orale réalisée dans un cadre professionnel. Une fois la présentation terminée, le travail du « rédacteur-orateur », qui autrefois pouvait s'arrêter là, se double maintenant de toutes les tâches nécessaires à la mise en ligne. On commencera par remarquer que ces tâches peuvent incomber à une autre personne, comme le webmestre du site. Si l'usage des outils de publication en ligne tend à se démocratiser dans les organisations, on est encore loin d'un mode opératoire où la présentation qui sert ici d'exemple aurait été conçue et présentée directement à l'aide d'un tel outil. Il se peut donc que la mise en ligne soit effectuée par quelqu'un d'autre que le producteur du contenu. On considèrera pour le moment que tous les aspects qui sont attachés à l'organisation des informations diffusées sur un site, ainsi que les préoccupations relatives au maintien de la cohérence d'une classification donnée, sont pris en charge collectivement, comme cela a été vu plus haut, par le webmestre et les contributeurs du site, et ne nécessitent donc pas d'être examinées dans toute leur complexité à chaque publication.

La mise en ligne d'un contenu informatif est associée à une caractérisation spécifique destinée à accompagner sa diffusion et sa réutilisation via le canal de l'intranet. Ce complément d'information peut revêtir des aspects très variables d'un intranet à l'autre et d'un site à l'autre. On présente ici une configuration rencontrée

¹ La notion de lecteur modèle « permet d'opposer le public effectif d'un texte à celui que ce texte implique par ses caractéristiques. On utilise parfois, avec une valeur équivalente, lecteur idéal. » in P. Charaudeau et D. Maingueneau (2002 : 338).

fréquemment pour la caractérisation d'un document qui n'a pas été initialement conçu pour être diffusé via l'intranet¹.

Le fichier de la présentation initiale est associé à une série d'informations qui sont entrées dans le système permettant de faire fonctionner le site. Elles ont pour fonction d'accompagner la mise en ligne du document et de le représenter sur toutes les pages où il est mentionné ainsi que dans la liste des résultats fournis par un moteur de recherche. On examinera ici cinq sortes d'informations associées au fichier initial.

Le libellé du lien hypertexte qui permet d'accéder au document, une fois celui-ci mis en ligne, peut être une variante du titre du document original. Cette modification est parfois rendue nécessaire lorsque l'intitulé n'est pas jugé assez explicite pour un type de diffusion médiata.

Même si l'usage montre qu'il aurait tendance à être négligé, un court texte introductif permet de fournir à l'intranaute, qui n'a pas nécessairement assisté à la présentation orale, les indications contextuelles utiles à la bonne compréhension du texte. Cette sorte de résumé est susceptible de supporter également la fonction de vulgarisation dont il a été question plus haut à propos de l'arborescence.

Cet élément de mise en contexte est fréquemment complété par un ou plusieurs contacts. Ceux-ci, accompagnés de coordonnées téléphoniques ou d'adresses de messagerie électronique, permettent à l'intranaute de poursuivre, par d'autres moyens, sa recherche d'informations.

Enfin, des liens hypertextes peuvent être tissés entre ce document et d'autres, à la manière des renvois que l'on trouve dans les dictionnaires et les encyclopédies. Ces liens ont également pour rôle d'alléger le premier niveau de lecture. Lorsque le texte est créé directement à l'aide d'un logiciel de gestion de contenu, les liens hypertextes allègent la première page, qui fait alors souvent office de résumé et de

¹ On se bornera à mentionner les compléments d'informations réalisés manuellement par un rédacteur. Ceux-ci sont en effet complétés par des ajouts qui sont pris en charge automatiquement par le système, comme la date de création ou de modification des ajouts, le nom du rédacteur, l'identifiant unique du complément d'information ainsi créé, etc. D'autre part, si le document avait été créé directement avec un logiciel de gestion de contenu, un outil de mise en page au format HTML aurait dû être utilisé.

table des matières. À l'aide d'une liste de signets, elle permet ainsi au lecteur pressé d'accéder directement aux paragraphes qui l'intéressent.

Ces quatre informations sont en général complétées par une série d'indications fournies de manière plus ou moins automatique par le système. Il s'agit, par exemple, de la date de publication, ou de la branche d'arborescence à laquelle le contenu est rattaché. D'autre part, lorsque l'information ne se présente pas sous la forme d'une page HTML, le nom du traitement de texte, ou de tout autre logiciel ayant servi à fabriquer le fichier est en général présenté automatiquement, ainsi que le poids informatique du fichier, mentionné en kilo-octets. Ces indications d'ordre pratique ont leur importance, puisqu'elles renseignent l'intranaute sur le logiciel qu'il doit posséder, et le temps de téléchargement auquel il doit s'attendre pour lire le fichier.

Ces cinq éléments, titre, court texte introductif, contacts, liens hypertextes, ainsi que la date de publication et les caractéristiques du fichier, sont les plus fréquemment utilisés pour donner une première idée d'un contenu informatif sur un site intranet. C'est ainsi que sont présentés la plupart du temps les listes de résultats d'une recherche d'informations, le signalement dans un « Quoi de neuf ? », ou simplement les contenus cliquables auxquels on peut accéder en naviguant sur les pages d'un site. Il en va de même pour les textes créés directement avec un logiciel de gestion de contenu. Néanmoins, dans ce dernier cas, l'accompagnement peut être plus poussé dans la mesure où il prend en compte la création du contenu jusqu'à sa publication.

La publication d'une information sur un intranet peut revêtir des formes diverses. Le rédacteur a la possibilité de publier directement l'information sur un ou plusieurs sites. Cette possibilité peut néanmoins lui être retirée par la mise en place d'un *workflow*¹ de validation et de publication, dans lequel deux autres agents ont la possibilité d'intervenir pour décider de l'opportunité de publier l'information. D'autre part, cette publication peut correspondre, dans certains cas, au signalement

¹ « *Technologie visant à informatiser les diverses étapes d'un processus ou la circulation des documents. Note : Il est courant de parler d'automatisation des processus documentaires, d'affaires, de gestion, etc.* » in *Dictionnaire de la comptabilité et de la gestion financière*, version 1.2, 2006, accédé à partir de l'adresse : <http://www.granddictionnaire.com>.

du nouveau contenu dans des zones dédiées à cette fonction sur les pages d'accueil des sites : il s'agit des rubriques « quoi de neuf », « nouveautés », « actualités » et autres. Ce signalement de surface est complété par un rangement logique plus ou moins complexe selon l'arborescence du site. En effet, un même document est souvent accessible par lien hypertexte à partir de rubriques différentes regroupant, par exemple, tous les supports en fonction de leur nature, de l'événement auquel ils se rattachent, ou encore, des thèmes qui caractérisent leur contenu. Enfin, si le site est inséré dans un ensemble plus vaste d'espaces de diffusion, le document mis en ligne peut être signalé dans une *newsletter* ou dans une liste de diffusion qui permet, pour les intranutes abonnés, de recevoir régulièrement des alertes liées à la publication de nouveaux contenus rattachés à une thématique particulière.

Qu'il s'agisse de l'accompagnement d'un fichier à mettre en ligne, de la création d'un contenu ou de sa publication, on a pu voir à travers l'exemple précédent, que le fait de travailler pour un réseau, à l'aide d'un intranet, engendre des tâches spécifiques destinées à accompagner l'information. Cet accompagnement contraint les webmestres et les rédacteurs à prendre en compte le destinataire, à faire sortir des limbes les lecteurs auxquels ces informations sont destinées. Cela ne signifie pas pour autant que cet objectif soit régulièrement atteint dans la réalité. Comment peut-il en être autrement lorsque toutes ces tâches sont le plus fréquemment accomplies sans qu'un réel décentrement ait été opéré par ceux qui produisent et organisent l'information ? Et comment ce changement d'optique pourrait-il avoir lieu dans des proportions significatives alors que ces tâches ne sont pas encore suffisamment distinguées des pratiques au service desquelles elles sont accomplies ? Démasquer l'implicite est, semble-t-il, la difficulté majeure à surmonter pour réaliser une « activité communautaire en entente » à l'aide d'un intranet. Le recours de plus en plus fréquent à la catégorisation des informations constitue de ce point de vue une avancée importante.

Les cinq ajouts d'informations présentés plus haut, même lorsqu'ils sont construits à des fins d'explicitation et présentés simultanément à l'intranute, ne suffisent pas à réussir une organisation optimale des informations. On a vu plus haut que l'organisation du site, principalement prise en charge par l'arborescence, est un

point capital pour l'accès à l'information. Mais elle constitue également un objectif difficile à atteindre tant que les tâches et les moyens nécessaires pour faire exister le point de vue du destinataire ne sont pas développés. Lorsque c'est le cas, la prise en compte de plusieurs profils d'utilisateurs se traduit par une augmentation des accès possibles pour une même information. Ces accès supplémentaires nécessitent que l'arborescence du site intègre des parcours différents, qui sont eux-mêmes générateurs de nombreux liens hypertextes. Lorsque cette solution est la seule à être adoptée et que rien ne vient la guider, les tâches de maintenance et de contrôle de l'arborescence se révèlent souvent dissuasives.

L'alternative qui a été trouvée pour alléger la maintenance d'un nombre de liens trop important consiste à ajouter aux cinq sortes d'informations présentées plus haut des indications complémentaires, destinées à préciser les caractéristiques du document. Ces critères relèvent des métadonnées, en ce sens qu'ils viennent caractériser le contenu principal du document. Ils se présentent fréquemment sous la forme de listes de critères alphabétiques ou hiérarchiques. À titre d'exemple, la nature du contenu du document, et par conséquent ses utilisations possibles, peuvent être distinguées. Cela permet au rédacteur d'indiquer s'il s'agit d'un support de formation, d'une étude, d'un mode d'emploi, etc. Il devient alors possible de procéder à des regroupements d'informations spécifiques constitués, non plus à partir de liens, mais à partir de requêtes portant sur ces critères. D'autre part, l'intranauta peut être orienté sur le site en fonction de ses besoins et se diriger directement vers les supports de formation, si c'est ce qu'il recherche. Lors de la consultation du contenu d'une rubrique ou d'un résultat de recherche, il dispose également d'un indice lui permettant d'anticiper le contenu du support avant d'avoir à ouvrir le fichier. Ces critères sont susceptibles de caractériser le thème du document, mais aussi les zones géographiques dont il traite, les groupes de travail dont il est issu, etc.

Cependant, il faut noter que le succès de ces dernières initiatives dépend pour une large part de la manière dont elles sont insérées dans une organisation du travail de type coopératif. En effet, le préalable d'une entente entre les différentes populations d'individus participant à la vie d'un site (webmestre, rédacteurs, utilisateurs de l'information mise en ligne), semble nécessaire. De même, la prise en

charge explicite de ces tâches relevant de l'activité d'information constitue aussi un préalable indispensable. Par ailleurs, on voit que l'utilité de ces critères déborde le cadre des contenus diffusés sur un seul site. Car à quoi servirait-il de se mettre d'accord sur la manière de caractériser les contenus pour un seul site sur un intranet qui peut en comporter plusieurs dizaines ?

Par conséquent, la navigation entre sites doit être également envisagée, et l'un des moyens d'y parvenir pourrait consister, dans ce cas, à adopter un vocabulaire commun pour désigner la même chose sur l'ensemble d'un intranet. Dans le cadre d'une organisation du travail en entente, cela revient à déterminer un consensus entre parties prenantes, qui acceptent de s'entendre pour définir des critères acceptables et compréhensibles par tous. Cette négociation du sens nécessite que tous les acteurs comprennent les raisons qui conduisent à la mise en place d'un tel dispositif. De la qualité de leur accord dépend l'acceptation des règles nouvelles, et le respect de celles-ci est la seule véritable garantie de l'efficacité de la solution adoptée. L'enjeu devient donc de parvenir à dépasser les catégorisations spontanées, en leur substituant une manière de catégoriser qui soit guidée par le souci de rendre explicite, pour chaque utilisateur d'un intranet, une organisation de l'information partagée.

Le bon fonctionnement d'un moteur de recherche, la construction d'une arborescence destinée à organiser la présentation des contenus publiés sur un site, ainsi que l'ajout d'informations permettant de préciser la teneur d'un document, sont des leviers utiles dès lors que l'on cherche à améliorer la diffusion d'une information auprès de ses publics. Cela suppose une véritable prise en charge des enjeux liés à la publication dans le cadre des activités d'information. Qu'il s'agisse de déjouer certains pièges de la langue, du choix des noms de rubriques et de leur degré de généralité, de l'établissement d'une sorte de résumé ou de la catégorisation d'un contenu, l'utilisation qui est faite du langage apparaît comme un point crucial. Cependant, le traitement de ce point est sous-évalué ou insuffisamment pris en compte dans le nouveau modèle d'organisation du travail qui sert de cadre au développement des intranets dans les entreprises. Or, dans un groupe comme le Crédit agricole, cette problématique revêt des aspects nombreux et divers qui

appellent un début de formalisation, afin d'être pris en compte dans les problématiques liées à la diffusion et à l'utilisation de l'information. Dès lors, il convient de se demander si ces questions n'ont pas déjà fait l'objet d'une théorisation dans un cadre suffisamment proche de celui que l'on vient de décrire, afin de vérifier le bien fondé d'une éventuelle transposition.

D. Le statut paradoxal de l'intranet au regard des systèmes d'organisation de l'information

1. Intranet et système d'organisation de l'information

Les définitions qui réduisent l'intranet à un réseau informatique privé ayant hérité des technologies de l'Internet, ne présentent qu'un aspect de ce phénomène récent. L'intranet est apparu au fil de ce chapitre comme étant l'un des lieux où peuvent s'exprimer de nouvelles pratiques d'information liées au développement d'un modèle d'organisation du travail de type coopératif. Ce déplacement de la perspective initiale a donc permis d'aborder la question de l'organisation de l'information dans les entreprises, et de montrer ce qu'implique le fait de travailler pour un réseau d'individus, reliés entre eux du fait de leur appartenance à une organisation commune, que ce soit à l'échelle d'une entreprise, ou d'un groupe bancaire comme le Crédit agricole. Dès lors, l'intranet, considéré principalement du point de vue des différents sites informatifs qui le composent, peut être conçu comme un espace au sein duquel se déploient des pratiques ayant pour but l'échange médiatisé d'informations à usage professionnel, entre des agents appartenant à une même entreprise et reliés entre eux par un réseau informatique privé ayant recours aux technologies du Web en usage sur Internet. Cette approche autorise une conception de l'intranet comme une sorte de système d'organisation de l'information. Ce faisant, il peut être appréhendé dans la perspective des sciences de l'information, qui ont une tradition déjà ancienne de théorisation du fonctionnement de tels systèmes. On peut alors se demander si, dans ce corpus de pratiques déjà recensées et décrites, certaines ne pourraient pas être transposées dans le cadre d'un intranet, afin de répondre aux questions soulevées dans la précédente section.

Afin de comparer les caractéristiques d'un intranet avec celles d'un système d'organisation de l'information, on a recours à une sorte d'idéal type d'un tel

système, qui a été proposé par E. Svenonius¹. Cet auteur s'est attaché à décrire les propriétés fondamentales d'un tel système, dont elle situe l'origine dans l'organisation qui a été mise en place pour les grandes bibliothèques publiques du XIX^e siècle, et en particulier celles des États-Unis. Cette approche a été jugée réductrice au regard des évolutions récentes apparues dans le domaine des sciences de l'information². Pour justifiée qu'elle soit, cette critique ne porte pas, dès lors que l'on recherche précisément à caractériser la logique fondamentale qui est à l'œuvre dans de tels systèmes, et non les amendements qui ont pu lui être apportés³.

Un système d'organisation de l'information complet, dorénavant SOI, se doit d'atteindre cinq objectifs présentés par E. Svenonius. Le premier d'entre eux consiste à rendre possible la localisation d'entités particulières (par exemple, un document), dans un fichier manuel ou dans une base de données. Cet objectif est atteint grâce à une recherche d'informations s'appuyant sur les attributs caractérisant ces entités, ou à l'aide de relations établies entre les entités elles-mêmes. Il doit alors être possible de retrouver aussi bien un document (*finding*) qu'un ensemble de documents possédant des caractéristiques similaires (*collocating*). Il peut s'agir, par exemple, de retrouver toutes les œuvres d'un auteur ou tous les documents traitant d'un même sujet. En outre, le système doit permettre d'identifier un document, c'est-à-dire de le comparer avec sa description ou avec un autre document ayant des caractéristiques proches. L'utilisateur doit également disposer de critères qui lui conviennent pour conduire sa recherche de documents (*choice*). Une fois le document trouvé, il doit être consultable par tous moyens, par exemple, sous la forme d'un prêt ou d'un fichier numérique dont le contenu est affichable sur un écran (*acquisition*). Enfin, un tel système doit permettre de naviguer à l'intérieur d'une base bibliographique (*navigation*), c'est-à-dire de repérer les œuvres associées

¹ E. Svenonius (2000).

² On renvoie pour plus de détails aux deux comptes rendus suivants : J. Andersen, in *Journal of documentation*, vol. 27, n°2, 2001, p. 312-315, et G. Campbell, in *Knowledge Organization*, 2000, n°3, p. 173-175.

³ On citera néanmoins les travaux de D.C. Blair (1990) et (2006). Cet auteur a proposé d'introduire dans le domaine des sciences de l'information la conception des usages ordinaires du langage développée par L. Wittgenstein, entre autres, dans les *Investigations philosophiques* (1961). On note cependant qu'une telle conception, essentiellement développée dans ses aspects théoriques, peine à s'imposer dans les systèmes d'organisation de l'information. Les développements de ce chapitre tentent d'apporter, semble-t-il, quelques éléments d'explication à cet état de fait.

à une œuvre donnée, au moyen de mécanismes de généralisation, d'association ou d'agrégation, ou bien à partir des relations associative, hiérarchique ou d'équivalence établies entre les attributs des œuvres.

Sans que cela remette en cause le bien-fondé de ces objectifs, E. Svenonius reconnaît qu'ils sont diversement appliqués dans la réalité, celle-ci offrant une sorte de *continuum* entre des systèmes minimalistes et des systèmes complets. Ainsi, cet auteur classe Internet dans la première catégorie. Concernant l'intranet, sa position n'est pas précisée. Compte tenu des observations des précédentes sections, on peut affirmer que les intranets tendent, *mutatis mutandis*, vers la réalisation de ces objectifs. Néanmoins, et c'est là un constat paradoxal, alors que ces objectifs semblent intégrés à la mise en place et au fonctionnement des intranets, leur mise en œuvre emprunte rarement les mêmes moyens que ceux qui sont utilisés pour les SOI. À la lumière des précédentes observations, on peut suggérer deux sortes de raisons pour rendre compte de ce décalage. La première tient à la remise en cause de l'ontologie des systèmes d'organisation de l'information dans les intranets. Une explication d'un autre ordre tient dans le caractère plus ou moins bien adapté du recours à une normalisation de la langue pour réaliser ces objectifs.

2. Une ontologie remise en cause

L'ontologie des SOI se compose, toujours d'après E. Svenonius, de six sortes d'entités de premier ordre. Celles-ci correspondent aux documents (*documents*), aux œuvres (*works*), aux super-œuvres (*superworks*), aux éditions (*editions*), aux ensembles constitués par les œuvres d'un auteur (*author sets*), et aux ensembles constitués par les documents partageant un même sujet (*subject sets*). On se propose donc d'examiner maintenant en quoi les caractéristiques de ces entités se rapprochent ou s'éloignent de celles des objets observés sur les intranets.

Les documents traités par les SOI relèvent en général d'une littérature publique ayant suivi un circuit éditorial traditionnel assurant le respect d'un certain formalisme et encadrant la diffusion. L'intranet propose également ce type de littérature, mais son traitement est généralement pris en charge par une structure dédiée, comme un centre de documentation. Or, pour l'immense majorité des

documents diffusés, les caractéristiques sont très différentes : il s'agit d'une littérature à usage interne, incorporant, certes, une part de littérature grise¹, mais où les écrits dominants sont assez hétérogènes en raison du fait qu'ils respectent des codes locaux à usage interne, garantis seulement par le soin de leurs auteurs et celui des responsables de la diffusion. Ce qui distingue les deux sortes de documents est fondamentalement leur rapport à l'activité. Quand bien même la littérature académique connaît des rythmes de publication variables selon les disciplines, son traitement dans les SOI, c'est-à-dire sa description et son analyse, dépend d'une temporalité propre qui est déconnectée de l'action. Les documents produits dans les entreprises et diffusés sur les intranets sont généralement étroitement liés à la réalisation de l'activité. Cette situation a pour conséquence, comme cela a été présenté plus haut, d'insérer ce traitement dans un ensemble d'activités d'informations qui ont pour propriété de ne pas être perçues et prises en charge comme telles par les agents et leur hiérarchie. À ces caractéristiques décrites s'ajoutent l'instabilité des documents publiés (par exemple, un document aura un titre différent de son nom de fichier, qui lui-même ne sera pas identique au libellé du lien hypertexte qui permettra de le consulter), ainsi que leur caractère éphémère (une publication de plus de trois ans est généralement considérée comme obsolète).

Par ailleurs, la notion d'œuvre², qui permet d'opérer une dissociation entre le support matériel et le contenu intellectuel d'un document, est peu usitée pour les documents produits dans le cadre d'une organisation. Un cadre de diffusion restreint à une organisation, une durée de vie courte ainsi qu'une relative instabilité formelle peuvent expliquer que cette distinction ne corresponde pas à un usage observé dans les intranets. Ce point est en relation avec la notion de version qui ne se rencontre pas plus sur les intranets.

¹ « Qualifiée aussi parfois de littérature souterraine, la littérature grise regroupe divers types de documents non publiés ou publiés de façon restreinte : actes de congrès, thèses, prépublications, rapports de recherche, rapports de l'administration, rapports annuels des sociétés, rapports des conseillers d'ambassade. Sans être confidentiels, ces textes, échappant aux circuits éditoriaux habituels, sont difficilement repérables et accessibles. » S. Cacaly et alii (2004).

² « Operationally, then, a work is a set or a family of documents in which each document embodies essentially the same information or shares essentially the same intellectual or artistic content. » E. Svenonius (2000 : 36).

En effet, seule la dernière version d'un document est mise en ligne, puisque c'est celle qui garantit son caractère opérationnel, c'est-à-dire le fait qu'il contienne les dernières informations à jour pour la description ou la réalisation d'une tâche ou d'un ensemble de tâches, elles-mêmes dépendantes d'une organisation du travail spécifique. D'une manière générale, une édition en chasse une autre qui est alors supprimée ou dotée d'un statut d'archive provisoire. Cependant, l'utilisation d'un document en entreprise prévalant toujours sur l'unité intellectuelle de son contenu, ce dernier est couramment répété dans des déclinaisons sans que les rédacteurs aient cherché à les relier les unes aux autres. Cette conception monolithique du document (un usage – un support), conduit à une atomisation de fait, qui rend très difficile le contrôle du caractère réellement à jour des informations en ligne. Il est par ailleurs assez difficile de tenter d'y voir plus clair en remontant les liens hypertextes jusqu'à leur auteur.

On a vu plus haut que la notion d'auteur est problématique dans le cadre des écrits d'entreprise car elle heurte la conception dominante d'une réalisation collective ou réalisée au nom d'un collectif. En effet, les documents produits dans les organisations sont perçus par leurs auteurs mêmes, comme relevant d'abord d'un travail collectif, d'un co-travail, dans lequel l'échelon hiérarchique vient s'insérer au même titre que les apports des pairs, des co-rédacteurs, des parties prenantes d'un projet, ou des membres de l'unité au sein de laquelle le document est produit. On remarque, dans ce cas, que la notion d'auteur est dissoute dans sorte une responsabilité collégiale. Il s'agit alors d'un émetteur aux contours plus diffus et qui peut correspondre à un service, un département ou une direction. Derrière la notion vague de « site émetteur », qui est parfois utilisée pour organiser les informations, se cache fréquemment celle « d'entité émettrice », qui renvoie à un secteur de l'activité de l'organisation, collectivement responsable des contenus diffusés sur tel site d'un intranet. Cependant la durée de vie de ces unités peut être courte, et leur remplacement s'accompagne fréquemment d'un changement de dénomination et d'organisation, qui suffisent le plus souvent à en faire perdre la trace. L'anonymat et la fugacité ne facilitent pas la saisie des auteurs dans les organisations, et sur les intranets en particulier.

Enfin, un système d'organisation de l'information doit permettre à ses utilisateurs d'obtenir des ensembles de documents associés au même sujet (*subject sets*)¹, à la suite d'une recherche menée à l'aide du SOI. E. Svenonius remarque que cette entité a tendance à supplanter la précédente (*author sets*) dans les usages. Un tel résultat repose sur un traitement des documents qui vise à en produire une analyse au moyen d'une langue normalisée et contrôlée². On parle alors d'indexation. Contrairement aux autres entités qui, même si elles résultent d'un travail d'abstraction, comme la notion d'œuvre, reposent sur une assise physique qui se prête à une description pouvant toujours être améliorée ou uniformisée, la notion de sujet ressortit en premier lieu au domaine intellectuel. Or, une conséquence essentielle de ce changement de nature est gommée dans la présentation qui est faite de la notion de sujet. Alors que les documents, les œuvres, les éditions et les regroupements de documents produits par un même auteur existent en dehors de leur description, les regroupements par sujets ne peuvent pas exister sous une forme stabilisée sans que les sujets aient été préalablement construits et interprétés par quelqu'un. Mais rien n'est dit au sujet des conditions matérielles sans lesquelles un tel processus ne peut exister³. C'est pourquoi pour l'exposition de ce point crucial on suivra les normes en usage dans le domaine⁴.

3. La question du recours à une normalisation de la langue pour l'intranet

Selon la norme ISO 5963-1985, le processus d'indexation est décrit de la manière suivante : « **3.8 Indexation** : Action de décrire ou d'identifier un document en ce qui concerne son contenu. » Cette définition est complétée par une description

¹ « *That all documents on the same subject must be displayed together is mandated by the collocation objective.* » E. Svenonius (2000 : 46).

² L'analyse peut correspondre à l'élaboration d'un résumé du document, à son classement dans un plan de classement et à son indexation. Le résumé se fait généralement dans la langue du document. Seules les deux autres composantes de l'analyse ont recours à une langue normalisée qui est visée par les développements de ce paragraphe.

³ On peut voir là une orientation délibérée de la part de l'auteur qui, par cette omission a évité l'un de principaux points débattus au sujet des SOI. L'approche que l'on a retenue ne permet pas de développer ce point ici. Mais on renvoie à la thèse publiée de M. Amar (2000), qui fait une analyse très éclairante sur la question de l'indexation et les présupposés linguistiques qui sous-tendent la conception dominante dans le domaine des langages documentaires. On revient plus bas sur la position d'E. Svenonius concernant ce point important.

⁴ ISO 5963-1985 et NF Z 47-102:1993.

du processus qui est ainsi résumé : « **4.2** *Lors de l'indexation, on extrait les notions des documents par un processus d'analyse intellectuelle, puis on les transcrit à l'aide du langage d'indexation. L'analyse et la transcription doivent toutes deux être faites avec l'aide d'outils d'indexation tels que thésaurus et plans de classement*¹. » L'indexation concerne aussi la recherche des documents dans la mesure où celle-ci doit s'effectuer en utilisant le même langage d'indexation, comme cela est explicité dans le schéma ci-dessous :

¹ La norme NF Z 47-102, ainsi que le *Vocabulaire de la documentation* ajoutent que cette indexation peut également se faire en langage naturel. Outre que cela ôte toute possibilité de normalisation, cette dernière option a été avantageusement remplacée par l'indexation automatique des textes diffusés à l'aide des moteurs de recherche. Ce point a d'ailleurs été retiré de la définition donnée par l'édition de 2001 *ISO 5127 : 2001 Information and documentation – Vocabulary*.

Thésaurus : « *Liste normalisée et structurée de terme acceptés à l'indexation (les descripteur) et d'équivalents. Les descripteurs sont reliés entre eux par des relations sémantiques exprimées par des signes conventionnels. Langage combinatoire, il se distingue de la classification qui relève d'une logique d'inclusion. Les descripteurs sont combinés entre eux pour décrire des documents de tout type (lors de l'indexation) et écrire les questions lors de l'interrogation d'une banque de données. Un thésaurus facilite les recherches en limitant le « bruit » et le « silence » documentaires, grâce au contrôle des homographes et au rapprochement des synonymes.* » D. Dégez, in S. Cacaly et alii (2004)

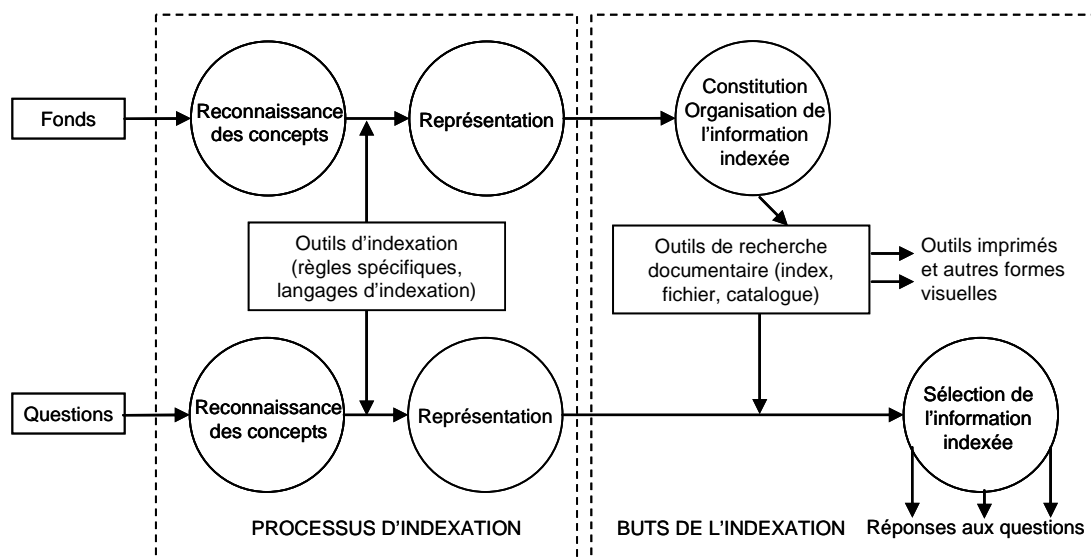


Figure 2.
L'indexation dans la chaîne documentaire¹

Ce schéma met en valeur la symétrie qui existe entre l'exploitation du fonds documentaire et les recherches : ces deux tâches sont médiatisées par le recours à une langue normalisée lors d'un processus d'indexation. Ce dernier produit deux sortes de résultats (les buts du processus) : d'une part, l'organisation du fonds et son « exposition » sous différentes formes (comme les catalogues), et d'autre part, les réponses aux questions adressées au système.

La norme, outre la description du processus, précise les qualités requises pour la réalisation du travail d'indexation relatif au traitement des documents (le premier parcours du schéma qui débute par la constitution ou l'alimentation d'un fonds documentaire). L'indexeur doit posséder les capacités d'analyse requises pour mener à bien la tâche de reconnaissance des concepts. Il lui faut une bonne connaissance du domaine, de manière à comprendre les termes rencontrés dans les documents². Par ailleurs, cet agent doit être « *familiarisé avec les outils d'indexation*³. » En effet, la compétence de l'indexeur et la qualité des outils sont considérés comme étant les principaux facteurs de la qualité et de la cohérence de

¹ Annexe A (informative) de la norme NF 47-102 : 1993 p. 2.

² ISO 5963-1985 , 8.3, p. 4.

³ ISO 5963-1985 , 7.2, p. 4.

l'indexation¹. L'indexeur doit, par ailleurs, être d'une « impartialité totale » qui garantit également la cohérence de l'indexation. Les normalisateurs reconnaissent cependant que ces qualités deviennent difficiles à maîtriser dans le cas d'équipes d'indexeurs : « *La cohérence est plus difficile à réaliser dans une grande équipe d'indexation ou lorsque l'indexation est faite par des équipes d'indexeurs travaillant dans des endroits différents, comme dans un système décentralisé. Dans de telles situations, une étape de vérification centralisée, avec rétroaction vers les indexeurs, est recommandée*². » Néanmoins, le contact avec les utilisateurs doit permettre d'améliorer cette pratique³.

Cette présentation du processus d'indexation met l'accent sur un certain nombre de points qui contrastent avec ce l'on a pu constater au sujet de l'intranet. Dans la norme, cette tâche fait l'objet d'une description explicite, qui indique qu'elle doit être réalisée par un agent spécialisé, c'est-à-dire formé et si possible expérimenté. Une telle situation n'existe pas dans le cadre de l'intranet que l'on a pris pour référence, à de très rares exceptions près, correspondant à des « niches » telles que les centres de documentation. D'une manière générale cette forme d'activité d'information n'est ni encadrée, ni soutenue par un programme de formation, ni même présentée ou nommée dans les descriptions de poste des agents. Autant dire qu'elle n'existe pas. Par ailleurs, la norme laisse entrevoir que ce processus est délicat à maintenir dès lors que les indexeurs sont nombreux ou éloignés physiquement les uns des autres. On a vu précédemment que la complexité de la structure du groupe bancaire constitue de fait un obstacle majeur à l'utilisation d'un outil d'indexation unique permettant de réaliser des indexations cohérentes. La mise en place d'un tel processus pourrait, au mieux, avoir pour contexte un site informatif. Mais quand bien même son usage serait limité à un tel cadre, la question de la prise en charge des visiteurs reste posée.

Il semble que, sur ce point, la norme repose sur un modèle de référence qui n'est explicité nulle part : son cadre d'application idéal semble être une structure de petite taille ou de taille moyenne, limitée à quelques agents dont l'activité est

¹ *Ibid.* 8.1.

² *Ibid.* 8.2.

³ *Ibid.* 8.4.

ournée vers un public restreint, qu'il est possible d'interroger à l'occasion sur sa perception de la qualité de l'indexation qui est réalisée. Le silence concernant la deuxième branche du processus (celle qui concerne l'indexation des questions), est éloquent. Il laisse le choix entre deux hypothèses : soit celui qui pose les questions doit être aussi compétent que l'indexeur, soit les questions sont posées par l'indexeur lui-même. Compte tenu des dates de publication des normes, 1985 et 1993, il est plus probable que la deuxième possibilité soit la bonne. Le rôle d'intermédiaire, traditionnellement joué par les professionnels de l'information, est alors valorisé, et la « maïeutique¹ » de l'expression du besoin d'information s'étend tacitement jusqu'aux équations de recherches², qui sont construites par ces professionnels et non par les utilisateurs des systèmes d'organisation de l'information.

En définitive, ce sont les mêmes contraintes pragmatiques qui pèsent sur les langues spécialisées et sur les langages d'indexation³ : « *Du point de vue de la pragmatique, il convient de mentionner un point qui plaide en faveur de l'unité des langues de spécialité, au-delà de leur diversité : la facilité avec laquelle les spécialistes se comprennent entre eux, même en parlant une langue différente. En effet, il paraît assez évident que le spécialiste d'une discipline donnée est capable de comprendre avec une relative facilité les textes de sa spécialité rédigés dans une autre langue, bien qu'il soit incapable de tenir la moindre conversation dans cette langue sur un sujet banal ou de comprendre une question ayant trait à la vie de tous les jours. On doit chercher la réponse à cette question dans les unités et les règles restreintes que possèdent les langues de spécialité, souvent volontairement établies*

¹ Y.-F. Le Coadic (2007 : 80).

² Une équation de recherche combine en général des termes d'indexation, des opérateurs booléens (ET, OU, SAUF) et un système de parenthèses, de manière à élaborer une question précise en respectant la syntaxe du système de recherche. Les bases de données interrogeables à distance via le Minitel pouvaient présenter de ce point de vue de redoutables difficultés.

³ Les auteurs qui ont tenté d'établir des ponts entre terminologie et langages d'indexation adoptent pour cadre ce schéma de l'utilisation des langages documentaires. On peut voir sur ce point W. Mustafa El Hadi (1992) et M. Holzem (1999). Du point de vue terminologique, le thésaurus est une sorte de terminologie : « *Dans un thésaurus documentaire, les données terminographiques sont des éléments de données terminologiques qui, d'une part, aident à maîtriser une langue d'indexation donnée, et d'autre part, indiquent les rapports entre notions.* » H. Felber (1987 : 168) ; E. Wüster (1974) semble considérer pour sa part que les thésaurus sont des sortes terminologies plutôt imprécises.

par les utilisateurs eux-mêmes, ainsi que dans leur caractère international. Ces éléments, ainsi que le sujet limité, facilitent l'intercompréhension. Plus une langue est spécialisée, plus le nombre de ses utilisateurs est restreint, et plus ses unités et ses règles sont internationales¹. » Dès lors qu'il s'agit de prendre en charge une diffusion étendue des informations, au-delà d'un public restreint de spécialistes, l'utilisation d'un langage d'indexation ne semble pas adaptée pour répondre aux problématiques de diffusion et d'organisation de l'information propres aux intranets. En effet, son utilisation ne ferait qu'ajouter une terminologie particulière à celles qui sont déjà véhiculées par les sites d'un intranet.

La forme actuelle d'organisation du travail dans laquelle prend place le développement de l'intranet du groupe bancaire que l'on a pris comme exemple impose des contraintes fortes, qui pèsent sur la réalisation des activités d'information des agents. D'une manière générale, on peut considérer que ces activités ne se sont pas encore suffisamment autonomisées pour être prises en charge de manière efficace. Il en résulte des dysfonctionnements bien identifiés, comme la difficulté que l'on éprouve à trouver une information en naviguant dans l'arborescence des sites, ou en utilisant un moteur de recherche. On constate que le point de vue des visiteurs est occulté par ceux du producteur et du diffuseur d'information. Ceux-ci sont mal outillés pour fournir à leurs destinataires les repères utiles à leurs recherches. Un point central semble résider dans la difficulté inhérente à l'explicitation de connaissances manipulées par les agents sur un mode fréquemment tacite, dans un cadre professionnel limité et contraignant, qui fait obstacle à la prise en charge de destinataires non prévus pour l'exercice habituel des activités. Cet état de fait, qui instaure un décalage cognitif entièrement à la charge des visiteurs d'un site intranet, ne résiste pas longtemps à un échange entre interlocuteurs dans une banale situation de face à face. En revanche, il peut se maintenir sans être réellement remis en cause en raison du caractère médiat des informations diffusées par les sites.

¹ M.-T. Cabré (1998 : 131-132).

On s'est alors demandé si les systèmes d'organisation de l'information, avec lesquels l'intranet partage au moins certains objectifs, ne recélaient pas des pistes de solutions réutilisables dans ce nouveau cadre. La confrontation des caractéristiques des intranets avec celles des systèmes d'information a permis de mettre au jour des différences importantes, qui semblent hypothéquer la possibilité d'une simple transposition. En effet, le fonctionnement des intranets révèle une remise en cause de l'ontologie des systèmes d'organisation de l'information. Le recours à une langue normalisée, qui aurait pu correspondre à une solution possible pour remédier aux dysfonctionnements de l'intranet, se révèle sinon impossible, du moins très compromis dans un tel cadre. En résumé, on se trouve donc dans une situation paradoxale. Car si l'intranet constitue un terrain difficile pour les questions d'organisation de l'information, on a également pu constater que certaines tâches, exécutées dans le cadre du fonctionnement des sites, pourraient constituer autant de points d'appui pour une prise en charge de la distance cognitive pouvant exister entre les contenus et leurs publics. Encore faudrait-il que ces formes d'activités d'information puissent être éclairées de manière à intégrer, d'une part, une proportion plus importante d'explicitation des connaissances diffusées, et d'autre part, les manières de catégoriser et de classer l'information propres aux publics des sites. C'est pour tenter de combler cette lacune que l'on a cherché à proposer une méthode de construction de référentiels terminologiques adaptables au contexte.

II. CAHIER DES CHARGES POUR UN RÉFÉRENTIEL TERMINOLOGIQUE ADAPTABLE AU CONTEXTE

A. Un observatoire des usages linguistiques propres à une entreprise

Les propositions qui vont suivre s'inscrivent dans un courant visant à prendre en compte certaines caractéristiques des usagers dans un processus d'organisation et de diffusion de l'information¹. Parmi les éléments pouvant être pris en compte, on a

¹ On ne cherche pas à opposer, comme c'est le cas fréquemment dans le domaine des sciences de l'information, une approche dite orientée-usager à une approche orientée-émetteur, ou système. Si certains points de vue ont eu tendance à occulter les autres en occupant une place par trop dominante, on pense qu'il est plus fructueux de chercher à combiner leurs acquis. Pour une vision du cadre

choisi de s'intéresser plus particulièrement à ceux qui relèvent du domaine linguistique. Plus précisément, et compte tenu des observations réalisées dans la première section, on a considéré que les pratiques de création, d'organisation et de diffusion de l'information sur un intranet sont susceptibles d'être améliorées, dès lors que l'on prend le soin de les informer à l'aide d'indications relatives aux manières de penser et de parler propres aux différents publics d'un site. L'inflexion des démarches informationnelles gravitant autour de la publication sur les intranets repose ici, non pas sur l'imposition d'une norme langagière commune, dont on a vu qu'elle ne semble pas adaptée, mais sur l'exposition formalisée de la diversité des usages langagiers en cours dans une organisation. Dans la mesure où ces usages peuvent être associés à des situations, des pratiques, et en définitive, à des publics différents, on postule que leur formalisation devrait permettre de mieux adapter la diffusion de l'information en fonction de ses différents destinataires.

Cette démarche ne peut prendre véritablement tout son sens que dans la mesure où elle n'est pas le fait d'expériences isolées ; elle revêt donc un aspect programmatique à l'échelle d'une organisation. Sa visée opérationnelle devra également tenir compte des normes qui sont de nature à encadrer son développement.

La mise en place d'une telle démarche suppose aussi que les usages langagiers puissent être situés, c'est-à-dire que les caractéristiques de leurs conditions de productions puissent être explicitées. Ces données permettront de délimiter les champs à l'intérieur desquels pourront être articulés situations, populations d'agents et pratiques discursives. On cherchera par conséquent à adopter un modèle de la communication qui soit à même d'intégrer ces différentes composantes.

général offert par la sociologie des pratiques informationnelles, on renvoie à l'article de J. Le Marec, in Cacaly (2004 : 219-222) et J. Le Marec (2004 : 353-372). Pour l'approche orientée usager, on renvoie aux ouvrages de Y.-F. Le Coadic, et en particulier (2001). Pour l'intégration de cette approche dans l'évaluation des systèmes de recherche d'information, on renvoie aux synthèses éclairantes de S. Chaudiron et M. Ihadjadene (2002 : 211-231), S. Chaudiron (2004 : 287-310), M. Ihadjadene (2004 : 337-351). Pour l'intégration de ces problématiques dans le cadre de l'intranet, on renvoie aux ouvrages de J. Nielsen (1993 et 2004) ainsi qu'au mémoire déjà cité plus haut de J. Paillaré (2005).

Dans cette perspective, il conviendra également de veiller à dépasser une conception prégnante en terminologie, qui cloisonne les langues spécialisées en fonction de domaines, ces derniers conférant leur cohérence aux productions terminographiques¹. On a vu plus haut que les publics d'un site intranet, non seulement ne se limitent pas aux agents oeuvrant au sein d'une forme d'activité, mais qu'ils peuvent également avoir des caractéristiques et des préoccupations différentes. Par exemple, on perçoit que les centres d'intérêt du personnel d'encadrement peuvent ne pas être les mêmes que ceux des opérationnels, ou encore qu'une recherche visant une procédure pour l'appliquer, ou seulement pour savoir ce qu'elle contient, n'induit ni les mêmes attentes, ni les mêmes comportements de la part du public d'un site. Les profils et les attentes des intranauts influent sur leur manière de rechercher l'information. Il conviendra donc d'identifier un cadre propice à l'intégration des différentes facettes linguistiques caractérisant les échanges langagiers ayant lieu dans une même entreprise. Il s'agira moins de remettre en cause la notion même de domaine, que de chercher à prendre en compte les passerelles permettant de passer d'un domaine à l'autre, ou encore, d'une spécialisation élevée à une spécialisation moindre dans l'usage de la langue.

Enfin, ce cahier des charges comportera également des contraintes liées à l'insertion de la réalisation de tels référentiels dans le contexte actuel des activités d'information relatives au fonctionnement des sites intranets.

Le besoin que l'on a identifié plus haut, et qui consiste dans la nécessité d'informer les activités d'information à l'aide d'une présentation des usages de la langue ayant cours dans une organisation, passe par la constitution d'une ressource particulière. Dans la mesure où elle est en prise avec l'activité d'une entreprise, une telle ressource doit s'appuyer sur un travail d'observation continu ou pouvant se répéter à intervalles réguliers. Enfin, il semble qu'un observatoire des usages linguistiques, du fait de sa position surplombante, soit à même de guider un usage réflexif de la langue.

¹ On en trouve une illustration, entre autres, dans la présentation du modèle wüsterien réalisée par H. Felber (1987).

En exposant les variations du lexique employé dans une organisation qui apparaissent dans telle ou telle situation de communication, il révèle aussi des territoires linguistiques. L'organisation de l'entreprise peut apparaître alors comme la juxtaposition de zones dialectales dont un tel observatoire devrait pouvoir proposer une sorte de cartographie. En objectivant les variations et en les illustrant, un tel outil doit donner à voir et à penser aux agents exerçant leur activité d'information. Comment les visiteurs de telle provenance abordent-ils telle ou telle thématique dans une situation de communication caractérisée ? Si l'observatoire des usages apporte déjà un début de réponse à une telle question, il sera à même de remplir son rôle d'informateur. S'il peut remplir un tel rôle, il fournira des indications utiles pour la création et l'organisation de l'information sur un site intranet. Il favorisera ainsi une prise en charge de nature à combler les écarts cognitifs pouvant exister entre émetteurs et destinataires des contenus publiés.

Cependant, la mise en place d'une telle ressource n'induit pas nécessairement que sa réalisation soit centralisée. Au contraire, elle devra pouvoir reposer sur l'agrégation d'observations réalisées à différents endroits du « territoire » de l'entreprise, et concerner différentes strates dans les échanges langagiers (du parler de métier à la « langue officielle » de l'entreprise, en passant par différentes terminologies en usage dans certains secteurs). Cette observation répartie des usages linguistiques doit pouvoir s'appuyer sur un élément unificateur, capable de garantir la qualité des observations, l'homogénéité de leur mise en forme, et partant, leur capacité à être échangées au sein de la même organisation. Il convient donc d'asseoir l'élaboration d'une telle ressource sur une procédure suffisamment robuste pour s'adapter à différents contextes et différentes sortes d'usages linguistiques, mais également précise, de manière à garantir une homogénéité des traitements nécessaires à son application. En effet, les résultats obtenus devront pouvoir être contrôlés et reproduits. Par ailleurs, que cette ressource constitue un ensemble autonome, ou bien qu'elle soit destinée à enrichir des outils lui préexistant, comme des terminologies, elle devra s'appuyer sur les normes existantes. Il conviendra donc d'organiser les données de cette ressource de telle

sorte qu'elles puissent venir compléter une terminologie respectant les usages standard exposés, entre autres, dans la norme relative au travail terminologique¹.

La constitution de telles ressources devra rechercher l'exemplarité, en identifiant des cas typiques, plutôt que l'exhaustivité, en visant la terminologie d'un secteur d'activité. Deux raisons motivent ce choix, qui a pour but de garantir le caractère opérationnel d'un tel dispositif. La première raison tient aux contraintes pesant sur cette tâche de nature terminographique : d'une part, le statut encore incertain des activités d'information dans le cadre de l'organisation du travail, et d'autre part, la nécessité de suivre de près les évolutions affectant l'activité de l'entreprise. En effet, la pertinence de la ressource résidera pour une bonne part dans sa capacité à offrir des pistes pour la prise en charge de sujets nouveaux, pour lesquels se pose fréquemment la question de leur catégorisation et de leur classement. La seconde raison tient au fait que cet outil restitue des usages langagiers à une population qui se trouve moins dans la position de les ignorer, que dans celle de ne plus les apercevoir, du fait du caractère d'évidence qu'ils ont acquis dans l'exercice quotidien des activités. On fait l'hypothèse, que dans une partie des cas au moins, les agents seront en mesure d'extrapoler, en puisant dans leur compétence passive, qui leur permet ordinairement de comprendre certains discours sans pour autant être à même de les produire.

Enfin, un référentiel terminologique adaptable au contexte privilégiera les informations relatives à la catégorisation et au classement des objets. En effet, il s'agit d'opérations cognitives fondamentales² qui sont à même d'éclairer les problématiques d'organisation de l'information. Mais on a vu plus haut que l'on devra rechercher les manifestations linguistiques de ces opérations, ainsi que leurs variations, en fonction de situations de communication différentes. Par conséquent, on privilégiera les traces écrites laissées par les discours produits dans le cadre de

¹ Norme NF ISO 704 : 2000, *Travail terminologique – principes et méthodes*.

² « *Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech. Every time we see something as a **kind** of thing, for example, a tree, we are categorizing. Whenever we reason about **kinds** of things – chairs, nations, illnesses, emotions, any kind of thing at all – we are employing categories.* » G. Lakoff (1987 : 5-6). Pour un éclairage pluridisciplinaire relatif à cette question, on renvoie aux ouvrages dirigés par D. Dubois (1991 et 1997).

situations caractérisées. Les informations collectées en vue de la constitution d'un référentiel terminologique adaptable au contexte devront donc reposer sur des formes attestées des manières de classer et de catégoriser, telles qu'elles se rencontrent dans les discours produits dans et par l'entreprise.

B. Prise en compte des situations de communication : une rupture en trois schémas

1. La transparence de la langue et le signal

On a vu plus haut que le fonctionnement d'un intranet s'insère dans une série d'activités d'information plus ou moins prises en charges par les agents, dans le cadre d'une évolution de l'organisation du travail, où la part des activités langagières va en augmentant. Ces activités (création, description, classement, catégorisation, publication et recherche de contenus sur des sites intranets), permettent d'assimiler l'intranet à un système d'organisation de l'information, au moins en ce qui concerne ses objectifs. Cependant, l'un des principaux moyens utilisés pour y parvenir, le recours à un usage normalisé de la langue, n'a généralement pas été mis en pratique.

Dans la mesure où les activités d'information liées au fonctionnement d'un intranet sont assimilables à des situations de communication, elles sont également susceptibles d'être orientées de manière à assurer une meilleure prise en compte des destinataires. On propose que cette orientation s'appuie sur l'utilisation d'un référentiel terminologique chargé de restituer les usages langagiers observés dans des situations de communication aux propriétés comparables. Le caractère adaptable des référentiels terminologiques réside ainsi dans leur capacité à construire une sorte d'image langagière de chacun des publics d'un site. Cette image doit être capable de varier en fonction de la nature de la communication et de son objet, mais aussi en tenant compte des caractéristiques des différents publics sous le rapport de leurs profils cognitifs (néophytes ou experts), de leur appartenance à un secteur d'activité ou à une entité du groupe bancaire. Dès lors que les ajustements de la communication ne semblent pas pouvoir reposer sur l'usage d'une langue normalisée – c'est-à-dire qui doit être la même pour les interlocuteurs et qui doit

permettre d'établir une relation biunivoque entre les mots et les choses¹ –, il devient nécessaire d'interroger le modèle de la communication qui peut rendre compte de ces échanges. En effet, dans la mesure où l'on compte s'appuyer sur l'observation des variations linguistiques induites par les situations de communication, il devient nécessaire de disposer d'un modèle qui permette de modéliser leurs caractéristiques.

Un tel modèle devra rendre compte non seulement de l'asymétrie caractérisant les échanges verbaux, mais aussi des conditions réelles dans lesquelles les activités d'information prennent place. En effet, l'objectivation des caractéristiques situationnelles propres à une forme de communication semble être un préalable nécessaire, pour que les producteurs et les diffuseurs d'information puissent rompre avec le sentiment d'évidence qui caractérise l'usage ordinaire de la langue, généralement non réflexif. Car ce sentiment fait obstacle à la prise en charge des situations de communication médiatisées, comme celles que l'on rencontre sur les sites intranets d'une grande entreprise. On cherchera donc un modèle de la communication, qui puisse rendre compte des échanges situés dont les propriétés devront guider l'élaboration de référentiels terminologiques adaptables au contexte. En suivant les principaux avatars du schéma de la communication de C.E Shannon, on présentera les propriétés essentielles qu'un tel modèle doit posséder lorsqu'il a vocation à accueillir les caractéristiques des situations de communication.

Le modèle le plus répandu de la communication langagière repose sur deux piliers : la croyance en une certaine forme de transparence de la langue, et l'oubli du caractère situé des échanges langagiers. Le premier point, qui s'enracine dans une longue tradition, a été caractérisé par T. de Mauro, comme relevant de « l'aristotélisme linguistique » : « *Rappelons quel est le centre de la conception aristotélienne. Le langage est considéré comme une écriture de l'âme. Il consiste à transcrire des séquences de données psychiques qui sont fidèlement transmises à l'extérieur par la transcription, sans être altérées*². » Si cette conception du langage comme *fidèle messenger* lui confère une importance certaine, elle l'enferme

¹ On sait que la recherche de la langue parfaite, qui anime une partie de la tradition terminologique représentée par E. Wüster, est une préoccupation ancienne dont l'histoire a été retracée par U. Eco (*La recherche de la langue parfaite dans la culture européenne*, Paris, Seuil, 1994, 436 p.).

² T. De Mauro (1969 : 43).

également dans un rôle purement instrumental : « [les formes linguistiques] *méritent l'attention parce qu'elles sont les portes par lesquelles pénétrer dans la structure de l'esprit et de la réalité, mais ce qui intéresse alors est évidemment l'esprit et la réalité et non la simple porte d'entrée*¹. » Malgré les différentes crises qui l'ont affectée, cette conception a subsisté jusqu'à nos jours. C'est celle que Ferdinand de Saussure a décrite en ces termes² : « *Pour certaines personnes la langue, ramenée à son principe essentiel, est une nomenclature, c'est-à-dire une liste de termes correspondant à autant de choses*³. » Cette conception d'une langue transparente s'accommode fort bien d'un schéma de la communication désincarnée, valable quel que soit le contexte. Ces deux facteurs, leur accord avec une conception dominante de la langue, et leur caractère générique, expliquent probablement, dans une large mesure, la postérité du modèle de C. E. Shannon.

Le schéma reproduit ci-dessous avait pour but d'illustrer l'un des apports de son auteur, C. E. Shannon, à un modèle pré-existant, décrit par R.V.L. Hartley, dans un article de 1928 intitulé « *Transmission of Information*⁴ ».

¹ *Ibid.*

² F. de Saussure (1995 : 97) [1916].

³ *Ibid.*

⁴ R.V.L. Hartley (1928).

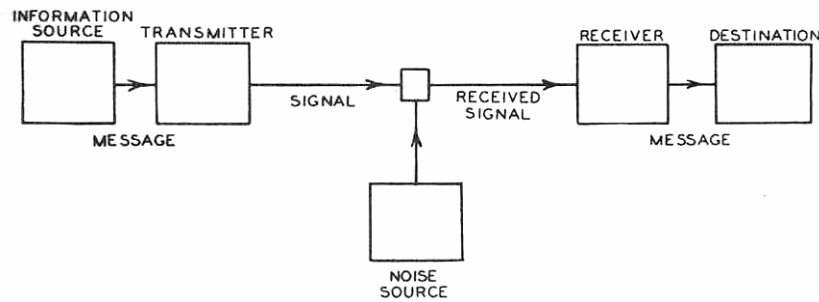


Figure 3.

Schematic diagram of a general communication system (C.E. Shannon)¹

R.V.L. Hartley, comme C.E. Shannon, ont proposé de mesurer une information qui, une fois codée, prend la forme d'un signal susceptible d'être utilisé pour la télégraphie, la téléphonie, la transmission des images ou la télévision. Le premier place sa réflexion sur un plan général, qui lui permet d'écarter des considérations sur les phénomènes qui peuvent altérer la qualité de transmission du signal. C.E. Shannon, en revanche, propose d'étendre la théorie générale de la communication « *de façon à y introduire de nouveaux facteurs, en particulier l'effet du bruit dans le canal et les économies possibles compte tenu de la structure*

¹ C.E. Shannon (1948 : 381) C'est le schéma original qui est reproduit ici.

La présentation des différents éléments composant le schéma sont repris de W. Weaver (1975) :

Information source : « Une source d'information, qui produit un message ou une séquence de messages communiqués au récepteur terminal. Le message peut être de différents types(...). Il peut s'agir, par exemple, d' « une séquence de lettres, comme en télégraphie ou dans les systèmes télétypes ; d'une fonction univoque du temps $f(t)$, comme en radio ou en téléphonie (...) » .

Transmitter : « Un émetteur, qui opère sur le message de façon à produire un signal adéquat pour la transmission dans le canal. En téléphonie, cette opération consiste simplement à changer la pression sonore en courant électrique proportionnel. En télégraphie, nous avons une opération de codage qui produit une séquence de points, de traits et d'espaces dans le canal, correspondant au message.(...) » .

Channel : « Le canal est simplement le milieu utilisé pour transmettre le message de l'émetteur au récepteur. Ce peut être une paire de fils métalliques, un câble coaxial, une bande de fréquence radio, un faisceau lumineux, etc. Durant la transmission ou à l'une des extrémités du canal, le signal peut être modifié par du bruit. Cela est indiqué sur la figure par la source de bruit agissant sur le signal transmis pour produire le signal reçu. »

Receiver : « Le récepteur, ordinairement, effectue l'opération inverse de l'émetteur en reconstruisant le message à partir du signal. »

Destination : « La destination est la personne (ou la chose) à laquelle le message est envoyé. »

statistique du message initial et de la nature de la destination finale de l'information¹. »

Pour ces deux auteurs, la source d'information dispose d'un ensemble fini de symboles dont elle souhaite transmettre un sous-ensemble quelconque d'occurrences à un destinataire. Si toutes les sortes de symboles sont acceptables, la mesure de l'information transmise ne peut se faire qu'au prix d'une réduction du message à ses seuls aspects quantifiables : « *Hence in estimating the capacity of the physical system to transmit information we should ignore the question of interpretation, make each selection perfectly arbitrary, and base our result on selecting any one symbol from that of selecting any other. By this means the psychological factors and their variations are eliminated and it becomes possible to set up a definite quantitative measure of information based on physical considerations alone². »* Depuis 1949, ce modèle, qui opère une réduction de l'information et de la communication à la transmission d'un signal, est toujours utilisé dans la théorie de l'information³. S'il ne permet pas de rendre compte de la communication humaine de manière satisfaisante, il a néanmoins été adapté au domaine de la linguistique par R. Jakobson.

2. Message et langage

On sait que R. Jakobson a emprunté au schéma de C.E. Shannon le canevas qui lui a servi à présenter les six fonctions du langage (les fonctions référentielle, expressive, conative, phatique, métalinguistique et poétique)⁴ : « *Pour donner une*

¹ W. Weaver (1975 : 65) [1949 pour l'édition américaine].

² R.V.L. Hartley (1928: 538).

³ Y.-F. Le Coadic in S. Cacaly (2004 : 233-234).

⁴ Les six fonctions sont décrites ainsi :

- « (...) la visée du référent, l'orientation vers le contexte – bref la fonction dite « dénotative », « cognitive », référentielle (...). »

- « La fonction dite « expressive » ou émotive, centrée sur le destinataire, vise à une expression directe de l'attitude du sujet à l'égard de ce dont il parle. »

- « L'orientation vers le destinataire, la fonction conative, trouve son expression grammaticale la plus pure dans le vocatif et l'impératif (...). »

- « Il y a des messages qui servent essentiellement à établir, prolonger ou interrompre la communication, à vérifier si le circuit fonctionne (« Allo, vous m'entendez ? ») (...). Cette accentuation du contact – la fonction phatique, dans les termes de Malinowski – peut donner lieu à un échange profus de formules ritualisées, voire à des dialogues entiers dont l'unique objet est de prolonger la conversation. »

idée de ces fonctions [du langage], un aperçu sommaire portant sur les facteurs constitutifs de tout procès linguistique, de tout acte de communication verbale, est nécessaire. Le destinataire envoie un message au destinataire. Pour être opérant, le message requiert d'abord un contexte auquel il renvoie (c'est ce qu'on appelle aussi, dans une terminologie quelque peu ambiguë, le « référent »), contexte saisissable par le destinataire, et qui est, soit verbal, soit susceptible d'être verbalisé ; ensuite, le message requiert un code, commun, en tout ou au moins en partie, au destinataire et au destinataire (ou, en d'autres termes, à l'encodeur et au décodeur du message) ; enfin, le message requiert un contact, un canal physique et une connexion psychologique entre le destinataire et le destinataire, contact qui leur permet d'établir et de maintenir la communication¹. » R. Jakobson tenait particulièrement au rapprochement entre la linguistique et les mathématiques à travers la théorie de la communication². Au-delà des emprunts mutuels (la notion de redondance, celle des possibilités préconçues et du rôle de la sélection, etc.), les deux disciplines devaient coopérer pour résoudre un certain nombre de questions importantes liées au langage. Dans « Linguistique et théorie de la communication », il prenait pour modèle inaugural de cette coopération, les travaux du chercheur russe B. Tomachevski qui, dans les années 1920, avait utilisé les chaînes de Markov pour l'étude statistique du vers. Il faut néanmoins convenir que l'emprunt de R. Jakobson relève d'une adaptation drastique du modèle initial de C.E Shannon, afin de le rendre compatible avec une prise en compte de la question du sens.

En effet, le destinataire et le destinataire, non seulement émettent et reçoivent le message, mais ils procèdent eux-mêmes aux opérations de codage et d'encodage. Ils partagent la connaissance du contexte du message, son référent, et ont établi entre eux une connexion psychologique qui assure la continuité de la communication. On est donc assez loin du modèle désincarné de C.E. Shannon : « *Les essais qui ont été*

- « *Chaque fois que le destinataire et/ou le destinataire jugent nécessaire de vérifier s'ils utilisent bien le même code, le discours est centré sur le code : il remplit une fonction métalinguistique (ou de glose).* »

- « *La visée (Einstellung) du message en tant que tel, l'accent mis sur le message pour son propre compte, est ce qui caractérise la fonction poétique du langage.* » R. Jakobson (1963 : 214-218).

¹ « Linguistique et poétique », R. Jakobson (1963 : 213-214).

² « Linguistique et théorie de la communication », R. Jakobson (1963 : 87-99).

*tentés de construire un modèle du langage sans relation aucune au locuteur ou à l'auditeur, et qui hypostasient ainsi un code détaché de la communication effective, risquent de réduire le langage à une fiction scolastique*¹. » Les opérations utilisant le code ne concernent plus une simple translittération de formes à formes, mais permettent l'association du signifié au signifiant d'un signe². On passe ainsi d'un code générique pouvant caractériser de simples signaux³, à un code linguistique reposant sur un mécanisme de double articulation entre morphèmes, plus petits éléments de sens, et phonèmes, plus petits éléments sonores. Ces éléments sont combinables entre eux pour produire des séquences plus importantes, comme les mots. D'une manière générale, ce sont les messages individuels qui sont pris en compte, en tant que réalisations de combinaisons du code, et non plus seulement les possibilités de sélection d'un message. Enfin, contrairement au modèle initial, l'encodage et le décodage ne sont pas des opérations symétriques : « *En gros, le processus d'encodage va du sens au son, et du niveau lexico-grammatical au niveau phonologique, tandis que le processus de décodage présente la direction inverse – du son au sens, et des éléments aux symboles*⁴. » Par ailleurs, le modèle transformé par R. Jakobson possède un *feedback*, qui dépend de cette asymétrie entre parole et écoute. À l'issue de cette comparaison, il ne reste du modèle de C.E. Shannon qu'un canevas, qu'accompagnent les vœux d'une coopération fructueuse entre les deux disciplines pour la description du code et de ses sous-codes, au-delà de la seule phonologie.

3. Langage et situation de communication

Dans le schéma de la communication linguistique proposé par R. Jakobson, les interlocuteurs et leur message ont en quelque sorte gagné en étoffe. Mais ce modèle n'offre pas encore de prise aux caractéristiques externes de la situation de communication. C'est à C. Kerbrat-Orecchioni que l'on doit de les avoir introduites dans le schéma initial.

¹ *Ibid.* p.95.

² *Ibid.* p.90.

³ « L'agencement de la communication verbale » R. Jakobson (1973 : 77-103).

⁴ *Ibid.* p.93-94.

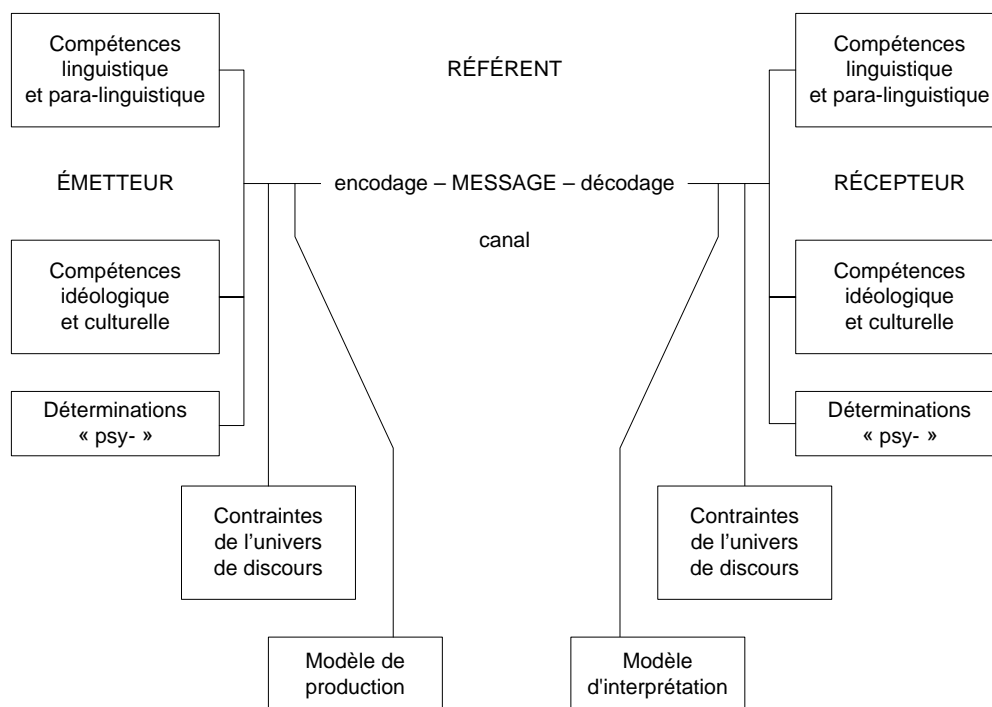


Figure 4.

Reformulation du schéma de la communication (d'après C. Kerbrat-Orecchioni)¹

Même si ce modèle a pour référence une situation simple d'interlocution, un tête-à-tête, il possède des caractéristiques suffisamment génériques pour lui permettre de situer la plupart des échanges verbaux. La langue ne peut plus être transparente, dès lors qu'elle dépend des « compétences linguistique et para-linguistique » des interlocuteurs. L'échange est situé, car caractérisé par les « contraintes de l'univers de discours² », mais aussi par les « compétences idéologiques et culturelles » des participants. Ces derniers sont par ailleurs rétablis dans leurs droits de sujets³ (« déterminations « psy » »). Les compétences de

¹ C. Kerbrat-Orecchioni (1999 : 22). Pour le nom de la figure on a repris le titre de la section dans laquelle le schéma est commenté. On renvoie à l'auteur pour les commentaires complets du schéma (1999 : 16-32).

² « Nous appellerons « univers de discours » l'ensemble : (1) (situation de communication) + (2) (contraintes stylistico-thématiques). » *Ibid.* p. 20 ; La situation de communication correspond aux conditions concrètes de la communication. Quant aux contraintes stylistico-thématiques, elles renvoient aux contraintes de « genre ».

³ L'émetteur et le récepteur sont des sujets déterminés psychologiquement et psychanalytiquement. « leurs compétences culturelles (ou « encyclopédiques », ensemble de savoirs implicites qu'ils possèdent sur le monde) et idéologiques (ensemble des systèmes d'interprétation et d'évaluation de

l'émetteur et du récepteur peuvent dès lors être inégalement réparties, ce qui est plus conforme à l'intuition que l'on peut avoir de la communication humaine. Les compétences de communication¹ intègrent les possibilités d'échec et de correction : « *Il faut en prendre son parti : l'intercommunication (les dialectologues l'ont depuis longtemps montré, et ce qui est vrai des confrontations de dialectes l'est aussi, toutes proportions gardées, des confrontations d'idiolectes) est un phénomène relatif et graduel*². » Ce schéma a été critiqué par son auteur, qui en a présenté certaines limites. De façon générale, C. Kerbrat-Orecchioni indique que les notions présentes dans le schéma devront faire l'objet d'investigations plus ou moins poussées, comme c'est le cas pour les modèles de production et d'interprétation : « *Les modèles de compétence linguistique explicitent l'ensemble des connaissances que les sujets possèdent de leur langue ; mais lorsque ces connaissances sont mobilisées en vue d'un acte énonciatif effectif, les sujets émetteur et récepteur font fonctionner des règles générales qui régissent les processus d'encodage et de décodage, et dont l'ensemble une fois explicité (ce qui est encore loin d'être le cas), constituerait les « modèles de production et d'interprétation »*³. »

L'auteur admet, par ailleurs, qu'un tel schéma représente mal l'existence d'émetteurs complexes contenant en fait plusieurs rôles, comme lorsqu'une agence de publicité est sollicitée par un émetteur qui peut lui-même renvoyer à un collectif, etc. De même, le récepteur peut-il renvoyer à une réalité plus complexe qu'un simple interlocuteur. Ce schéma permet cependant de préciser un autre point essentiel. C'est que les données situationnelles (échange écrit ou oral, canal de transmission, organisation de l'espace communicationnel) « *ne sont pertinentes que sous la forme d'« images », de représentations que les sujets énonciateurs s'en construisent, et qu'il faut en particulier admettre dans leurs compétences culturelles les images (I) que l'émetteur (A) et le récepteur (B) se font d'eux-mêmes et de leur*

l'univers référentiel) qui entretiennent avec la compétence linguistique des relations aussi étroites qu'obscurcs, et dont la spécificité vient encore accentuer les divergences idiolectales. » (Ibid. p. 20-21).

¹ D. H. Hymes (1991).

² C. Kerbrat-Orecchioni (1999 : 18).

³ Ibid. (1999 : 21).

partenaire discursif, c'est-à-dire les quatre éléments que Michel Pêcheux (1969) symbolise comme suit :

I_A (A) (Image de A pour A) : « qui suis-je pour lui parler ainsi ? »

I_A (B) (Image de B pour A) : « qui est-il pour que je lui parle ainsi ? »

I_B (B) : « qui suis-je pour qu'il me parle ainsi ? »

I_B (A) : « qui est-il pour qu'il me parle ainsi ? »¹. »

Ces images et le référent du discours sont convertis en contenu du message. Autrement dit, toute situation de communication laisse des traces dans les discours, qu'il devient possible d'analyser et d'interpréter. C'est cette approche qui est décrite par J.-B. Grize à travers la notion de schématisation.

Le modèle de la communication proposé par J.-B. Grize prend en compte les différentes critiques adressées au précédent modèle et lui substitue une approche permettant d'articuler les ingrédients de la situation de communication avec les caractéristiques des discours.

L'utilisation de cette notion dans le cadre de la constitution d'un référentiel terminologique sera approfondie plus loin. On se contentera ici d'en indiquer la provenance et l'arrière plan théorique, qui conditionnent l'approche que l'on a retenue. Selon J.-B. Grize « *Cette notion est élaborée parallèlement à la notion de 'schéma' de Gonthier, utilisée pour rendre compte de l'argumentation scientifique : les deux mettent en évidence une description sommaire qui vise l'essentiel, défini selon les buts poursuivis ; une description qui pourrait toujours être complétée par de nouvelles déterminations ; et qui possède une signification extérieure, grâce à laquelle elle s'inscrit dans le contexte des actions pratiques².* » La notion correspond donc à une construction discursive, produite dans et par les discours, dont elle intègre les éléments présents dans la situation de communication : « *Une schématisation a pour rôle de faire voir quelque chose à quelqu'un, plus précisément, c'est une représentation discursive orientée vers un destinataire de ce que son auteur conçoit ou imagine d'une certaine réalité³.* » Dans cette définition, on retrouve en filigrane la version « actualisée » du schéma de la communication,

¹ *Ibid.* p. 23. Michel Pêcheux : *Analyse automatique du discours*, Paris, Dunod, 1969, 142 p.

² J.-B. Grize (1974 : 189).

³ J.-B. Grize (1996 : 50).

avec les interlocuteurs et leurs compétences, l'univers de discours, et ce dont il est question – le référent. Mais cette fois, la perspective adoptée est celle du résultat de l'échange, le discours. On donne ci-dessous le schéma de la communication proposé par J.-B. Grize, dans lequel la schématisation occupe une position centrale :

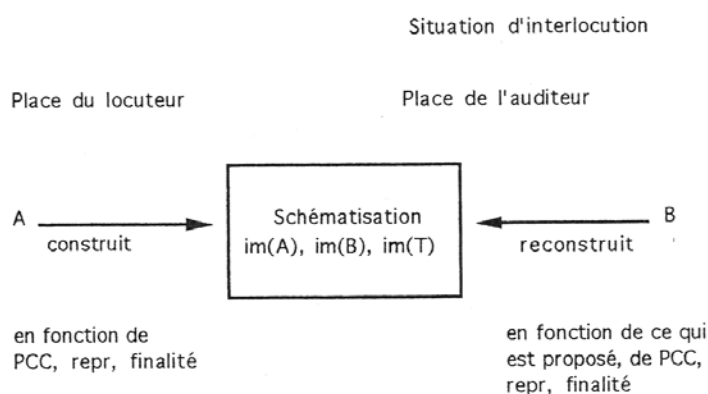


Figure 5.
Schéma de la communication (selon J.-B. Grize¹)

Comme l'illustre la figure ci-dessus, la schématisation résulte des opérations de choix successifs que l'énonciateur et son interlocuteur opèrent relativement à un objet extérieur. Ces constructions produisent des images du référent portées par le discours et son interprétation. Le référent externe n'est alors présent que sous la forme d'un objet de discours. Celui-ci est accessible par l'intermédiaire de notions aux contours flous, ce qui assure à cette approche une certaine robustesse pour l'exploration de nombreux genres de discours. Pour l'étude de ces genres et des schématisations qu'ils véhiculent, on conviendra de tenir compte d'un certain

¹ Schéma de communication proposé par J.-B. Grize (Grize, 1996) et J.-M. Adam (Adam, 1999) : « Une schématisation comporte au moins six types d'images de base qui sont proposées par le discours et sont autant de sortes de versions du monde :

- des images de la situation d'interaction sociodiscursive en cours ;
- des images de l'objet de discours (que l'on appellera aussi bien thème que référent) ;
- des images de A (schématisateur) ;
- des images de B (co-schématisateur).

A ces quatre cas répertoriés par J.-B. Grize, il faut certainement ajouter encore des images de la langue de l'autre ou de celle que l'autre attend que l'on produise. Cette question fondamentale qui traverse les études sociolinguistiques et la réflexion de Pierre Bourdieu sur le 'capital linguistique' des sujets s'étend également jusqu'aux images du support et/ou du canal de transmission de la schématisation. » (Adam, 1999, p. 107). Nous avons retiré de la citation les abréviations utilisées dans un graphique que nous ne reproduisons pas ici.

nombre de postulats adoptés par J.-B. Grize, et qui sous-tendent son modèle du fonctionnement de la communication. Il s'agit tout d'abord du postulat du dialogisme bakhtinien, selon lequel toute production discursive est traversée par les productions antérieures qui la façonnent tout en s'incorporant à elle. La situation d'interlocution fournit un deuxième postulat. Comme on l'a vu plus haut, elle comprend une partie externe, la situation concrète, l'époque à laquelle elle se passe, etc., et une partie interne qui correspond à sa conversion sous la forme du contenu de l'échange. Elle relève d'une pragmatique, telle que rappelée dans l'introduction. Un troisième postulat concerne l'existence de représentations. Ce sont celles que C. Kerbrat-Orecchioni évoque en citant M. Pêcheux. Elles sont complétées par l'image que les interlocuteurs ont de ce dont il s'agit, « im (T) » dans le schéma, les images de la situation d'interlocution, ainsi que celles qui portent sur les relations qui s'établissent entre images, comme la représentation que se fait A de sa relation avec B, notée $\text{repr}_A (A - B)$, sur l'image de la langue de l'autre et sur celle du canal de transmission. L'existence de préconstruits culturels (PCC), quatrième postulat, renvoie aux compétences idéologiques et culturelles du schéma de C. Kerbrat-Orecchioni. Le cinquième postulat concerne le résultat des schématisations, c'est-à-dire la construction d'objets de discours : « *Le discours est création de sens et construit pour cela des objets de pensée à partir de la signification des termes dont il se sert. De tels objets relèvent à la fois des objets des signes et des référents auxquels ils renvoient, ils constituent les **référentiels** des discours*¹. »

On a indiqué plus haut que l'adoption d'un tel modèle de la communication constitue un préalable pour rompre avec la conception la plus largement partagée d'une langue-nomenclature et de situations de communications interchangeable sans grande incidence. Un tel modèle permet de mettre en évidence, par exemple, qu'un document, produit pour un public donné et que l'on expose ensuite à de nouveaux publics en le mettant en ligne sur l'intranet d'une grande entreprise, doit bénéficier d'une prise en charge particulière. Les éléments de contexte auxquels son contenu fait allusion doivent être explicités pour le nouveau public, faute de quoi ce dernier ne sera peut-être même pas en mesure d'y accéder avant d'en évaluer la

¹ J.-B. Grize (1996: 67).

teneur opérationnelle. De même, la notion de compétence linguistique propre à chaque interlocuteur permet-elle de problématiser les inégalités pouvant apparaître dans l'usage du code. Dans la mesure où un échange médiatisé par des pages HTML bloque les ajustements métalinguistiques naturels qui apparaissent dans une conversation courante, un tel modèle met en évidence la nécessité d'accompagner de gloses la publication et l'organisation de l'information sur les intranets. Par ailleurs, un tel modèle ne se contente pas de problématiser les difficultés propres à une communication médiatae.

En effet, il suggère également des pistes susceptibles de pallier les lacunes qui ont été présentées plus haut, en guidant les tâches de publication et d'organisation de l'information. Dans la mesure où la correspondance entre le contexte extralinguistique et la langue est achevée dans et par les discours – la signification en langue ne correspondant qu'à une virtualité qui a besoin d'être actualisée dans les discours –, les discours situés constituent la source à laquelle doit puiser la construction de référentiels terminologiques adaptables au contexte. L'apport de tels référentiels tient d'abord dans leur capacité à rendre compte des variations susceptibles d'affecter un sociolecte selon les contextes de production des discours. En exposant ces variations, ils contribuent à les rendre visibles pour des agents qui les ignorent ; en informant les agents, ils doivent leur fournir les modalités d'une diffusion et d'une organisation des contenus sur les sites intranets, qui prennent mieux en compte les caractéristiques des visiteurs.

En résumé, la constitution de référentiels terminologiques adaptables au contexte dépend directement de ressources textuelles, sélectionnées à partir de leurs caractéristiques communicationnelles. Afin de guider la collecte et d'offrir une vue d'ensemble sur les ressources disponibles ou exploitées, il est nécessaire d'adopter un cadre général dans lequel chaque ressource pourra être située. Autrement dit, dans la mesure où les discours sont produits dans le cadre de situations de communication précises, il convient de pouvoir apprécier les strates du sociolecte auxquelles ils donnent accès sous la forme d'échantillons.

C. Restitution des variations d'un parler d'entreprise

1. La notion de parler d'entreprise

On a vu que le choix des discours devant servir de sources pour la constitution de référentiels terminologiques adaptables au contexte repose sur la sélection de situations de communication caractéristiques. Cela rend nécessaire l'identification de critères explicites. Par ailleurs, la collecte réalisée pour un référentiel devra pouvoir être rapportée à l'ensemble plus vaste que forme le sociolecte en usage dans l'entreprise. C'est là une condition importante pour guider et coordonner la constitution parallèle de référentiels. Il est donc nécessaire de se doter d'un cadre théorique qui puisse rendre compte des principales caractéristiques du sociolecte propre à une entreprise.

La restitution des usages langagiers proposée par un référentiel terminologique adaptable au contexte ne peut raisonnablement pas prétendre rendre compte de toutes les situations de communication, mêmes réduites au contexte que l'on a retenu. De même, si l'on admet, dans une optique sociolinguistique, que la langue ne constitue pas un ensemble homogène, il ne s'agira pas non plus de chercher à présenter avec les référentiels terminologiques le résultat de toutes les causes de variation affectant l'usage d'un lexique chez des individus différents : *« En effet, chacun est dépendant de caractéristiques variables fixées par l'âge, le sexe, l'instruction, la profession, les revenus, la localisation, etc., toutes choses incluses dans de multiples situations de communication qui dégagent à leur tour différents modèles de styles et de niveaux de langue hiérarchisés. Les variétés d'usage peuvent ainsi être modulées selon les énoncés, étant entendu que plusieurs types peuvent se trouver mêlés dans un discours suivi et que des moyens d'expression connus, mais non utilisés par un locuteur, n'en agissent pas moins sur sa manière de communiquer¹. »* On cherchera plutôt à présenter des variations exemplaires d'un sociolecte, motivées par les caractéristiques propres à telle ou telle situation de communication caractéristique.

¹ B. Quemada (1978 : 1147).

Ce sociolecte a pour contexte général un secteur d'activité, au sein duquel existe déjà un *continuum* entre terminologie scientifique et technique et parlars professionnels ou des métiers¹. Ce secteur est défini par les activités qui lui sont propres et le vocabulaire spécialisé qui leur est associé : « *Un vocabulaire technique se définit à partir d'un ensemble d'activités particulières trouvant leur expression linguistique dans un ensemble de termes correspondants. Ceux-ci sont unis par un ensemble de rapports qui peut recevoir le nom de champ sémantique. Chaque activité technique doit pouvoir se définir sur le plan lexical par un vocabulaire particulier et il semble possible d'étudier ainsi les glissements de termes d'un champ sémantique à un champ sémantique connexe quand les activités créatrices s'interpénètrent. Mais on peut se demander aussi si la fonction de communication entre les hommes se livrant aux mêmes activités spéciales ne confère pas au langage technique un caractère spécifique. Notre hypothèse est que le langage technique ne semble pas comporter une grammaire différente du langage général, mais seulement un vocabulaire spécifique dont le système morphologique s'insère dans le système général de la langue*². »

On retiendra de cette conception que les langues spécialisées se caractérisent surtout par leur lexique, que l'on pourra étendre à une phraséologie. Cependant, cette approche centrée sur un domaine d'activité, qui rend bien compte de l'existence des langues spécialisées, conduit fréquemment à deux limitations que l'on cherchera à éviter. D'une part, elle privilégie les termes spécialisés au détriment des unités qui se trouvent à la périphérie d'un tel ensemble, soit qu'elles appartiennent aussi à un autre domaine, soit qu'elles relèvent de zones du lexique mieux partagées, voire de la langue commune. Or, il semble que les unités de cette sorte, qui assurent la transition d'un secteur à l'autre à l'intérieur d'un même sociolecte, présentent un intérêt certain dans une perspective communicationnelle. D'autre part, cette approche tend également à gommer les différences pouvant apparaître dans des situations de communication typiques ayant pour contexte telle ou telle organisation, au profit de la terminologie d'un domaine type, c'est-à-dire non

¹ *Ibid.* p.1155.

² L. Guilbert (1965 : 9).

situé. On a vu plus haut que l'on propose au contraire de prendre appui sur un contexte précis, dont il convient de ne pas négliger les caractéristiques. Il semble que la notion de parler d'entreprise proposée par D. de Vecchi permette de rendre compte plus finement des réalités linguistiques que l'on cherche à restituer.

D. de Vecchi¹ a dégagé la notion de parler d'entreprise en comparant les éléments de la terminologie d'une entreprise de transport aérien, d'une part, avec les unités relevant de la langue commune, et d'autre part, avec les terminologies en usage dans des entreprises appartenant au même secteur d'activité. Il ressort de la première comparaison que certaines unités terminologiques, attestées dans le contexte étudié, ne sont pas répertoriées dans les dictionnaires courants. C'est là un indice banal du caractère spécialisé d'une partie de la langue qui n'est pas recensée par les dictionnaires consacrés à la langue commune. Cette séparation n'est cependant pas étanche, car on sait également qu'une circulation s'établit entre langue spécialisée et langue commune, les néologismes pouvant être incorporés au trésor commun, alors que ce dernier, par le phénomène de la terminologisation, fournit une partie de leurs termes aux langues spécialisées. Il reste que « (...) *le français de l'automobile n'est en aucune façon une sorte de langue ou de dialecte, et l'usage de termes analogiques comme « technolecte » induit un parallélisme fallacieux. Le français de l'automobile est l'usage du français pour rendre compte de connaissances en matière d'automobile, ce qui est bien différent (...)* »². Ce premier constat atteste donc l'existence de sociolectes au sein d'un ensemble linguistique commun.

La deuxième confrontation a consisté à soumettre la terminologie attestée dans une entreprise à des agents travaillant au sein d'une autre entreprise appartenant au même secteur d'activité. Il ressort de cette expérience que seul un quart des unités sont facilement reconnues, alors que 38% d'entre elles sont plus ou moins bien comprises et que les 38% restantes sont inconnues³. Ces observations mettent ainsi en évidence un découpage de second niveau au sein de la terminologie d'un secteur d'activité. Le sociolecte d'une entreprise se compose d'un *continuum*

¹ D. De Vecchi (1999).

² P. Lerat (1995 : 18).

³ D. de Vecchi (1999 : 232).

allant de la langue commune à des unités qui lui sont propres, en passant par la terminologie du domaine d'activité. Si l'on se place à une échelle encore plus fine, en considérant l'organisation interne d'une entreprise, chaque pan d'activité sert également de contexte à l'usage d'unités relevant de la langue commune, de la terminologie d'un domaine d'activité, et du sociolecte propre à cette entreprise¹. La langue parlée à l'intérieur d'une entreprise ne peut donc pas être réduite à la terminologie d'un domaine d'activité, d'une part, parce qu'elle incorpore d'autres strates de la langue, comme la langue commune et, d'autre part, parce qu'elle se compose fréquemment de plusieurs domaines d'activité. L'auteur constate ainsi que la terminologie qui pourrait être partagée entre les entreprises exerçant dans le même secteur d'activité « *ne l'est plus à partir du moment où chaque entreprise conceptualise ses produits, s'organise et donc nomme à sa manière et en fonction de ses besoins*². » C'est pour rendre compte de cette particularité que D. de Vecchi a proposé la notion de parler d'entreprise qui permet de « (...) *désigner l'ensemble des processus linguistiques qui actualisent les répertoires linguistiques des membres d'une communauté, définie en fonction de l'appartenance à une entreprise. Autrement dit, la cristallisation linguistique de tout moyen de communication mis à la disposition d'une entreprise, pour des conceptualisations ayant des origines diverses*³. » Ce parler est caractérisé principalement par son lexique, et en particulier par différentes sortes d'unités nominales⁴ : noms communs simples et complexes, noms propres et codes (internes, internationaux, informatiques, etc.). Sa description relève de la terminologie⁵.

Dans la mesure où elle permet de prendre en compte des phénomènes de variations linguistiques caractéristiques du fonctionnement des organisations, la notion de parler d'entreprise semble à même de servir de cadre pour le développement de référentiels adaptables au contexte des situations de communication. On en donne ci-dessus une illustration à l'aide de l'objet « livret

¹ *Ibid.* (1999 : 368).

² *Ibid.* (1999 : 363).

³ D. de Vecchi (1999 : 316).

⁴ *Ibid.* p. 320.

⁵ *Ibid.* p. 356 et suiv.

jeune » présenté dans quatre contextes linguistiques différents¹. Le tableau ci-dessous présente ces différents contextes de façon schématique, avec les variantes de la dénomination et les catégorisations spécifiques de l'objet.

*Tableau 2.
Variations d'un parler d'entreprise autour de l'objet « livret jeune »*

<p>JURIDIQUE</p> <p>Titre II Les produits d'épargne (...) Chapitre 1 Produits d'épargne générale à régime fiscal spécifique (...) Section 3 Le livret jeune</p> <p>Dénominations : livret jeune / livrets jeunes</p>	<p>DOCUMENTAIRE</p> <p>ressources bancaires (...) épargne (...) livret d'épargne livret d'épargne jeune</p> <p>Dénomination : livret d'épargne jeune</p>
<p>COMPTABLE</p> <p>Collecte Collecte épargne (...) Épargne livret (...) Livrets jeunes</p> <p>Variations de la dénomination : Livrets jeunes / Livret Jeune Mozaïc / LMZ</p>	<p>COMMUNICATION INSTITUTIONNELLE</p> <p>L'objet dénommé « livret jeune » est :</p> <ul style="list-style-type: none"> - présenté comme faisant partie de la gamme de produits « Mozaïc » - catégorisé comme étant une sorte de livret - est rattaché à la clientèle des jeunes <p>Variations de la dénomination : livrets Jeunes / livrets jeunes / livret Jeunes / Livrets Jeunes / Livret Jeune Mozaïc / Livret Jeunes</p>

Lecture du tableau 2 : les quatre cartouches matérialisent chacun un domaine d'activité (juridique, documentaire, comptable et communication institutionnelle). À l'intérieur de chaque cartouche on présente le contexte d'apparition du terme « livret jeune » dans un document appartenant à chaque secteur d'activité. Le secteur juridique est représenté par le *Code monétaire et financier*, le secteur documentaire l'est à l'aide du *Thésaurus Resagri*, le secteur comptable est représenté par une norme comptable, enfin, le secteur de la communication institutionnelle est représentée par un corpus de rapports d'activité. Les variantes rencontrées dans chaque contexte sont mentionnées en bas de chaque cartouche.

Lorsque l'on compare les dénominations recensées dans les quatre secteurs mentionnés ci-dessus, deux faits ressortent particulièrement. Tout d'abord, la dénomination terminologique « livret jeune », observée dans le *Code monétaire et financier*, ainsi que dans le *Décret d'application*, ne se retrouve pas dans les autres

¹ Le contexte juridique est représenté par un petit corpus de 2000 occurrences composé des articles L 221-24 à L 221-26 du Code monétaire et financier et du Décret n° 96-367 du 2 mai 1996 relatif à la mise en place du « livret jeune ». Le secteur documentaire est illustré à l'aide du thésaurus RESAGRI (édition 1997), qui se compose de descripteurs utilisés pour réaliser une indexation manuelle des informations. Un échantillon de référentiels internes a été prélevé pour le domaine comptable. Dans tous les cas, il s'agit de corpus ad hoc n'ayant bénéficié que d'un traitement léger, voire nul pour l'extrait de thésaurus. Enfin, pour le secteur de la communication institutionnelle, on a utilisé un corpus de plus de 200000 occurrences qui sera présenté dans les chapitres suivants.

secteurs, ou tout au moins avec sa graphie exacte. On note que cette répartition reflète la séparation entre extérieur et intérieur par rapport l'entreprise, alors que l'on peut dire que toutes les sources recensées ont un rapport assez étroit avec l'activité bancaire. D'autre part, il faut bien constater qu'entre secteurs d'une même entreprise la situation n'est guère meilleure, puisque l'échantillon représentant l'un d'entre eux ne possède aucune forme commune avec les autres. On va voir que cela peut s'expliquer par la nature des utilisations qui sont faites de l'objet « livret jeune ».

La dénomination « livret d'épargne jeune », présente dans le thésaurus du secteur documentaire, est ignorée, entre autres, par les textes juridiques, qui représentent pourtant une source on ne peut plus légitime. Ce descripteur correspond en fait à une unité artificielle créée pour la construction d'un thésaurus et destinée à l'indexation documentaire. Le thésaurus est, en effet, une construction hiérarchisée de termes qui doivent représenter non seulement le domaine, mais aussi son organisation. Celle-ci s'incarne en particulier dans le mécanisme « terme de tête – expansion ». Le livret jeune étant un livret d'épargne, il est admis qu'un terme artificiel puisse être créé dans le seul but de rendre explicite cette information.

C'est un peu le même mécanisme qui caractérise le faux sigle « LMZ », pour « livret Mozaïc ». Cette dernière dénomination n'est pas non plus répertoriée. En revanche, il existe d'autres sigles à trois lettres concernant les livrets d'épargne : LEE, pour le livret d'épargne entreprise et LEP, pour le livret d'épargne populaire. Un sigle conforme aux usages « locaux » a donc été créé de toutes pièces : il comporte trois lettres, il est facile à prononcer et évocateur, qualités absentes de « LJ », qui est par ailleurs le sigle correspondant à un développé caractérisant une forme de lettre circulaire en usage dans le groupe bancaire.

Les flottements graphiques les plus remarquables au sujet de la dénomination « livret jeune » se rencontrent cependant dans le corpus de communication institutionnelle. Ils revêtent deux aspects : l'incertitude sur l'attribution de la majuscule à « livret » et à « jeunes », et l'utilisation presque systématique du pluriel pour « jeunes », même si « livret » est au singulier. Le livret jeune s'adresse à une clientèle de jeunes. Aussi, la dénomination « livret jeune » est-elle nécessairement amenée à côtoyer dans les textes celle de « jeunes ». Compte tenu de ces éléments,

on interprète la majuscule comme la marque du souci de délimiter la dénomination du produit par rapport à son cotexte.

Reste à rendre compte du pluriel qui est appliqué aussi quand « livret » est au singulier. L'explication semble liée à ce que l'on vient de dire. En effet, pour le législateur, le livret jeune se comprend comme un livret unique dont peut être titulaire « *une personne physique âgée de douze à vingt-cinq ans et résidant en France à titre habituel* », ce qui correspond à une définition spécialisée du substantif « jeune ». En revanche, dans le contexte bancaire, « jeune » au singulier est presque ignoré, car ce qui importe, c'est le segment de clientèle, le marché des jeunes. Dans ce contexte, il se comprend donc comme le livret d'épargne destiné aux jeunes.

Par ailleurs, les catégorisations de l'objet « livret jeune » éclairent de façon complémentaire ces phénomènes de variation : « livret », « produit d'épargne générale à régime fiscal spécifique », « livret d'épargne », « épargne livret », constituent autant de manières de catégoriser, propres à la communication institutionnelle et aux secteurs d'activité juridique, documentaire et comptable. Cet exemple a permis d'illustrer l'usage d'un sociolecte particulier et les variations fines dans les manières de dénommer et de catégoriser que différentes formes d'activité peuvent générer en fonction de leurs besoins. Il a également permis de montrer que la constitution de référentiels terminologiques adaptables au contexte doit pouvoir s'appuyer sur des ressources linguistiques caractéristiques, d'une part, des secteurs d'activité d'une entreprise, et d'autre part, des situations de communication qui mobilisent le parler d'entreprise dans de tels cadres.

2. Ressources textuelles des référentiels terminologiques

La démarche retenue s'inscrit dans le cadre d'une linguistique de discours qui trouve son origine dans les travaux de Z. Harris¹ et que Rostislav Kocourek² a résumée au sujet de la méthode terminologique : « *On peut formuler une **hypothèse** qu'il est linguistiquement légitime de choisir un ensemble de textes, délimités d'une manière externe, dans le but de déterminer les ressources linguistiques sous-*

¹ Z. Harris ([1951], 1969).

² R. Kocourek, (1991 : 24).

tendues, de dégager les propriétés, les principes, les tendances de ce sous-ensemble de textes, et, en ce faisant, d'enrichir, de préciser, d'approfondir la connaissance et la compréhension de la langue entière. » Cette démarche s'inscrit donc dans un courant terminologique à propos duquel on a parlé de « virage méthodologique », et dont les principes ont été énoncés ainsi :

« Proposition 1 : objet empirique d'une linguistique textuelle, le texte est le point de départ de la description lexicale à construire. On va du texte vers le terme. Les bases théoriques de la terminologie doivent être ancrées dans une linguistique textuelle.

Proposition 2 : le terme est un construit. Il est le produit d'un travail d'analyse, mené par le linguiste terminologue, dont les choix sont guidés par une double contrainte de pertinence:

- Pertinence vis-à-vis du corpus. Il s'agit de retenir et de décrire des structures lexicales qui présentent des caractéristiques à la fois spécifiques et stables. C'est à ce stade qu'intervient la validation par l'expert.

- Pertinence vis-à-vis de l'application. Les unités finalement retenues doivent l'être en fonction de leur utilité dans l'application visée, qui s'exprime en termes d'économie et d'efficacité. La validation est à chercher du côté des utilisateurs de l'application¹. »

Les ressources textuelles sur lesquelles doit pouvoir s'appuyer la construction de référentiels terminologiques dédiés aux activités d'organisation de l'information sur les intranets doivent être organisées en fonction des caractéristiques d'un parler d'entreprise. Les déclinaisons du parler propres aux différents secteurs doivent ainsi être prises en compte au même titre que la « langue de l'entreprise »². Par ailleurs, on cherchera pour ces déclinaisons des situations de communication caractéristiques propices à l'observation de leur fonctionnement. Enfin, ces ressources devront être organisées de manière à rendre compte des évolutions des strates du parler qu'elles incorporent. Par conséquent on privilégiera

¹ M. Slodzian et D. Bourigault (1999 : 31).

² D. de Vecchi opère une distinction entre parler d'entreprise en langue de l'entreprise, cette dernière faisant « l'objet d'une décision des hautes instances de l'entreprise comme langue de communication pour le personnel. » D. de Vecchi (1999 : 318). On considèrera que cette langue est une composante parmi d'autres du parler de l'entreprise.

une organisation des corpus sous la forme de séries ouvertes et homogènes (car les discours rassemblés le seront dans le cadre d'une situation de communication particulière).

On adopte comme définition d'un corpus celle qui est donnée par J. Sinclair et qui a été reprise dans *Les linguistiques de corpus* : « un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage¹. » Le suivi du fonctionnement d'un parler lié à un secteur de l'entreprise et à une situation de production discursive caractéristique doit pouvoir être enrichi au fil de l'eau, de manière à rester un témoin pertinent de la manière dont sont verbalisées les évolutions des activités de l'entreprise. Le corpus sera donc un *monitoring corpus*, ou corpus de suivi, c'est-à-dire qu'il pourra être enrichi en permanence, ou à échéances régulières, par de nouvelles données textuelles. Dans la mesure où les textes rassemblés appartiendront à la même série, le type de corpus qui sera utilisé correspondra en fait à une série textuelle chronologique, caractérisée par des données langagières homogènes ordonnées chronologiquement.

Les textes, rassemblés dans un format électronique homogène, de type texte brut ou HTML, doivent être sélectionnés et organisés en corpus à partir des exigences qui ont été formulées plus haut. Le texte brut sera en revanche documenté de manière à pouvoir distinguer les grandes parties composant le corpus (elles correspondent généralement à des textes complets), ainsi que des unités textuelles plus fines (intertitres, paragraphes). Chaque ressource devra être documentée de manière à ce qu'il soit possible d'identifier la situation de communication dans laquelle elle a été produite, et partant, la mise en jeu du parler d'entreprise qu'elle permet de documenter².

¹ J. Sinclair (1996 : 4) cité par B. Habert, A. Nazarenko et A. Salem (1997 : 144).

² Le *Dublin Core* « ensemble de métadonnées utilisées pour décrire les documents électroniques » Cacaly (2004), constitue de ce point de vue un début de description (titre, auteur, sujet, résumé, éditeur, contributeurs, date, type de document, format, une identification unique de type ISBN ou URI, source, langue, contenus associés, coordonnées spatio-temporelles du contenu, droits), auquel il faudrait ajouter la caractérisation des conditions de production et de la situation de communication concernées. Les activités et les recommandations de la Dublin Core Metadata Initiative se trouvent à cette adresse : <http://dublincore.org/>.

Généralement disponibles au format électronique, les textes ne subiront pas a priori de traitements poussés nécessitant un contrôle minutieux de la part de l'agent désireux de les exploiter. Lors de la phase d'exploitation de ces données textuelles, les logiciels utilisés devront permettre au moins une exploitation du texte brut (format .txt). En outre, ils devront permettre une navigation fluide entre les textes et les unités que l'on aura souhaité extraire. En effet, cette tâche répétitive mais indispensable, car c'est en définitive sur elle que repose la fiabilité des données collectées, doit être dans la mesure du possible simplifiée pour l'opérateur. En ce sens, l'intégration des technologies utilisées doit non seulement être possible au sein d'une suite bureautique classique, mais surtout, ne pas constituer une rupture importante nécessitant un apprentissage poussé.

Comme cela a été indiqué plus haut, la constitution de corpus sous la forme de séries textuelles chronologiques doit permettre de suivre finement l'évolution d'échantillons d'un parler d'entreprise. Celui-ci sera appréhendé à travers les traces qu'il a laissées dans les discours produits dans le cadre de situations de communication sélectionnées pour leur caractère exemplaire. Ce découpage de la cartographie d'un parler d'entreprise rend également nécessaire l'adoption de points de repères, susceptibles de mettre en évidence les variations induites par des situations de productions différentes. En effet, il s'agit de trouver le moyen de comparer rapidement les propriétés d'un parler qui sont mobilisées dans telle ou telle situation, de manière à en tirer les enseignements utiles pour informer une activité d'information. En effet, si la notion de parler d'entreprise permet de cerner avec la finesse requise ce type réalité, il reste que l'exploration de toutes les variations possibles à l'échelle des productions discursives caractéristiques d'un groupe bancaire à une époque donnée relève de la gageure. On a donc recherché un moyen d'aborder ce phénomène de telle sorte que l'on puisse l'illustrer à l'aide d'exemples caractéristiques permettant la restitution de différentes strates d'un parler d'entreprise.

D. Référentiel et référentiels

1. Référentiel et objets de discours

Pour atteindre ce but, on propose de s'appuyer sur une sélection d'objets susceptibles d'apparaître dans de nombreuses situations de communication ayant pour cadre un contexte précis, qui correspond dans cette étude aux intranets d'un groupe bancaire. On a vu plus haut que le référent est présent dans les discours sous la forme d'objets de discours, dont les propriétés sont construites dans et par les discours. On fait l'hypothèse que dans le cadre des discours produits à l'aide d'un parler d'entreprise, certains objets sont plus spécifiques que d'autres, en ce sens qu'ils touchent à l'ontologie de l'entreprise, c'est-à-dire à ce qui fait que telle entreprise est différente de toutes les autres. Autrement dit, on va tenter d'articuler de cette manière les éléments d'un référentiel d'objets¹ spécifique avec les unités composant un référentiel terminologique adaptable au contexte.

Afin d'opérer une sélection parmi les éléments dont se compose une situation de communication, on a choisi de prendre pour point de départ certains référents présents dans les discours sélectionnés sous la forme d'objets de discours, et ayant de bonnes chances d'appartenir au référentiel associé à l'exercice des activités propres à une organisation. Cette composante de la situation de communication est de nature à établir un lien entre les interactions temporaires des agents et les réalités au sujet desquelles leurs discours particuliers sont produits dans le cadre d'une entreprise particulière. Cependant, la relative stabilité pouvant provenir de l'ensemble des choses auxquelles on peut faire référence dans un contexte déterminé est contrebalancée par l'étendue de cet ensemble. En effet, dès lors qu'il est sorti d'un cadre communicationnel spécifique dans lequel il est mobilisé, un référentiel correspond à un ensemble potentiellement infini. On a donc cherché le moyen de délimiter les portions de référentiel utiles relativement aux problématiques d'exploration d'un parler d'entreprise.

¹ C'est à L. Depecker que l'on doit le signalement de la notion de référentiel. Sur ce sujet, on peut se reporter à L. Depecker (2002 : 82-85).

Chaque agent est détenteur à des degrés variables d'un référentiel d'entreprise. Pour le caractériser, on part de la définition que Josette Rey-Debove¹ donne du référentiel : « *Ensemble de tout ce dont un locuteur peut parler dans une langue donnée (objets réels ou imaginaires, concrets ou abstraits, appelés référents).* » On restreint ensuite le champ ainsi ouvert, en ajoutant que ces objets doivent entretenir un rapport direct ou indirect avec l'accomplissement des tâches pour lesquelles le locuteur est rémunéré dans le cadre de son activité. Ce référentiel possède différents niveaux, qui comprennent les activités liées au métier, les domaines qui lui sont liés, mais aussi les orientations stratégiques de l'entreprise. Par ailleurs, on a vu que les sujets relatifs à chaque strate sont verbalisés à l'aide des composantes d'un parler d'entreprise. Celles-ci sont sélectionnées parmi différents registres, allant des échanges informels aux textes de communication institutionnelle (journaux internes, rapports annuels, communiqués de presse, etc.). Par conséquent, chaque situation de communication agit comme un filtre susceptible de réduire l'étendue du référentiel évoqué et les moyens linguistiques utilisés pour le faire.

Dans ce cadre, le référentiel fait office de médiateur entre les agents communiquant par l'intermédiaire des sites intranets. Son rôle social est explicité par F. Gonseth, auquel J. Rey-Debove a emprunté cette notion : « (...) *le mot référentiel en vient à signifier l'ensemble (explicite ou implicite) des préalables faute desquels telle ou telle activité systématique ne pourrait pas avoir lieu. Le référentiel peut alors s'offrir comme un cadre où cette activité prendra place. C'est ainsi que tout ce qui fait pour nous, en telle ou telle circonstance, la « réalité » d'un terrain de jeu peut servir de référentiel commun au joueur et au spectateur*². » Transposé au contexte de l'intranet, et plus particulièrement aux activités d'information touchant à la publication, l'organisation ou la recherche d'informations, le référentiel peut être schématiquement représenté sous la forme de trois ensembles.

¹ J. Rey-Debove (1998: 289).

² F. Gonseth (1975 : 22).

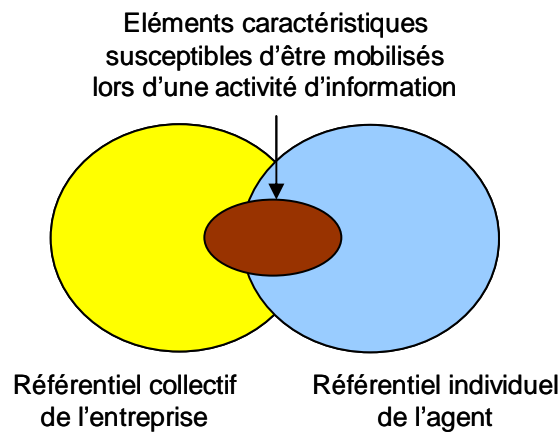


Figure 6.

Référentiel et éléments caractéristiques susceptibles d'être mobilisés lors d'une activité d'information

Le référentiel collectif est propre à d'une entreprise particulière à un moment donné. Il correspond à un ensemble de sous-référentiels, comme la liste des membres de l'entreprise, la nomenclature des produits, la liste des implantations, la liste des clients, les normes comptables, les lois et décrets qui s'appliquent à l'entreprise, les règles de procédures applicables à tel ou tel processus, la liste des logiciels et applications spécifiques maintenus, la liste des filiales et entités rattachées, la liste des sites intranets, etc. Il s'agit d'un ensemble ouvert, en permanente évolution, composé d'un *continuum* allant du particulier au concept général, du spécifique à l'universel, et dont la connaissance exhaustive est théoriquement possible, mais inaccessible dans la pratique.

Le référentiel individuel d'un agent appartenant à une entreprise est composé au moins de deux sous-ensembles, celui qui est sans rapport avec le référentiel de l'entreprise, et celui qui correspond à un sous-ensemble du référentiel de l'entreprise. On postule que ce sous-ensemble se compose lui aussi de strates, allant de ce que le sujet perçoit comme indispensable pour mener à bien les tâches essentielles liées à son métier, aux connaissances générales sur l'entreprise et son contexte économique.

Enfin, le troisième ensemble est formé par les éléments de référentiel sur lesquels le diffuseur peut être conduit à intervenir de manière à assurer une prise en charge adaptée aux différents publics du site. Parce qu'ils sont susceptibles d'être partagés par les agents d'une même entreprise à propos de tel ou tel sujet, la prise en

compte des différentes manières de verbaliser ces éléments du référentiel permettra d'orienter les recherches d'information. Le diffuseur de l'information devra alors chercher à faire correspondre les informations apportées par les référentiels terminologiques avec les caractéristiques des différents publics d'un site, comme la provenance des intranutes, leurs profils, qui varient selon qu'il s'agit de néophytes ou d'experts, la nature de l'information mise en ligne, etc. Cette prise en charge de la diffusion étendue des informations sur un intranet doit pouvoir s'appuyer sur des indications relatives aux différentes manières de verbaliser les éléments caractéristiques d'un référentiel, sans qu'il soit nécessaire d'avoir une vue d'ensemble sur les usages langagiers des différents publics d'un site. On compte en effet que les référentiels terminologiques jouent plutôt le rôle de déclencheurs propres à activer les compétences linguistiques des agents, que d'inventaires exhaustifs d'un parler en situation.

En effet, le cadre, somme toute délimité, des échanges pouvant avoir lieu sur un intranet, ainsi que la nécessité d'être en prise avec l'activité, ont suggéré de remplacer un traitement exhaustif, souvent périmé à peine achevé, par des restitutions exemplaires des fonctionnements discursifs propres à fournir des repères aux activités d'information. Ces indications seront destinées à permettre au producteur et au diffuseur d'information d'extrapoler à partir de la présentation de fonctionnements discursifs typiques enregistrés pour un certain nombre d'objets. Il convient maintenant de préciser la méthode qui sera utilisée pour identifier ces derniers lorsqu'ils sont présents dans les textes du corpus sous la forme d'objets de discours. C'est pour restreindre le champ des explorations à ce qui a paru suffisant, afin de répondre à la nécessité de faire un retour sur les fonctionnements langagiers ordinaires dans une entreprise, que l'on a été conduit à rechercher un point de repère stable dans les discours, qui soit également en prise avec le référentiel de l'entreprise qui sert de contexte aux activités d'information qui doivent être guidées.

2. Hypothèse 1 : un référentiel peut être saisi au moyen de ses noms propres

Sans pour autant postuler que les noms propres¹ constituent nécessairement des objets de discours susceptibles de mettre sur la voie d'observations de nature à restituer les usages langagiers propres à une organisation, ces unités ont paru posséder des vertus utiles liées à leur pragmatique. Même si l'analyse pragmatique des noms propres n'occupe pas une place centrale dans les travaux de G. Kleiber, ce dernier lui a néanmoins prêté attention dans sa thèse consacrée aux problèmes de référence², et l'on rappelle ici les principaux points de son argumentation. G. Kleiber commence par remarquer que si tous les particuliers peuvent théoriquement être nommés, tous ne le sont pas pour autant. Autrement dit, il existerait des règles propres à l'utilisation des noms propres, qui décideraient de l'opportunité de leur emploi dans une langue et une culture données.

En suivant P. F. Strawson, G. Kleiber distingue trois conditions qui doivent être remplies par un particulier afin qu'il puisse bénéficier d'un nom propre. Il doit être l'objet de fréquentes prédications au sein d'un groupe de locuteurs, ce qui exclut les particuliers éphémères, pour lesquels l'attribution d'un nom spécifique ne serait pas rentable en termes de communication. Le groupe doit aussi ressentir la nécessité d'identifier de cette façon le même particulier. Enfin, le groupe de locuteurs ne doit pas disposer d'une description définie facile à manipuler, qui pourrait suffire à désigner le particulier en question. Cette troisième condition n'est cependant pas exclusive, et G. Kleiber reconnaît que le fait d'éprouver le besoin de nommer, par exemple, un objet familier comme une maison ou une voiture, s'en affranchit

¹ La question des propriétés des noms propres et de leur fonctionnement discursif sera abordée dans le chapitre 8. On peut se contenter à ce stade de la définition proposée par la *Grammaire générale et raisonnée* de A. Arnauld et C. Lancelot, qui convient à l'usage référentiel des noms propres dont il sera principalement question dans ce travail : « *Ils [les hommes] ont appelé **noms propres** ceux qui conviennent aux idées singulières, comme le nom de **Socrate**, qui convient à un certain philosophe appelé Socrate ; le nom de **Paris**, qui convient à la ville de Paris. (...) Ce n'est pas qu'il arrive souvent que le mot propre ne convienne à plusieurs, comme **Pierre**, **Jean**, etc., mais ce n'est que par accident, parce que plusieurs ont pris le même nom ; et alors il faut y ajouter d'autres noms qui le déterminent, et qui le font rentrer dans la qualité de nom propre, comme le nom de **Louis**, qui convient à plusieurs, est propre au roi qui règne aujourd'hui, en disant **Louis quatorzième**. Souvent même il n'est pas nécessaire de rien ajouter, parce que les circonstances du discours font assez voir de qui l'on parle. » A. Arnauld et C. Lancelot (1997 [1660] : 28-29).*

² G. Kleiber (1981). Sur ces points, G. Kleiber (1981 : 317-319) suit la réflexion de P. F. Strawson (1974 : 42-46).

aisément. Il reste que ce besoin s'insérera généralement dans une norme sociale, qui non seulement pèsera sur le choix du particulier nommable, mais aussi sur la manière de le nommer (on reviendra sur ces points dans la troisième partie).

La notion de groupe de locuteurs est surtout mise en avant par P. F. Strawson que l'on suivra dorénavant. Ce point est important, car il permet de faire la jonction avec la notion de référentiel. En effet, en dehors des noms propres culturels, tels que les noms des principaux lieux, comme les continents, les états, les capitales, ou ceux des personnages célèbres, la majorité des noms propres possèdent un champ de circulation généralement plus restreint que celui de la plupart des unités appartenant au lexique d'une langue donnée¹. Leur utilisation en fait même une unité caractéristique d'un sociolecte restreint. C'est ce que suggère la situation imaginée par P. F. Strawson², dans laquelle des interlocuteurs ressentent le besoin de nommer. Même si deux individus qui assistent à un match de football ou de basket n'éprouvent par ailleurs qu'un intérêt limité pour de telles manifestations sportives, ils vont ressentir la nécessité de nommer les joueurs, afin de se faire part de leurs commentaires. Sans connaître le nom véritable des sportifs, et même en ignorant tout de leurs carrières, les deux spectateurs vont créer les noms nécessaires à leurs échanges : « *The two watchers, for the duration of the game and perhaps a little time afterwards, constitute now something like a name-needing circle in relation to the players concerned*³. »

En reformulant les choses de manière à les mettre en accord avec l'approche que l'on propose, les noms propres apparaissent bien comme pouvant donner accès au référentiel d'un groupe déterminé de locuteurs, situé dans l'espace et dans le temps⁴. On constate donc, au moins théoriquement, qu'ils correspondent à un type

¹ Le caractère généralement limité du cadre d'usage d'un nom propre avait déjà été noté par Dumarsais, qui précisait ainsi la définition donnée par la grammaire de Port-Royal citée plus haut : « *Le nom propre ; c'est le nom qui n'est dit que d'un être particulier, du moins dans la sphère où cet être se trouve ; ainsi Louis, Marie, sont des noms propres, qui, dans les lieux où l'on en connoît la destination, ne désignent que telle ou telle personne, et non une sorte ou espèce de personnes.* » C.C. Dumarsais (1797 : 192).

² P. F. Strawson (1974 : 44).

³ Idid.

⁴ On peut même aller plus loin et considérer qu'un groupe de locuteurs peut se définir, indépendamment de la langue, à partir des noms propres qu'il utilise et dont ses membres connaissent au moins une partie.

d'unité à prendre en compte, dès lors que l'on cherche à cerner la manière dont un référentiel peut être verbalisé dans telle ou telle situation de communication. On fait alors l'hypothèse que ces unités possèdent des propriétés linguistiques pertinentes qui en font de bonnes candidates pour l'observation et la restitution des usages d'un parler d'entreprise à partir des traces laissées dans les discours. De ce point de vue, non seulement les noms propres présentent un intérêt en soi pour des référentiels terminologiques, mais ils peuvent aussi peut-être permettre de capter les usages d'un parler, au moyen des relations qu'ils entretiennent avec les autres sortes d'unités présentes dans les discours.

3. Hypothèse 2 : un vocabulaire peut être capté à partir de ses noms propres

En effet, la première hypothèse, relative aux liens unissant noms propres et référentiel, doit être doublée d'une seconde hypothèse, portant cette fois sur les relations établies entre les noms propres et les discours. Cette hypothèse s'appuie sur les propriétés pragmatiques et linguistiques des noms propres, afin de suggérer qu'ils pourraient également constituer de bons candidats pour la collecte d'une partie du vocabulaire des corpus rassemblés en vue de la constitution des référentiels terminologiques. La vérification de cette hypothèse nécessitera de s'assurer au préalable qu'une telle collecte est réalisable. En effet, il faudra vérifier que les noms propres ont une présence significative dans les textes du corpus. Leur absence constitue, en effet, un cas rédhibitoire. On sait cependant qu'une partie d'entre eux a déjà été identifiée comme une composante importante dans un parler d'entreprise : *« L'ensemble de ces données [noms de marque et de produits] identifie et constitue la partie la plus représentative et extérieure du parler d'entreprise¹. »*

Par ailleurs, si la présence des noms propres est bien attestée, il faudra également s'interroger sur le rôle qu'ils occupent dans les textes sélectionnés. Lorsque celui-ci s'avère mineur, il faut pouvoir évaluer la pertinence de la collecte, même si ce point n'est plus un obstacle. Enfin, lorsque des noms propres sont à la fois présents et dotés d'un rôle non négligeable dans les discours, cela ne signifie

¹ D. De Vecchi (1999 : 322). On ajoutera également que si les noms propres de marques et de produits constituent une partie extérieure de parler d'entreprise, leur utilisation à l'intérieur de l'entreprise est également très courante.

pas encore qu'ils permettront aussi de capter des unités représentatives du vocabulaire d'un corpus. En effet, on peut imaginer qu'ils soient regroupés sous la forme de simples listes alphabétiques coupées du reste du vocabulaire.

Les contrôles de la présence, du rôle et de l'insertion des noms propres dans les textes du corpus correspondent à autant de tâches préalables qui permettront de garantir la possibilité d'une collecte d'unités destinées à alimenter un référentiel terminologique. Celle-ci devra néanmoins être évaluée au regard de sa productivité (nombre d'unités collectées à partir d'un nom propre), mais aussi de la restitution des propriétés générales du vocabulaire du corpus. Comme on peut le voir, l'hypothèse d'une collecte réalisée à partir des noms propres est encadrée par une série de contrôles. Ceux-ci sont destinés, d'une part, à valider l'hypothèse à l'échelle d'un corpus mais aussi, d'un point de vue plus général, en ce qui concerne l'observation des parlars d'entreprise, et d'autre part, à garantir que les collectes pourront être conduites utilement, à partir du moment leur faisabilité sera avérée.

4. Contraintes pesant sur la collecte

Pour la construction des référentiels terminologiques adaptables au contexte, on doit pouvoir s'appuyer sur des solutions logicielles robustes capables de travailler avec du texte brut ou bien du texte au format HTML. En effet, ces formats sont facilement manipulables pour un agent dont les compétences informatiques se limitent à l'utilisation de suites logicielles bureautiques et d'applications orientées utilisateur final, c'est-à-dire sans compétences techniques spéciales. Compte tenu du contexte d'exploitation retenu, on peut énumérer certaines caractéristiques que doivent posséder les programmes susceptibles d'être utilisés :

- la solution logicielle doit être maintenue par une équipe et disposer d'un plan de développement,
- elle doit être supportée par les systèmes d'exploitation les plus répandus (Windows, MacOS, Linux),
- elle doit être disponible dans une version stabilisée (sans bugs bloquants),
- elle doit être accompagnée d'un manuel d'utilisation explicite, donnant des indications sur les méthodes employées,

- elle doit permettre un apprentissage rapide pour l'utilisateur final (de l'ordre d'une demi-journée), dans un environnement de type bureautique,
- elle doit pouvoir être rapidement mise en œuvre,
- elle doit garantir l'autonomie de l'utilisateur, la conservation et l'intégrité de son travail effectué sur les données,
- son ergonomie doit être adaptée à une problématique où chaque unité isolée par un traitement doit pouvoir être observée dans son contexte d'origine,
- elle doit être robuste et posséder des performances comparables pour tout type de discours et tout volume de données (en restant néanmoins dans la limite de quelques millions d'occurrences),
- elle doit permettre de restituer les résultats obtenus sous la forme d'états imprimables ou exportables dans un format utilisable par d'autres logiciels,
- les traitements qu'elle effectue sur les textes ne doivent pas empêcher l'application ultérieure de traitements de niveau supérieur.

Dans le cadre de cette étude on a opté pour des logiciels de textométrie sachant travailler avec du texte brut¹. Dans la mesure où les traitements offerts peuvent reposer exclusivement sur la manipulation des formes graphiques (suites de caractères comprises entre deux espaces ou signes de ponctuation), ces programmes se situent en amont des exigences du traitement automatique du langage naturel (TALN) et permettent une souplesse et une rapidité de mise en place adaptées au but poursuivi. Par exemple, ils dispensent d'une analyse morpho-syntaxique, qui consiste à réduire les textes à une succession de lemmes. Cette sorte de traitement, outre qu'elle risquerait de peser sur la phase d'exploitation en raison des contrôles qu'elle requiert, n'a pas paru indispensable pour mener les différentes collectes qui seront présentées dans les chapitres de ce travail. Les logiciels utilisés constituent ainsi une approche complémentaire à celle du TALN² pour l'extraction d'informations, la constitution de lexiques, d'ontologies ou de terminologies, dans des contextes variés.

¹ Il s'agit ici de Lexico version 3.4 (A. Salem *et al.*) et d'Hyperbase version 5.1 (É. Brunet).

² Parmi les nombreux travaux ayant adopté cette approche on peut citer ceux de D. Bourigault et C. Jacquemin (Bourigault, 2000).

Les résultats de la collecte peuvent être stockés dans une base de données ou dans un tableur d'usage courant. Dans les deux cas, il s'agit de permettre des traitements en série et d'imposer un certain formalisme aux données (comme l'existence de listes de contrôle, pour éviter les doublons au moment de la saisie, la nécessité de réaliser des ajouts et des suppressions de façon explicite, etc.). Ce point est important dans la mesure où il garantit la possibilité d'exporter ces données dans un autre applicatif en cas de besoin. Les données étant rangées en colonnes et en lignes, des exports de type csv (comma separated values), ou texte au format tabulé, sont toujours possibles. Ces formats relativement standards d'échange de données doivent permettre de constituer des collections de référentiels au même format.

Enfin, la restitution des données collectées doit être adaptée aux deux besoins principaux identifiés pour améliorer l'organisation et la diffusion de l'information. Ces données devront pouvoir être importées dans un dictionnaire permettant à un moteur de recherche de réaliser des expansions de requêtes. Cela signifie, entre autres, que les relations établies entre les unités présentes dans un référentiel devront être explicites et suivre les recommandations des normes afférentes. Par ailleurs, afin que ce type de ressources permette de guider utilement les activités d'information relatives à la publication d'informations sur les intranets, il est nécessaire qu'il soit d'une consultation facile pour les agents. Celle-ci devra donc pouvoir se faire sous la forme d'une navigation arborescente combinée avec un système de recherche à l'aide de mots-clés.

La réalisation de la collecte des noms propres, puis de certaines unités caractéristiques du vocabulaire du corpus, doit être la plus automatisée possible. Néanmoins, il s'agit également de trouver un juste équilibre entre les exigences d'automatisation et celles relatives à la portabilité, qui ont été évoquées plus haut. En effet, l'expérience montre, dans le domaine du traitement automatique du langage naturel, mais dans bien d'autres domaines également, que plus le programme est automatisé, plus il est sensible aux modifications de contextes. Dans la mesure où la collecte doit pouvoir être réalisée sur des corpus relativement différents les uns des autres de part leurs propriétés linguistiques, cette contrainte a

pesé sur l'automatisation des phases de collecte. On accepte donc que les collectes soient plus assistées qu'automatisées à proprement parler.

En revanche, il a paru indispensable d'explicitier les étapes qui ont été retenues pour la réalisation de la collecte. En effet, cette explicitation garantit le caractère reproductible des résultats obtenus, ce qui doit être considéré comme un gage de leur qualité. Par ailleurs, on a vu que, compte tenu de la taille et de la diversité du groupe bancaire qui sert de contexte à cette étude, la production de référentiels doit se faire, dans la mesure du possible, de telle sorte que plusieurs personnes parviennent à des résultats comparables si elles exploitent le même corpus en respectant les mêmes étapes. On s'assure ainsi que les améliorations pourront être valablement répercutées à l'ensemble des agents pratiquant de telles analyses. On voit que la robustesse et la fiabilité de la méthode peuvent être de puissants facteurs pour son développement.

Enfin, la réalisation de tels référentiels doit prendre en quelque sorte le pas de l'organisation dans laquelle elle se développe. Elle doit s'insérer dans les pratiques, sans les remettre en cause, mais en se présentant sous la forme d'une évolution cohérente. Selon le profil des contributeurs et des webmestres, la constitution d'une telle ressource pourra être réservée à des experts ou rendue disponible auprès des diffuseurs de contenus qui en feront la demande. On a vu plus haut que l'activité d'information pâtit de son caractère invisible et donc difficile à chiffrer et à formaliser. La constitution de telles ressources doit permettre d'ajouter une dimension nouvelle et explicite à cette activité. Elle contribue ainsi à la renforcer et à asseoir sa visibilité dans les organisations.

La constitution des référentiels terminologiques adaptables aux contextes de diffusion de l'information de sites intranets doit également s'insérer dans la temporalité propre au déroulement des activités de l'entreprise. Il s'agit donc de ressources qui doivent pouvoir être régulièrement tenues à jour et partagées. Elles doivent être en prise avec les évolutions d'un référentiel dont il leur incombe de restituer les différentes formes de verbalisation dans un parler d'entreprise. Cette forte contrainte temporelle rend nécessaire de viser des restitutions exemplaires des variations observées : apparition et disparition de noms propres, association de

certaines d'entre eux avec des manières de classer et de catégoriser variables selon les secteurs de l'entreprise et les publics des sites. Les cas exemplaires ainsi recensés devraient permettre aux producteurs et aux diffuseurs d'information d'extrapoler à partir de leur propre connaissance du parler en usage dans leur entreprise.

Les échanges d'informations réalisés à l'aide des sites intranets, à l'intérieur d'un groupe bancaire rassemblant plusieurs centaines d'entreprises et 150 000 collaborateurs, sont actuellement confrontés à une situation paradoxale. D'une part, leur volume est en augmentation constante, car le recours à des systèmes d'information médiatisés s'inscrit dans une évolution tendancielle de l'organisation du travail. D'autre part, les activités consistant à prendre en charge l'organisation de ces informations, de manière à en assurer la diffusion auprès de leurs destinataires à l'intérieur de l'entreprise, sont généralement mal intégrées à la nouvelle organisation du travail. De fait, les activités d'information partagent, avec l'utilisation du langage, la propriété de ne pas être perçues comme des opérations en tant que telles par les agents qui les pratiquent quotidiennement dans leurs activités. Cette caractéristique a pour effet de maintenir ces pratiques dans un état largement tacite, qui fait obstacle à une prise en charge explicite des « ratés » de la communication électronique se faisant par pages HTML interposées : difficulté à trouver les informations pour les visiteurs des sites, déception devant les résultats souvent trop volumineux et imprécis fournis par les moteurs de recherche, mais également, absence de repères pour guider le travail des émetteurs de l'information.

La mise en place des systèmes d'organisation de l'information vise habituellement à résoudre ce genre de difficultés. Si les sites intranets partagent les objectifs assignés à de tels systèmes, leurs caractéristiques diffèrent, au point de rendre très difficile l'adoption des techniques mises au point dans le cadre des seconds. En effet, alors que les systèmes d'organisation de l'information sont traditionnellement placés en marge de l'activité des organisations, ce qui leur permet de développer leurs pratiques selon une temporalité qui leur est propre, la diffusion des informations sur les intranets est étroitement liée au rythme de l'activité de l'entreprise, en raison de leur nature essentiellement opérationnelle.

La participation directe des intranets à la réalisation des activités d'une entreprise a donc deux conséquences qui pèsent sur la réalisation des activités d'information. La première tient au caractère essentiellement éphémère, « jetable » pourrait-on dire, et instable, des informations et de leurs supports. Les intranets ignorent la sédimentation des informations qui caractérise les fonds des

bibliothèques ou des centres de documentation. Cette instabilité des fonds complique la réalisation de descriptions secondaires des contenus, destinées à permettre leur organisation et leur recherche. Or, la remise en cause de l'ontologie des systèmes d'organisation de l'information n'est pas la seule conséquence de l'insertion des intranets dans l'activité des entreprises. En effet, la deuxième conséquence tient au caractère inapproprié du recours à une normalisation de l'usage de la langue passant par la mise en place de langages documentaires. Le recours à ces derniers, qui apporte une réponse aux variations affectant les usages d'une langue, repose sur l'adoption explicite d'une terminologie propre à un corps de gestionnaires de l'information. Outre que celui-ci n'existe généralement pas dans le contexte des intranets, un tel usage normalisé de la langue ne ferait qu'ajouter une terminologie de plus aux nombreuses composantes d'un parler d'entreprise.

Dès lors, il apparaît que la mise en place de ressources terminologiques visant à faciliter l'organisation et la diffusion de l'information sur les intranets devra tenir compte des contraintes particulières caractérisant ce type de contexte. Mais on a tiré un autre enseignement de la présentation des pratiques d'information qui sont réalisées dans un tel cadre : leur efficacité gagnerait grandement à être guidée par un usage explicite de la langue. Dans la mesure où celui-ci ne semble pas pouvoir passer par le recours à une normalisation – dont l'efficacité peut également être sujette à caution –, on a fait l'hypothèse que cette explicitation pourrait provenir d'une exposition de la diversité des usages langagiers propre au sociolecte de l'organisation dans laquelle les intranets prennent place. Cette exposition, en permettant aux diffuseurs d'information de situer leurs manières de dire par rapport à celles des publics de leurs sites, devrait fournir aux premiers un échantillon d'expressions à utiliser, afin de tenir compte des manières de catégoriser et de classer propres aux seconds. La mise en place d'une telle ressource suppose que l'on se dote d'un cadre théorique à partir duquel une forme de cahier des charges puisse être élaborée.

La mise en place d'une sorte d'observatoire des usages linguistiques propres à une organisation suppose d'adopter un modèle de la communication verbale qui soit à même d'intégrer les caractéristiques des situations de communication réelles. Ces dernières sont observables, entre autres, à partir des traces qu'elles laissent dans

les discours. Un référentiel terminologique adaptable au contexte doit pouvoir être constitué à partir d'observations réalisées sur des échantillons de discours produits dans et par les organisations. On propose donc la constitution de corpus ouverts regroupant des séries textuelles homogènes organisées chronologiquement, de manière à permettre un suivi de la manière dont les activités de l'entreprise sont verbalisées. Chaque corpus correspond alors à une série textuelle produite dans le cadre d'une situation de communication qu'il est possible de caractériser finement par rapport aux autres situations rencontrées au sein d'une organisation. De cette manière, chaque référentiel terminologique devient un échantillon d'un ensemble plus vaste correspondant à une forme particulière de sociolecte, un parler d'entreprise. Cette notion permet de rendre compte du fait que le sociolecte en usage dans chaque entreprise est traversé, non seulement par les terminologies de différents secteurs, ou domaines d'activité, mais aussi par un *continuum* s'étendant des unités relevant de la langue commune à celles qui sont propres à une organisation donnée.

Un référentiel terminologique est donc doublement situé : d'une part, en raison des caractéristiques des conditions de production des discours dont il est tiré, d'autre part, parce qu'il doit représenter les différentes composantes de l'usage particulier d'un parler d'entreprise spécifique. Il reste que la mise en place de tels référentiels ne doit pas nécessairement viser une restitution exhaustive du parler d'une organisation, si tant est que cette tâche soit réalisable. Compte tenu du fait que cette sorte d'outil doit permettre de guider des activités d'information, elles-mêmes prises dans la réalisation des activités de l'entreprise, il semble préférable de viser des restitutions exemplaires sous le rapport des conditions de production des discours et de la restitution d'un parler. On propose d'illustrer ce caractère exemplaire en explorant les textes servant de sources aux référentiels terminologiques à partir d'unités linguistiques dotées de propriétés pragmatiques singulières, les noms propres. Ces derniers permettent à un groupe plus ou moins étendu de locuteurs de dénommer des objets ou des êtres dotés d'une certaine stabilité, au sujet desquels des prédications répétées ont lieu dans les échanges, et pour lesquels il n'y a pas de descriptions définies satisfaisantes. Ce faisant, les noms propres désignent des particuliers remarquables dans le référentiel d'un groupe

donné. Ils constituent de ce point de vue des unités susceptibles de figurer dans un référentiel terminologique. Mais on fait également l'hypothèse que la stabilité de leur lien dénomiatif permettra non seulement des comparaisons entre des usages variables d'un parler selon les situations de communication, mais aussi de capter les modifications de ces usages dans le temps. Les noms propres deviennent alors les pivots autour desquels seront articulées, d'une part, l'évocation des référentiels mobilisés dans différentes situations de communication, et d'autre part, les variations affectant leur verbalisation, telles que les référentiels terminologiques devront les restituer.

La réalisation de référentiels terminologiques adaptables au contexte doit permettre d'éprouver ce modèle théorique. Chaque étape de la constitution de cette ressource, dédiée à l'organisation de l'information sur les intranets, nécessitera non seulement un approfondissement de certains points abordés dans ce chapitre, mais aussi des propositions permettant de démontrer sa faisabilité, et donc de son caractère opérationnel, à partir de données textuelles tirées du contexte retenu pour cette étude. Ce dernier point intégrera donc les contraintes pesant sur les activités d'information dans le cadre des intranets, qu'elles concernent la constitution et l'exploitation des corpus, la caractérisation de leur vocabulaire au regard de l'usage d'un parler d'entreprise, ou encore la réalisation de la collecte des unités destinées à peupler les référentiels, mais aussi leur organisation en vue de leur utilisation et de leur restitution. Les deux chapitres suivants sont consacrés au premier de ces points, relatif à la constitution de ressources textuelles au format électronique.

CHAPITRE 2

UN RÉFÉRENTIEL À L'ÉPREUVE DES DISCOURS

Le choix d'un corpus se fait généralement en fonction de questions que le chercheur se pose à partir d'une problématique extérieure qu'il s'agit de vérifier. Des travaux d'historiens peuvent ainsi s'appuyer sur un corpus de textes afin de mettre en lumière la manière dont les forces politiques de gauche et de droite s'affrontent par discours interposés à une époque donnée¹. Pour ce qui est de ce travail, on cherche à savoir dans quelle mesure les noms propres présents dans un texte ou un groupe de textes, produits dans le cadre d'une organisation, permettent de capter un échantillon caractéristique de l'usage d'un parler d'entreprise, et partant, d'un référentiel spécifique. Par conséquent, la vérification est en quelque sorte interne, dans la mesure où l'on interroge d'abord le vocabulaire du corpus pris pour lui-même, comme trace particulière laissée par une situation d'interlocution. Si l'on s'en tient à cette seule considération, ce sont à peu près tous les textes produits dans le cadre d'une organisation qui sont dignes d'intérêt pour la constitution d'un référentiel terminologique adaptable au contexte.

Dès lors, le choix du corpus doit être effectué en fonction des objectifs assignés au référentiel terminologique à constituer. Ceux-ci dépendent de l'activité d'information qu'il s'agit de documenter, à l'aide d'une présentation des caractéristiques de l'usage d'un parler d'entreprise, compte tenu du, ou des publics visés. Lorsque les principales caractéristiques du public ont été déterminées, il s'agit de sélectionner un corpus de textes, de préférence homogène, et doté d'un minimum de profondeur temporelle, de manière à pouvoir identifier, dans la mesure du possible, ce qui relève du vocabulaire caractéristique, de la tendance ou de l'accidentel. La question générique, relative à la présence et au rôle des noms propres dans les textes retenus comme sources pour la constitution du référentiel, se

¹ Ainsi A. Prost dans *Vocabulaire des proclamations électorales de 1881, 1885 et 1889* : « La conclusion principale de l'enquête est cependant d'une autre nature : elle porte, globalement, sur le vocabulaire des proclamations électorales ; nous pensons, en effet, avoir prouvé que ce vocabulaire était caractérisé par l'opposition fondamentale et permanente des deux pôles classiques de la droite et de la gauche. » (1974 : 153). Plus récemment, D. Mayaffre, dans *Le poids des mots – Le discours de gauche et de droite dans l'entre-deux-guerre expose sa problématique* : « Plus grossièrement, le choix de deux hommes de droite et de deux hommes de gauche relève de la problématique de cette étude. L'objectif est d'étudier la guerre civile verbale que se livrent les Français et les forces politiques, jusqu'à affaiblir mortellement la France républicaine. Mais en quels termes les Français s'opposent-ils ? Sur quels thèmes les forces politiques se déchirent-elles ? » (2000 : 18).

pose maintenant en des termes plus précis : à quels pans du référentiel de l'organisation ces unités donnent-elles accès dans l'échantillon du genre de discours retenu pour le corpus ? Quant à la pertinence du référentiel terminologique à constituer, elle devra être évaluée à la lumière des améliorations apportées dans la réalisation des activités d'information. Ce point n'est pas abordé dans le cadre de ce travail, qui traite essentiellement des questions relatives à la constitution d'une telle ressource, compte tenu du cahier des charges qui a été donné dans le premier chapitre.

Après avoir présenté le type d'activité d'information à documenter, on discutera le choix du corpus qui a été retenu. Ce dernier doit lui-même faire l'objet d'une documentation de manière à permettre, d'une part, un début de localisation parmi les productions discursives de l'entreprise, et, d'autre part, une première évaluation des informations relatives au parler d'entreprise qu'il permettra de capter. Il s'agit dans les deux cas d'asseoir la phase de collecte sur un choix explicite et dont on aura présenté les limites. On commencera par donner une description des principales caractéristiques des documents rassemblés, qui sera complétée par la présentation de la manière dont ils ont été organisés sous la forme d'un corpus au format électronique.

I CORPUS ET RÉFÉRENTIEL

A Intranetes et offre d'information sur l'intranet d'un groupe bancaire

Le groupe bancaire qui a été retenu pour servir de contexte à ce travail est composé de nombreuses entités, dont la taille varie de quelques dizaines d'agents à plusieurs dizaines de milliers. Chaque entreprise est organisée en directions, qui correspondent à la plus importante unité interne d'organisation du travail. Chaque direction est elle-même composée de plusieurs métiers diversement organisés, ce qui explique que, selon les cas, une direction puisse ne posséder aucun site intranet, ou bien avoir la responsabilité d'en animer un ou plusieurs, selon l'importance des métiers. Par ailleurs, les intranets sont diversement accessibles aux intranetes d'un tel groupe. À l'intérieur d'une même entreprise, certains sites ou certaines informations peuvent être réservés à une partie seulement du personnel. De même, les intranets de certaines entreprises peuvent n'être visibles que de leurs seuls salariés, alors que d'autres sites sont susceptibles d'être consultés par l'ensemble des agents travaillant dans le groupe bancaire. Enfin, il est nécessaire d'établir une différence entre l'offre d'information théorique disponible sur les intranets d'une entreprise pour un agent donné, et la connaissance que cet agent possède d'une telle offre lors des navigations qu'il réalise. Cette simplification assez grossière d'un intranet permet néanmoins d'identifier des situations caractéristiques que l'on a résumées dans le schéma ci-dessous.

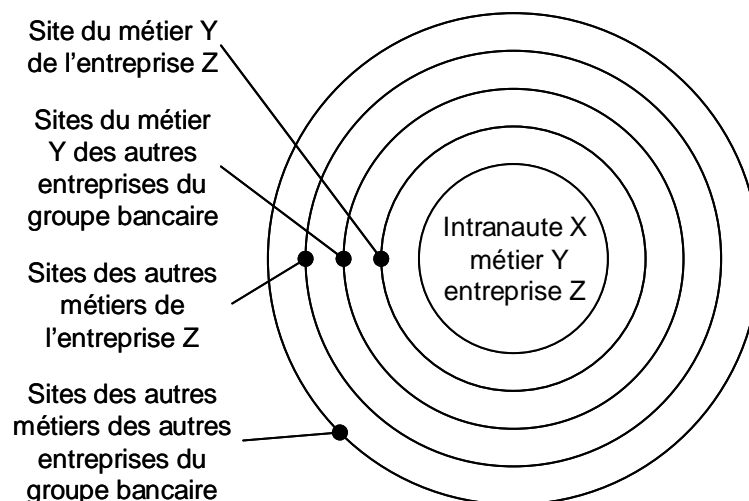


Figure 7.

Pour un intranaute caractérisé par son métier et son entreprise de rattachement, situation des différents sites intranets disponibles

Lecture de la figure 7 : Le centre de la figure représente un intranaute quelconque caractérisé par son métier, et le secteur d'activité auquel il se rattache, ainsi que l'entreprise dont il est salarié. Sur les cercles concentriques sont présentés les différents sites consultables par l'intranaute en fonction de leur degré d'éloignement par rapport, d'une part, au référentiel théoriquement nécessaire à l'intranaute pour l'exercice de son métier, et d'autre part, à l'entreprise émettrice d'informations via un site intranet. Pour ce schéma, on a considéré que la proximité du métier prime sur celle de l'entreprise. En fait, cette situation est susceptible de varier pour chaque site en fonction des sujets qu'il aborde et de l'ancienneté de l'intranaute. Par exemple, le site du comité d'entreprise sera théoriquement plus accessible pour un agent ayant de l'ancienneté que pour un nouvel embauché. Il s'agit d'une hypothèse basée sur des observations empiriques qui pourront être contrôlées de façon plus systématique.

Le schéma de la figure 7 met en évidence l'inégale position des sites de l'intranet du groupe bancaire par rapport au référentiel d'un intranaute quelconque, défini principalement par son métier et l'entreprise dans laquelle il exerce son activité. En effet, on peut avancer que plus le site est éloigné de l'intranaute en fonction des deux critères que l'on a retenus – pour être plus complet il faudrait ajouter d'autres caractéristiques liées, non seulement à l'intranaute, comme son ancienneté, son statut, le type de métier exercé, etc.¹, mais aussi à l'entreprise,

¹ Il serait ensuite possible de descendre à un niveau de précision encore plus grand. Il faudrait alors indiquer le type de besoin d'information d'un intranaute précis, étant donné qu'il peut varier d'une

comme le fait d'être une filiale ou une banque indépendante, le fait de concevoir des produits distribués par les réseaux bancaires ou d'exercer une activité d'ingénierie financière, etc. –, plus il est probable que l'agent devra fournir des efforts cognitifs importants pour trouver une information. Or, on sait par ailleurs que ce type de démarche est soumis à une « loi » dite de Moers : « (...) *un usager tend à ne pas utiliser un système d'information lorsqu'il lui apparaît qu'il est plus pénible d'obtenir l'information grâce à ce système que de ne pas l'avoir en ne l'utilisant pas*¹. » Si elle n'est pas prise en charge explicitement, la diffusion des informations sur un site intranet se trouve dans une situation paradoxale : sa raison d'être est de toucher un public élargi, qui risque cependant de renoncer à consulter ce qui sort du cercle de ses repères habituels.

Lorsque l'on effectue une lecture du schéma donné ci-dessus, non plus en prenant pour point de départ le centre représentant l'intranaute, mais en partant des sites intranets, qui incorporent les activités d'information nécessaires à leur fonctionnement, il est possible de dégager une image simplifiée des conditions d'accès à ces sources d'informations. Plus ces dernières se trouvent éloignées des visiteurs (géographiquement, culturellement, relativement à une forme d'activité, etc.), plus la prise en charge de la situation de communication liée à leur diffusion doit viser à combler un écart cognitif important. Il en est vraisemblablement ainsi lorsqu'un site intranet de l'organe central du groupe bancaire, pris comme contexte opérationnel pour ce travail, s'adresse à des agents rattachés à une filiale ou à une banque régionale, et dont le métier s'exerce dans un autre secteur d'activité que celui auquel le site appartient. On peut alors se demander si, dans une telle situation,

recherche d'information à l'autre. Pour une approche cognitive de cette question et la présentation des travaux réalisés dans le domaine des sciences de l'information, on renvoie, entre autres, aux publications de P. Ingwersen (1992), (1996) et (1999). On note, cependant, que dans cette perspective, la réalité des échanges linguistiques est quelque peu négligée au profit du seul plan conceptuel. De façon tout à fait classique dans ce domaine, la langue est susceptible de représenter un obstacle entre la représentation mentale, conceptuelle, du besoin d'information, et sa formulation (cet écart est nommé « *delta problem* »). L'usage de la langue y est pratiquement réduit à un « *label effect* » : « *The phenomenon that request formulations may often consist of one or several concepts which are of a more general nature or out of the context which constitutes the real information need.* » P. Ingwersen (1992 : 116-118 et 229). Par ailleurs, on notera aussi qu'un tel modèle peine à appréhender une réalité telle qu'un sociolecte, pris qu'il est entre une représentation unitaire de la langue et l'idiolecte de l'utilisateur d'un système d'information.

¹ Cacaly et alii (2004), article *Loi de Moers*.

la plupart des visiteurs du site ne doivent pas être assimilés à des néophytes. C'est cette situation limite que l'on a retenue dans le cadre de ce travail, dans la mesure où elle a semblé de nature à embrasser des difficultés, qui dans des contextes d'éloignement moindre, sont probablement plus faciles à résoudre. Il n'en reste pas moins que la question de la diffusion d'informations pour un public de non initiés se pose pour chaque position représentée par l'un des cercles du schéma. La question afférente de la documentation des activités d'information doit donc être également envisagée pour un site rattaché à un secteur d'activité dans une entreprise donnée et diffusant des informations pour un public « métier » au sein de la même organisation. Il convient maintenant d'envisager la sélection des ressources susceptibles de documenter la situation type retenue, celle de la diffusion des informations à partir d'un site intranet de l'organe central en direction des filiales et des banques du groupe bancaire.

B. Ressources destinées à documenter les activités d'information

1. Propriétés remarquables des discours de communication d'entreprise

La documentation des activités d'information d'un agent chargé de publier, à partir d'un site de l'organe central, des informations destinées aux personnels des entités dont se compose un groupe bancaire, peut reposer sur plusieurs référentiels terminologiques. Ceux-ci pourraient représenter plusieurs situations de communication caractéristiques du *continuum* qui a été présenté dans la précédente section. Cependant, un tel dispositif suppose que le parler d'entreprise concerné ait déjà fait l'objet d'un recensement concerté et assez étendu¹. Dans la mesure où l'on a pu voir qu'une telle description n'a pas encore été entamée dans le contexte retenu,

¹ Comme cela a été rappelé dans le précédent chapitre, l'étude des productions discursives au sein des organisations constitue un terrain d'investigation relativement nouveau. Si l'on s'en tient aux seules informations écrites, un tel contexte recèle des genres et des sous-genres dont le recensement et la classification sont toujours en cours : « *Un relevé des écrits mentionnés par les chercheurs du réseau Langage et Travail en 1993 donnait la liste suivante : « bilan, comptes de profits et de pertes, règlements d'activité, livret ouvrier, registre matricule, fichier du personnel, contrats, contentieux, relances, statuts, audits, projets, études, circulaire, bordereaux, fiches de paye, autorisations d'absence, de déplacement, bons de retrait d'un magasin, consignes, modes d'emploi, documentations techniques, cahier de suivi d'un malade, commandes, bons d'approvisionnement, bons de fabrication. » On pressent derrière cette énumération un continent méconnu, des inventaires à entreprendre, des typologies à construire... »* A. Borzeix et B. Fraenkel (2001 : 238-239).

il a néanmoins paru nécessaire d'amorcer ce travail, en proposant l'examen d'une première sorte de ressources. Celles-ci relèvent de la communication d'entreprise qui assume, dans le cadre du fonctionnement d'une organisation, la production d'un certain nombre de discours en prise avec l'activité.

En effet, chaque entité est ordinairement en charge des tâches de communication relatives à son activité¹. Celles-ci sont assurées par des personnels spécialisés qui se répartissent en fonction des formes d'activité concernées : communication institutionnelle relative à l'entreprise comme entité économique dotée d'une image pour le marché, marketing traitant de la communication liée à la commercialisation des produits, relations publiques concernant les relations avec certains publics utiles à l'entreprise, enfin, communication interne dédiée à la diffusion d'informations relatives à l'activité à l'intérieur de l'entreprise. Dans tous les cas, il s'agit de discours « autorisés » en ce sens qu'ils ne peuvent être diffusés sans l'aval de la direction de l'entreprise et qu'à ce titre, ils manifestent l'existence d'une autorité en même temps qu'ils l'incarnent. Ces discours sont donc généralement soumis à un formalisme plus contraignant que les parlers des métiers ; il conviendra donc d'évaluer cette caractéristique. Cependant, on a vu plus haut que cette « langue de l'entreprise » constitue néanmoins l'une des composantes d'un parler d'entreprise, et qu'à ce titre elle constitue une source possible pour la construction de référentiels terminologiques.

Les discours de la communication d'entreprise présentent également une autre particularité pertinente, dès lors que l'on cherche à documenter la sorte d'activité d'information que l'on a retenue dans ce travail. En effet, les principaux types de communication doivent prendre en charge la difficulté soulevée par l'évocation du référentiel propre à une organisation auprès d'un public qui en ignore tout, ou presque. Cela se rencontre pour la communication institutionnelle, lorsqu'elle s'adresse aux journaux ou directement à des millions de petits porteurs d'actions de l'entreprise, c'est aussi le cas pour la communication commerciale qui cherche à conserver plus d'une vingtaine de millions de clients ; enfin, les publications de la communication interne s'adressent pour certaines à plus de

¹ On reprendra ici la classification proposée par T. Libaert (1998).

130 000 salariés. Ces discours font rarement l'objet d'une attention très poussée de la part de leurs destinataires : « *Vivre dans le monde contemporain, c'est être confronté à une multitude de textes aussi éphémères qu'envahissants : tracts, annuaires, journaux, affiches, guides touristiques, « mailings » publicitaires... Énoncés survolés, feuilletés, consultés, rarement lus, au sens plein du mot*¹. » Le fait que ces discours soient négligés par leurs destinataires, ou bien même, pour certains d'entre eux, qu'ils ne soient pas destinés à être lus², ne constitue pas un défaut rédhibitoire, dès lors que ces textes sont produits dans le cadre de l'entreprise et qu'ils participent de son parler. Leur caractère « jetable » les rapproche des productions discursives de l'entreprise. En revanche, il est de nature à entraver leur sélection en vue de la constitution de corpus.

En effet, les informations véhiculées par ces différentes formes de la communication d'entreprise sont soumises à deux formes de contraintes liées à l'activité : l'actualité et le renouvellement permanent. Cette dernière contrainte est illustrée de manière exemplaire par la commercialisation régulière de nouveaux produits destinés à remplacer les produits existants en fonction d'une logique de marché concurrentiel, où l'effet de mode prime sur la réalité du renouvellement. Par ailleurs, l'actualité de l'entreprise modèle les informations mises en forme par la communication interne. Ces deux phénomènes se rencontrent dans les productions de la communication institutionnelle à qui il incombe de rendre compte de l'actualité des différentes formes de l'activité de l'entreprise. Mais alors que les deux premières formes de communication traitent plutôt de faits nouveaux qui se chassent les uns les autres, la communication institutionnelle doit combiner cet aspect avec des éléments plus permanents, caractéristiques de l'entreprise.

Cependant, la communication institutionnelle ne forme pas un ensemble homogène. Pour un groupe bancaire coté en bourse, celle-ci comporte au moins des opérations ponctuelles, telles que celles qui accompagnent les OPA (offres publiques d'achat) ou les augmentations de capital, des assemblées ordinaires d'actionnaires, réunies chaque année, des annonces légales destinées à informer le

¹ D. Maingueneau (2000).

² A. Borzeix et B. Fraenkel (2001 : 123).

marché de certaines opérations financières, des lettres périodiques aux actionnaires permettant d'entretenir avec ces derniers des liens privilégiés en les tenant informés de l'activité de l'entreprise, enfin, des rapports annuels, qui constituent un outil de base pour cette forme de communication. Ce sont ces documents qui proposent une présentation des différentes activités de l'entreprise, en combinant l'actualité remarquable de l'année écoulée avec les éléments plus permanents qui distinguent telle entreprise de ses concurrentes. Dans la mesure où ces informations sont diffusées en direction d'un large public, aussi bien interne qu'externe, l'étude de leur traitement discursif a semblé de nature à documenter les activités d'information retenues dans le cadre de cette étude.

En effet, dans une perspective de documentation d'activités d'information liées à la publication d'information sur les intranets, ce type de document présente plusieurs intérêts. La matière principale y semble bien relative au référentiel de l'entreprise émettrice. En raison de sa nature, ce type de document s'adresse en priorité à un public plutôt externe. Cependant, les rapports d'activité, on le verra plus loin, s'adressent également de façon directe ou indirecte aux salariés de l'entreprise émettrice. Enfin, ils constituent une trace de discours d'entreprise produits dans une situation de communication pour laquelle on peut penser qu'une forme de prise en charge des lecteurs non initiés a été utilisée.

Ces remarques appellent deux commentaires, qui assignent des limites aux résultats que l'on peut escompter de l'exploitation de tels discours aux fins d'organisation de l'information. Tout d'abord, leur utilisation dans ce cadre ne signifie pas que l'on porte un jugement quelconque sur la qualité des réalisations de la communication institutionnelle. Il s'agit, en effet, d'un usage détourné des buts premiers assignés à ces documents, qui sont pris tels qu'ils sont. Ceux-ci sont considérés comme une production discursive spéciale, pour laquelle l'entreprise fournit des efforts particuliers afin de communiquer en direction d'un public qui n'est pas censé partager son référentiel. Compte tenu de ce qui a été présenté plus haut relativement aux activités d'information sur les intranets, ils constituent une forme d'exception propre aux activités de communication. En second lieu, adopter comme source ce type de documents ne signifie pas qu'il n'existe pas d'autres discours sur l'activité d'une entreprise. C'est pourquoi on s'attachera à les situer par

rapport à ce que pourrait être la restitution globale d'un référentiel d'entreprise ainsi que sa verbalisation à l'aide d'un parler d'entreprise lorsque celle-ci a pour cadre d'autres situations de communication. Mais avant d'entreprendre ce travail de mise en contexte et de présentation d'un genre de discours donné, il reste à évaluer la période pertinente à retenir dès lors qu'il s'agit aussi de suivre les évolutions ayant affecté la manière de restituer l'activité d'une entreprise.

2. Référentiel d'entreprise et période couverte par un corpus de textes

L'organisation de l'information, sa diffusion et sa recherche sont des tâches exercées en permanence sur un réseau intranet. L'étude d'un référentiel tel qu'il peut être utilisé dans une situation de communication particulière suppose donc des mises à jour régulières. Pour cela, on a vu qu'il est nécessaire de se doter d'un corpus de suivi, ou « *monitor corpus* », qui a la propriété d'être ouvert afin d'accueillir des données textuelles fraîches au fur et mesure qu'elles sont produites. On verra que cela implique de formaliser les traitements appliqués aux données composant le corpus, afin de garantir la fiabilité et la pertinence des observations dans la durée.

Par ailleurs, l'exploration d'un référentiel spécifique à partir d'un corpus suppose de pouvoir distinguer ce qui relève de l'accident, de la tendance ou de l'usage récurrent. Or, un corpus est constitué de discours datés qui, pris isolément, ne peuvent fournir que des indications partielles. Dans le but de rendre lisible ce qui change au cours du temps, on a opté pour la constitution d'une série textuelle chronologique¹. La constitution d'un corpus rend ainsi nécessaire l'explicitation du type de relations pouvant articuler deux temporalités différentes : celle des activités d'information qu'il s'agit de documenter (utilisations ponctuelles en 2004), et celle du référentiel d'entreprise que l'on souhaite prendre en compte. En effet, la portée de celui-ci ne se limite pas aux utilisations ponctuelles que l'on en fait. Il a donc été nécessaire de définir une durée adaptée, entre la temporalité courte qui caractérise les tâches d'organisation, de diffusion et de recherche de l'information ainsi que la

¹ « Nous réservons donc l'appellation de « série textuelle chronologique » aux séries périodiques homogènes constituées par des textes produits dans des situations d'énonciation similaires, si possible par le même locuteur (individuel ou collectif) et présentant des caractéristiques lexicométriques comparables. » A. Salem (1988 : 107).

durée de vie des documents mis en ligne sur un intranet, qui oscille entre quelques jours et deux à trois ans maximum, et la durée de vie d'un référentiel dont certaines composantes ont plus d'une centaine d'années.

Si l'on souhaite embrasser un référentiel dans toute son étendue, il serait certainement nécessaire de remonter au moins jusqu'à la naissance de l'institution, ce qui, dans le cas du Crédit Agricole, correspond à une période très éloignée. En effet, les actes fondateurs de cette banque remontent à 1894 pour les caisses locales, 1899 pour les caisses régionales, et 1920 pour l'organe central qui produit le rapport annuel. Cet organe a lui-même évolué : l'Office national du Crédit agricole (future Caisse Nationale de Crédit Agricole), devient un EPIC¹ en 1979, puis, en 1988, une société anonyme détenue majoritairement par les caisses régionales de Crédit Agricole. En 2001, la Caisse Nationale de Crédit agricole devient une société cotée en bourse et possédant une nouvelle dénomination : « Crédit Agricole SA ». Si l'on envisage de traiter plus particulièrement l'activité bancaire, un autre découpage pourrait s'appuyer sur les extensions de compétence du Crédit agricole en tant que banque. Mais dans ce cas, faut-il partir du financement des prêts à long terme aux coopératives agricoles (1910), et reposer la question à chaque nouvelle extension, jusqu'au financement des entreprises accordé par le législateur en 1991, et qui achevait ainsi la transformation de l'établissement en banque universelle² ?

On a vu plus haut que le fait d'aligner le point de départ chronologique du corpus sur une durée correspondant, approximativement, au temps de conservation des documents des sites sur un intranet ne présentait qu'un faible intérêt. En effet, on ne peut pas réduire le référentiel mobilisable à un période donnée, aux seules informations contenues dans les documents utilisés sur un intranet. Cela reviendrait, d'une part, à négliger l'usage que les intranauts peuvent faire de leur mémoire, et d'autre part, à postuler une sorte de connaissance universelle instantanée, pour laquelle toutes les informations disponibles sur un intranet seraient immédiatement utilisables par tous. En d'autres termes, ce serait adhérer sans réserve à une idéologie

¹ Établissement public à caractère industriel et commercial.

² Pour l'histoire du Crédit Agricole, on revoie aux ouvrages d'A. Gueslin (1978, 1984 et 1985), et pour la période récente A. Choinel (1996 et 2002). Enfin, les rapports d'activité de la période constituent une autre source relevant de la littérature grise. Des éléments complémentaires sont fournis par la bibliographie.

de l'instantanéité qui, si elle accompagne le développement des nouvelles technologies, ne peut néanmoins pas être acceptée telle quelle.

Pour tenter de concilier les exigences liées à la prise en compte de ces deux aspects d'un même référentiel, l'ancienneté de certaines composantes potentiellement présentes dans la mémoire longue des agents, d'une part, et d'autre part, la partie de celui-ci qui est effectivement mobilisée lors d'une activité d'information datée, en l'occurrence l'année 2004, on a opté pour un moyen terme. Pour cela, on a recherché l'événement le plus récent et le plus susceptible d'avoir eu des répercussions importantes sur le référentiel à étudier. Lorsque ce travail a été initié, en 2001, l'acquisition de la banque Indosuez, en 1996, est apparue comme un bon point de départ. En effet, on peut tenter de suivre la façon dont un événement majeur de cette sorte a été traité sur la durée. Avec la cotation de Crédit Agricole S.A. à la fin de l'année 2001, le corpus disposait d'une clôture provisoire ponctuée par l'apparition d'un deuxième événement majeur. Néanmoins, il s'est avéré utile de reculer le point de départ d'une année, afin de disposer d'une sorte de témoin de la période précédente, inaugurée en 1991. Par ailleurs, le Crédit Agricole ayant procédé au rachat du Crédit Lyonnais en 2003, on a repoussé les limites du corpus de manière à pouvoir intégrer ce troisième événement majeur. En définitive, on a retenu une période allant de 1995 à 2003. Celle-ci permet l'observation de trois bouleversements d'importance ayant affecté l'institution et, par ricochet, le référentiel étudié. Il convient maintenant de caractériser plus finement le genre de discours qui a été retenu, afin de constituer la source du référentiel terminologique adaptable au contexte que l'on projette de mettre en place.

II. RAPPORT ANNUEL ET RAPPORT D'ACTIVITÉ

A. Un « outil » de communication institutionnelle

1. Le rapport annuel est un document financier

Le corpus principal que l'on utilise dans le cadre de cette étude se compose de rapports d'activité. Ceux-ci sont généralement associés à un document plus complet dénommé, le rapport annuel : « *Document qui décrit la situation ainsi que les activités d'une entreprise ou d'un organisme durant l'exercice écoulé et qui doit*

être établi annuellement par les administrateurs ou les dirigeants¹. » Les rapports annuels sont des documents publics produits par les entreprises et dont la diffusion en plusieurs langues peut dépasser les 100 000 exemplaires². Depuis le développement de l'Internet dans la deuxième moitié des années 1990, ils figurent souvent en bonne place sur les sites de leurs émetteurs. Ils regroupent deux sortes d'informations qui ne sont pas de même nature mais qui, probablement pour cette même raison, se trouvent être complémentaires. On commencera par présenter les principales propriétés de cet outil de communication institutionnelle, avant d'envisager l'usage d'un genre de discours particulier.

Historiquement, le rapport annuel est une compilation d'informations relatives aux aspects comptables et financiers du fonctionnement de l'entreprise. Il comprend le rapport de gestion et les comptes, le bilan consolidé, les notes nécessaires à la compréhension des comptes, les rapports des commissaires aux comptes et un certain nombre d'informations juridiques et administratives. Ces informations sont encadrées par des textes législatifs et leur présentation fait l'objet de normes diffusées à l'échelle internationale.

En ce qui concerne la France, l'appel à l'épargne publique, à travers le mécanisme de la mise en vente d'actions représentant des parts du capital de sociétés de capitaux (banques, entreprises industrielles et commerciales), est encadré par le législateur depuis 1867. Dès cette époque, celui-ci a édicté des règles relatives à la publication d'informations financières destinées à renseigner les détenteurs de parts de capital ainsi que les investisseurs potentiels. Ces dispositions sont reprises dans le *Code de commerce*³ pour les sociétés anonymes en général et dans le *Code monétaire et financier*⁴ pour les établissements de crédit, les entreprises d'investissement et les compagnies financières⁵.

¹ Office Québécois de la langue française, *Grand dictionnaire terminologique*, (consultable à l'adresse suivante : <http://www.granddictionnaire.com>).

² A. Guimard (1998 : 57-63).

³ *Code de commerce*, Paris, Dalloz, 98e édition, 2002, LIII-2482 p.

⁴ *Code monétaire et financier*, actualisé en permanence sur le site : <http://www.legifrance.gouv.fr>.

⁵ Pour la production des comptes sociaux et des documents comptables, les établissements de crédit et les entreprises d'investissement sont soumis aux dispositions des articles L.232-1 et L. 232-6 du code de commerce (art. L. 511-35 à L. 511-37 du *Code monétaire et financier*). Ces articles précisent les informations qui doivent être annexées au bilan.

Les données comptables et financières fournies par le rapport annuel font de ce dernier une source primaire précieuse pour la connaissance économique des entreprises. Ces informations font l'objet d'analyses, de commentaires mais aussi de contrôles engageant la responsabilité des dirigeants de l'entreprise. Leur importance est telle que pour les milieux financiers, les économistes spécialisés ainsi que pour la presse financière et économique, l'intérêt du rapport annuel est fréquemment réduit à ces seules données¹. C'est certainement une conception réductrice, au moins dans la mesure où elle ne tient pas compte de l'intention de l'émetteur, qui a recherché à l'occasion de cette présentation annuelle et obligatoire, le moyen de communiquer également sur ses activités.

2. Le rapport annuel est un support de communication

À partir de la fin des années 1960², ces éléments financiers ont été fréquemment introduits par une présentation plus générale de l'entreprise et de son contexte économique (les métiers exercés en son sein, les produits et services qu'elle propose, son organisation, le secteur d'activité dans lequel elle opère, etc.). Agrémentée de graphiques, d'histogrammes et de photographies, cette première partie se veut agréable à regarder et présente l'activité sous un jour simplifié, plus économique que directement financier, comme c'est le cas pour la présentation des comptes. Progressivement, cette longue introduction aux comptes (elle représente fréquemment une cinquantaine de pages, voire plus), a été pensée et construite comme une véritable vitrine de l'entreprise. Celle-ci y est présentée sous un jour favorable à des publics divers : actionnaires, investisseurs, salariés, journalistes,

Comptes et documents comptables sont présentés au cours de l'assemblée générale ordinaire selon une périodicité réglée :

« Art. L. 225-100 – *L'assemblée générale ordinaire est réunie au moins une fois par an, dans les six mois de la clôture de l'exercice, sous réserve de prolongation de ce délai par décision de justice.*

(L. n° 2001-420 du 15 mai 2001) « *Le conseil d'administration ou le directoire présente à l'assemblée son rapport ainsi que les comptes annuels et, le cas échéant, les comptes consolidés.* » (...). » Ils sont rendus publics dans le mois qui suit l'approbation des comptes par l'assemblée générale par dépôt au greffe du tribunal (art. L. 232-23 du *Code de commerce*).

¹ On peut citer à titre d'exemple la définition proposée par le *Dictionnaire de banque et bourse* : « **Rapport annuel** : composé du rapport de gestion, du rapport financier (comptes) et du rapport des commissaires aux comptes, ce document est publié à la fin de chaque exercice par les sociétés cotées à destination de leurs actionnaires (actuels et potentiels). » Y. Crozet et alii (1993).

² F de Teyssier (1998 : 53-59) et T. Libaert (1998 : 8).

pouvoirs publics, concurrence, mais aussi clientèle. Dans ce dernier cas, il peut s'agir de grandes entreprises, comme de particuliers désireux de se renseigner avant de proposer leur candidature pour un emploi. C'est dire que la première partie des rapports annuels, appelée également « rapport d'activité », est destinée à un large public. Dans la mesure où, dans les rapports annuels, cette partie est la seule à posséder les caractéristiques communicationnelles que l'on souhaite exploiter, le corpus principal a été constitué d'une série de rapports d'activité. Ce découpage reflète par ailleurs l'utilisation des rapports d'activité qui tantôt rassemblent ces deux parties, tantôt les isolent tant elles peuvent fonctionner de manière autonome. Les versions électroniques téléchargeables à partir des sites Internet reprennent ce découpage en proposant fréquemment des fichiers distincts pour chaque partie.

À lui seul, le rapport d'activité constitue souvent le document le plus complet sur l'activité d'une entreprise¹. Il présente, pour chaque année d'exercice, les principaux métiers de l'entreprise, les biens et services qu'elle commercialise, mais aussi sa structure ainsi que les principales caractéristiques de son organisation. L'élaboration de ce rapport s'appuie sur des données fournies par les différentes directions de l'organe central, mais aussi par les filiales ainsi que d'autres instances faisant partie du groupe bancaire. Ainsi, chaque entité fournit-elle les renseignements de base sur le secteur d'activité dont elle a la charge. Ces informations font ensuite l'objet d'une synthèse, souvent élaborée par une agence de communication externe. Cette première version est ensuite validée par chaque émetteur, puis par la Direction générale. Un comité de lecture, issu du Conseil d'administration, ainsi que la cellule de communication financière, harmonisent le texte définitif qui sera présenté lors de l'Assemblée générale.

Outre l'objectif de fabriquer une vitrine de l'entreprise destinée à son environnement économique, ce travail de synthèse poursuit au moins un deuxième but d'ordre communicationnel. En effet, il s'agit aussi de produire un document dans lequel chaque métier, chaque entité puisse se retrouver. Cela se traduit, par exemple,

¹ Cette présentation s'appuie sur le processus réel d'élaboration des rapports d'activité rassemblés dans le corpus. Elle concerne le seul rapport d'activité. Pour ce qui est du rapport financier, il faut préciser que ce ne sont pas les mêmes unités qui collectent, contrôlent et mettent en forme les données financières et comptables. Les deux parties du rapport annuel correspondent ainsi à deux publics, mais aussi à deux circuits d'élaboration distincts au sein de l'organisation.

par le souci qui semble animer le comité de lecture de ne laisser imprimer que les termes anglo-saxons indispensables. En effet, alors que ce groupe bancaire possède par certains aspects une dimension internationale, il n'en demeure pas moins historiquement enraciné dans l'espace territorial français (jusqu'à dans le milieu des années 1980, il y avait autant de caisses régionales de Crédit agricole que de départements). Par souci de ce public interne, dont les préoccupations peuvent être parfois assez éloignées de celles des marchés financiers, la présentation de l'activité bancaire est expurgée d'une bonne part du jargon financier qui est largement dominé par les termes anglo-saxons.

Cette présentation complète de l'activité de l'entreprise fait l'objet d'une validation de la part des instances dirigeantes. En plus de sa complétude, ce document possède donc la propriété d'être un discours autorisé sur les différents aspects de l'activité qu'il présente. On a vu plus haut que cette présentation fait l'objet d'une négociation interne qui se déroule dans les limites fixées par la nécessité de produire un document destiné à véhiculer une (bonne) image, voire à soutenir un cours de bourse. C'est dire que cette complétude doit être comprise dans le cadre d'une communication institutionnelle réalisée dans un contexte d'économie de marché : « *Mais l'entreprise évoluant dans un environnement concurrentiel veillera, simultanément, à préserver la confidentialité de ses opérations dans ce document, public par nature*¹. » Cette présentation de l'activité est donc soumise à deux sortes de restrictions. Au filtrage introduit par le souci de présenter l'aspect positif de certaines réalités au détriment de points maintenus dans l'ombre (donner une bonne image), s'ajoutent les effets d'une autocensure destinée à ne pas livrer à la concurrence des indications susceptibles de divulguer des orientations stratégiques (investissements prévus, prises de participations envisagées, etc.). Ces indications relatives au processus d'élaboration d'un tel discours éclairent de l'intérieur les caractéristiques d'un genre.

¹ A. Guimard (2001 : 57).

B. Le rapport d'activité : un sous-genre de discours économique ?

1. Finalités

Afin de mieux préciser les contours de ce type de discours relativement peu étudié en linguistique de corpus, on peut maintenant chercher à en organiser la présentation sous l'angle des genres de discours. Ces derniers ne sont pas considérés ici comme des patrons prêts à l'emploi, mais plutôt, dans une perspective bakhtinienne¹, comme des productions discursives soumises à un certain nombre de conditions intégrées par le destinataire et le destinataire, et dont dépend la réussite de l'acte de communication. À la suite de D. Maingueneau et de P. Charaudeau², on propose de regrouper ces conditions en fonction de la finalité du discours, de son propos, du statut des partenaires légitimes (énonciateur et destinataire), du lieu et du moment légitimes, du support matériel et de l'organisation textuelle proposée. Une telle grille de lecture est susceptible de s'appliquer à tous les types de textes rassemblés en vue de constituer un corpus. Le niveau de détail fourni, plus important que celui proposé par certaines normes de description telle que le *Dublin Core*³, a néanmoins paru nécessaire pour situer ce type de production par rapport à l'ensemble des discours de l'entreprise, mais aussi par rapport aux corpus de genres discursifs déjà constitués.

Le rapport d'activité, parce qu'il propose une restitution de l'organisation, du fonctionnement et de l'activité économique d'une entreprise, peut s'apparenter à un sous-genre de discours économique. Dans cette catégorie il représenterait un aspect particulier de la communication d'entreprise, appelée communication institutionnelle. De ce point de vue, il peut être opposé à la publicité destinée à faire vendre un produit ou un service. Le rapport d'activité, on l'a vu, est destiné à promouvoir l'image d'une entreprise, c'est-à-dire à présenter sous ses aspects positifs les multiples facettes d'une entité qui n'est pas susceptible de faire l'objet d'une consommation immédiate. Il existe donc bien un rapport d'analogie entre ces

¹ M. Bakhtine (1984).

² D. Maingueneau (2000 : 50 et suiv.) et P. Charaudeau (2001).

³ Les activités et les recommandations de la Dublin Core Metadata Initiative se trouvent à cette adresse : <http://dublincore.org/>.

deux sortes de discours de communication d'entreprise, mais ils ne peuvent pas être confondus.

Avec les restrictions déjà mentionnées plus haut, le rapport d'activité est le support destiné à l'extérieur qui rassemble le plus de données relatives à une entreprise. On peut alors s'interroger sur le statut à accorder à cette diffusion d'informations en direction d'un public externe. Le rapprochement avec les discours de vulgarisation n'est pas incongru, dans la mesure où la communication institutionnelle, comme ces derniers, vise à diffuser des connaissances en direction d'un public de non initiés : « *La vulgarisation scientifique (désormais V.S.) est classiquement considérée comme une activité de diffusion, vers l'extérieur, de connaissances scientifiques déjà produites et circulant à l'intérieur d'une communauté plus restreinte ; cette diffusion se fait hors de l'institution scolaire-universitaire et ne vise pas à former des spécialistes, c'est-à-dire à étendre la communauté d'origine*¹. » Le discours de la communication d'entreprise, comme celui de la vulgarisation scientifique, poursuit un but d'acculturation des lecteurs. La communication opérée en dehors du cercle des initiés produit un discours possédant un statut particulier et que J. Authier propose d'appeler discours second, parce qu'il résulterait de la reformulation d'un discours premier : « *Transmission d'un discours existant en fonction d'un nouveau récepteur, la V.S. se donne donc d'emblée comme **pratique de reformulation** d'un discours source (désormais D1) dans un discours second (D2). Par là, elle s'inscrit dans un ensemble qui comprend traduction, résumé, contraction de texte, et aussi textes pédagogiques adaptés à tel ou tel niveau, analyses politiques reformulées en direction de tel ou tel groupe social, messages publicitaires réécrits en fonction de la cible visée, etc.*² »

Cependant, le parallèle trouve ses limites dans la différence de visée perlocutoire qui existe entre ces deux types de productions discursives. En effet, si la vulgarisation scientifique fait de l'accroissement des connaissances du destinataire sa justification première, il n'en est pas de même pour le rapport d'activité. Ce dernier s'inscrit dans le cadre de relations juridiques et économiques que le

¹ J. Authier (1982 : 34).

² *Ibid.* (1982 : 34-35).

destinateur cherche à entretenir ou à faire naître. Le message produit a donc une portée opérationnelle directe. Certes, ce dont il est question n'est pas réductible à un simple bien de consommation, mais ces informations sont destinées aux calculs de l'*homo œconomicus* ainsi qu'à leur substrat, qui n'est pas toujours rationnel. Pour ce qui est des rapports d'activité, on peut préciser qu'ils visent plus particulièrement à toucher la confiance de l'acteur économique : tout ce qui est bon pour engendrer chez lui un sentiment de sécurité, de gestion rationnelle et de conformité aux attentes du secteur, est mis en avant. On y cherche moins à mettre le lecteur devant ce qu'il ne sait pas encore, qu'à le rassurer sur ce qu'il savait déjà, même très vaguement. Il découle de ces remarques que la finalité du rapport d'activité impose moins à son destinateur de faire œuvre didactique que de rechercher un effet de connivence avec le lecteur. Les informations qu'il propose ne sont donc pas présentées sous une forme très différente de celle qui caractérise la communication interne de l'entreprise. Il s'agit d'une communication économiquement finalisée visant de façon prioritaire l'adhésion du lecteur plutôt que l'accroissement de ses connaissances.

À ce titre, le rapport d'activité représente une forme particulière de diffusion des connaissances. Celle-ci repose sur des implicites considérables, elle table sur un savoir de surface (un « sens commun » économique ?), et la transmission des informations s'y fait à coups de projecteur. Le rapport d'activité n'est ni un récit, ni une étude scientifique partant d'hypothèses, exposant les étapes d'un raisonnement ainsi que la méthode suivie, et discutant les résultats obtenus dans la résolution d'un problème. Il livre un bilan synthétique d'actions et d'évènements révolus, ayant pour cadre l'activité économique d'une entreprise durant une année. Le propos lui-même se prête à une juxtaposition de données chiffrées et d'informations compartimentées. Par ces aspects, ce document illustre également une manière caractéristique de diffuser l'information dans les organisations : au « travail en miettes¹ » correspond l'émiettement de l'information.

¹ G. Friedmann (1956), *Le travail en miettes – spécialisation et loisirs*, Paris, Gallimard, 347 p.

2. Le propos

La présentation de l'activité d'une entreprise peut être l'occasion de livrer des informations de natures assez différentes, même quand on reste dans le domaine de l'économie. Lorsque l'émetteur d'un rapport d'activité traite de son entreprise, en l'occurrence un groupe bancaire qui rassemble plus de trois cents entités, il fait des choix. Ceux-ci ne sont pas seulement destinés à sélectionner ce qui « présente bien » et à éliminer ce qui doit être passé sous silence, mais ils touchent également au niveau de détail, ou plus généralement, à ce qu'il convient de dire dans de telles circonstances. On ne présente pas chaque agence, le détail de chaque produit d'épargne vendu, ou bien encore, le fonctionnement des distributeurs automatiques de billets. L'information fournie dans ces rapports offre une vision globale, des tendances caractéristiques de l'activité, des « faits marquants », pour lesquels certaines indications chiffrées sont communiquées. Celles-ci sont destinées à quantifier l'activité et visent à donner des éléments de comparaison avec les résultats obtenus par d'autres entreprises du même secteur économique. On parle, par exemple de contrats, de portefeuille, de nombre de prêts, mais aussi de produits et de services commercialisés par certaines entités. En revanche, les moyens mis en œuvre pour que ces opérations commerciales puissent avoir lieu sont rarement présentés. L'infrastructure informatique, les procédures comptables, les dispositifs de sécurité ou l'organisation du travail restent dans les coulisses. Néanmoins, on constate que ces autres aspects de l'activité ont tendance à bénéficier d'une meilleure couverture depuis que la thématique du « développement durable » a fait son apparition dans les rapports d'activité.

3. Statut des partenaires légitimes

Le rapport d'activité ne peut pas être produit par n'importe qui : il est l'émanation de la direction d'une entreprise. Il constitue une source d'information primaire susceptible d'être reprise dans la presse économique. Les données figurant dans le rapport financier font l'objet d'un contrôle de la part de commissaires aux comptes et elles engagent la responsabilité des dirigeants de la société.

De même que le contenu du rapport annuel se laisse facilement diviser en un rapport d'activité et un rapport financier, de même le public hétérogène de ce

document se laisse-t-il ranger en deux ensembles relativement distincts. Les analystes financiers, les agences de notation et les journalistes de la presse spécialisée, consultent prioritairement le rapport financier : il s'agit des comptes consolidés, des chiffres clés, des comptes sociaux et de la présentation de l'exposition aux risques¹. En revanche, les actionnaires externes et internes à l'entreprise², les journalistes de la presse économique, les entreprises clientes, les fournisseurs, les représentants des pouvoirs publics (élus nationaux et locaux), et d'une manière générale, tous ceux qui cherchent à s'informer sur l'entreprise ont d'autres attentes. Pour ce public, les éléments les lus appartiennent majoritairement au rapport d'activité : les faits marquants (qui mettent l'accent sur les événements remarquables s'étant produits durant l'année écoulée), les comptes simplifiés, le profil du groupe (fiche d'identité de la société), ainsi que les informations boursières et la présentation des activités³.

Le rapport d'activité cible donc de façon privilégiée un public plutôt généraliste, possédant le vernis économique évoqué plus haut. Afin de s'adapter aux besoins de ce public, le rapport d'activité peut faire l'objet de présentations résumées (« abrégé » résumant la présentation de l'activité et donnant les principaux chiffres, « chiffres clés » se limitant à la présentation de ces derniers). Lorsque le rapport d'activité est diffusé sur Internet, il fait fréquemment l'objet d'un découpage en plusieurs fichiers qui respecte la structure du sommaire. Outre la préoccupation toute matérielle qui peut présider à un tel découpage (le poids des fichiers à télécharger est ainsi plus léger), celui-ci invite l'internaute à considérer que l'information est consommable sous la forme d'une série d'options, à la manière d'un menu de restaurant. Le public du rapport d'activité est alors susceptible de s'élargir à toute personne désireuse de s'informer sur une entreprise à partir de l'Internet. Il se rapproche ainsi de la figure déjà évoquée de l'*homo-œconomicus* : « (...) *sujet abstrait, sans épaisseur sociale (sans pays, sans religion, sans sexe, sans âge, etc.), informé, rationnel (i.e. calculateur) dans sa recherche du maximum de*

¹ Synthèse de l'enquête réalisée sur les rapports annuels sous la responsabilité de N. Coiffait (1999).

² En 2001, on compte plus ou moins deux millions d'actionnaires individuels (parmi lesquels 60 000 salariés du Crédit Agricole sur un effectif global de 130 000 agents environ).

³ *Quel rapport annuel en l'an 2000 ? La Vie française et alii* (1998).

plaisir (maximisation de la production pour le producteur, ou des satisfactions pour le consommateur) pour le minimum de souffrance (minimisation des coûts)¹. » À l'intérieur de l'entreprise, le rapport d'activité joue le rôle d'un discours d'autorité servant de référence lorsqu'il s'agit d'évoquer l'activité de l'entreprise et son organisation à l'occasion d'une communication qui peut être aussi bien externe qu'interne. Dans ce contexte, il est également l'un des rares documents à présenter un bilan à jour sur un grand nombre de sujets. Cela explique que son contenu soit fréquemment repris sous la forme d'extraits sur les sites intranets des directions de l'entreprise.

4. Lieu et moment légitimes

La diffusion de ce document est annuelle et précède l'assemblée générale des actionnaires (celle-ci est également l'occasion de la distribution de ce document), au cours de laquelle le conseil d'administration de la société présente un bilan de l'année écoulée et met au vote un certain nombre de mesures pour l'année en cours. Cette diffusion, calée sur le cycle comptable de l'entreprise, est unique en son genre et se démarque également par cet aspect d'une démarche publicitaire. Elle boucle en quelque sorte une année de communication financière dont les principales étapes se succèdent de cette manière² :

- janvier – février : publication des résultats prévisionnels,
- mars – avril : publication des résultats définitifs, parution du rapport annuel, de la lettre aux actionnaires, réunions d'analystes et d'actionnaires,
- mai – juin : publication du chiffre d'affaires du premier trimestre, assemblée générale des actionnaires,
- juillet – août : publication du chiffre d'affaires du premier semestre,
- septembre – octobre : publication des résultats du premier semestre, d'une nouvelle lettre aux actionnaires, réunions d'analystes
- novembre – décembre : publication du chiffre d'affaires des neuf premiers mois, réunions d'actionnaires, réunions d'analystes.

¹A. Silem, J.-M. Albertini (2004 : 383).

²A. Guimard (2001 : 94-95).

Alors que le rapport annuel ne peut paraître avant la publication des résultats définitifs, passée cette date, sa diffusion peut s'étaler tout au long de l'année. Elle emprunte alors différents canaux, de la remise en main propre lors de l'assemblée générale au téléchargement via Internet, en passant par le pli postal (représentants des pouvoirs publics), la visite d'actionnaires ou l'entretien d'affaires. Les rapports d'activité étant publics, ils peuvent être également demandés directement aux entreprises émettrices. Le support papier contribue à entretenir des liens, ne serait-ce que pour la mise à jour d'un fichier de destinataires, que la diffusion via Internet a tendance à rendre anonymes en même temps qu'elle les fait devenir universels. Comme on l'a déjà indiqué, les rapports d'activité sont également diffusés sur les intranets sous la forme d'extraits ou *in extenso*.

5. Supports matériels

Le texte du rapport annuel du Crédit Agricole, comme celui des autres banques, est en général imprimé luxueusement et se présente sous la forme d'un ouvrage de cent vingt pages environ, à peu près équitablement partagées entre le rapport d'activité et les états financiers. Cette description est valable pour les rapports annuels des années 1990 et l'année 2000. Pour les rapports annuels 2001 et 2002, si le volume du rapport d'activité est comparable à celui des années précédentes, en revanche, celui des états financiers a triplé. Cette augmentation s'explique par l'application des modèles de présentation des états financiers de la COB, absorbée entre-temps par l'AMF (Autorité des marchés financiers)¹. Enfin, le rapport d'activité 2003 a doublé de volume du fait, entre autres, de la présentation

¹ La Commission des Opérations de Bourse (COB) a été créée par une ordonnance du 28 septembre 1967. C'est une autorité administrative indépendante, qui a pour mission de veiller : à la protection de l'épargne, à l'information des investisseurs et au bon fonctionnement des marchés d'instruments financiers. La COB a été intégrée à partir de 2003 à l'AMF qui regroupe la COB, le conseil des marchés financiers (CMF) et le Conseil de discipline de la gestion financière (CDGF). L'AMF est une autorité publique indépendante ayant en charge la réglementation des marchés financiers. Le *Guide d'information - Le document de référence* (COB, mai 2002, 17 pages) donne les indications suivantes : « *Le document de référence contient l'ensemble des informations juridiques, économiques et comptables concourant à une présentation exhaustive d'une société pour un exercice donné. Sa structure est fondée sur les schémas d'information requis par la réglementation européenne (directive 80/390/CE) et les règlements français qui en découlent (règlement COB n°98-01 pour les Premier et Second marchés et règlement COB n°95-01 pour le Nouveau marché).* »

des activités d'un groupe bancaire important ayant fait l'objet d'une acquisition¹. L'édition principale du rapport d'activité est présentée sous des formes abrégées. Toutes ces versions existent en français et en anglais, les versions abrégées étant également traduites en espagnol et en italien². La version électronique diffusée sur Internet est conforme à l'original, au détail près, évoqué plus haut, que la version complète est fréquemment découpée en fichiers de plus petite taille correspondant à ses principales parties.

6. Organisation textuelle

L'organisation textuelle de ce document peut varier selon que l'entreprise est cotée en bourse ou non. Dans le premier cas, les entreprises sont vivement encouragées à suivre un plan respectant les normes européennes (on parle alors de « document de référence »). Dans le second, le plan peut être plus lâche, mais il faut distinguer ici deux niveaux. On remarque qu'en l'absence de norme explicite, les rapports annuels produits dans un même secteur économique contiennent tous à peu près les mêmes sortes d'informations, regroupées dans le rapport d'activité et le rapport financier. On trouve là une confirmation de l'existence d'un genre, chaque entreprise imitant ouvertement les innovations présentes dans les rapports des concurrentes. L'intervention de cabinets de conseil en communication financière et institutionnelle contribue également à homogénéiser ces documents.

En ce qui concerne le rapport d'activité, l'organisation des rubriques du sommaire peut néanmoins connaître des changements d'une année sur l'autre. Seules quelques rubriques sont relativement récurrentes, même si elles ne sont pas nécessairement présentées dans le même ordre. Alors que le message du président, le profil de l'entreprise et les chiffres clés figurent fréquemment en début de rapport, plusieurs autres rubriques importantes peuvent voir leur ordre modifié par bloc ou de

¹ Il s'agit du Crédit Lyonnais acquis par Crédit Agricole S.A. à la suite d'une OPA amicale ayant eu lieu en décembre 2002.

² Le rapport annuel est imprimé à 20 000 exemplaires en moyenne pour la version complète en français et à peu près autant pour la version anglaise. Les rapports abrégés (français, anglais, espagnol et italien) sont tirés à 15 000 exemplaires environ. Les indications sur la diffusion des versions électroniques ne sont pas assez précises ou peu fiables. En revanche, il n'est pas exagéré de penser que la diffusion totale de ce document dans ses différentes versions passe les 100 000 exemplaires.

façon individuelle. Le rapport annuel peut également être scindé en deux, le rapport d'activité étant diffusé parallèlement au rapport financier. Tout se passe comme si les blocs d'informations regroupés par les rubriques étaient dotés d'une certaine autonomie autorisant permutations, ajouts, suppressions au gré des usages (version longue, version abrégée) et des périodes (on constate une tendance à l'augmentation du nombre de rubriques). Parmi cette diversité d'usages, il est néanmoins possible de repérer deux dispositifs spécifiques, selon que les informations sont présentées sous la forme d'énoncés suivis respectant un ordre linéaire, ou qu'elles sont regroupées au sein de tableaux, d'organigrammes, d'histogrammes ou d'autres formes de listes. Ces points relatifs à l'organisation textuelle font l'objet d'une présentation détaillée dans la section suivante.

III. LE DISPOSITIF TEXTUEL DES RAPPORTS D'ACTIVITÉ

A. Les rapports d'activité sont composés d'éléments hétéroclites

1. Identification de trois sortes de contextes

Lorsque l'on feuillette un rapport d'activité récent, comme ceux dont se compose le corpus, on ne peut manquer d'être frappé par le fait qu'il s'agit d'un document qui offre plusieurs dispositifs de présentation de l'information. En effet, le texte est complété par des photographies qui occupent une place importante, de même que les listes, les schémas ou les tableaux. Un simple feuilletage permet également de découvrir que les rapports d'activité sont divisés en rubriques, dont les principales sont mentionnées dans le sommaire de l'ouvrage. La diversité des modes de présentation se double donc d'une segmentation du texte qu'il conviendra de présenter puis d'interroger.

La sélection d'une série de rapports d'activité a été réalisée en fonction des objectifs de documentation d'activités d'information et de constitution d'un référentiel terminologique adapté au contexte. Il a néanmoins paru utile de situer cet échantillon au sein du genre de discours auquel il appartient, et dont les textes du

corpus ne présentent qu'une évolution récente¹. En effet, celle-ci n'est pas toujours de nature à bien faire comprendre la logique qui anime l'évolution d'un tel genre de discours. L'ancienneté du genre « rapport d'activité » invite à prendre un petit peu de recul, afin d'observer la manière dont la restitution des informations relatives à l'activité ont pu évoluer dans un tel cadre.

Si l'on s'en tient aux textes relatifs à l'activité du Crédit agricole, on constate que, depuis la fin de la seconde guerre mondiale, la différenciation des modes de présentation de l'information n'a fait que s'amplifier. Le texte austère du *Rapport sur la situation du Crédit agricole mutuel au 31 décembre 1944* est déjà illustré de schémas dessinés à la main. Dix ans plus tard, deux couleurs d'encre sont utilisées pour imprimer textes et graphiques. Le papier glacé apparaît au début des années 1960, avec la quadrichromie, et une première photographie « pleine page » orne le rapport annuel de 1967. Avec celui de 1968, le changement de maquette se traduit par l'utilisation de deux couleurs de papier, des polices de caractères très frappantes, des pictogrammes et des schémas en couleur. À partir du début des années 1970, la photographie devient de plus en plus présente, et l'on peut considérer que les principaux dispositifs utilisés dans les rapports d'activité actuels ont été mis en place à cette époque.

Les textes des rapports d'activité comportent trois sortes de contextes qui peuvent être distingués : le premier rassemble les éléments textuels composés d'énoncés suivis. La position dominante de cette composante a été progressivement remise en cause par l'apparition de deux autres sortes de contextes. Ceux-ci regroupent, d'une part, un ensemble lui-même hétérogène, composé de graphiques, de schémas, de tableaux et de listes, et d'autre part, des photographies. Si celles-ci font parfois écho au texte, lorsqu'elles représentent des salariés ou des installations

¹ Parmi les productions discursives d'une entreprise, les rapports d'activité bénéficient d'une histoire relativement facile à reconstituer. En effet, leur statut légal, et leur rôle de texte d'autorité ont généralement favorisé leur conservation. Ce n'est évidemment pas le cas de toute la littérature grise produite par les entreprises, ni a fortiori, des divers documents liés encore plus étroitement à la réalisation de l'activité, et dont une partie a été citée à titre d'exemple dans le premier chapitre. Cependant, si les écrits d'entreprises ne sont pas tous logés à la même enseigne pour ce qui est de leur histoire et de leur description, il a néanmoins paru nécessaire d'aborder cette question dès lors que l'on s'est fixé pour contrainte de documenter un corpus de manière à faciliter l'intelligence des textes qu'il rassemble.

de l'entreprise, elles sont en général déconnectées des réalités évoquées dans les rapports d'activité. La répartition de ces trois ingrédients est relativement stable pour la période couverte par le corpus, qui va de 1995 à 2003. En effet, si l'on compare la surface occupée par chaque type de contexte, on retrouve un dosage très proche aux deux extrémités de la période :

*Tableau 3.
Répartition des trois sortes de contextes dans les rapports d'activité du corpus en début et fin de période*

Année	Texte	Graphiques et tableaux	Photographies
1995 (pour 64 pages)	53,1%	24,6%	22,3%
2003 (pour 101 pages)	54,2%	26,7%	19,1%

Dans la mesure où l'on cherche d'abord à étudier le fonctionnement d'unités langagières, on éliminera du corpus tout l'appareil des photographies, dont on conservera néanmoins les légendes (il s'agit principalement des photographies du président et du directeur général, qui accompagnent traditionnellement le message introduisant les rapports d'activité). Si ce choix s'explique assez naturellement, en revanche, la question de retenir ou d'éliminer les contextes composés de tableaux, de graphiques et autres schémas, n'appelle pas de réponse évidente. Ce point est rarement abordé lors de la présentation des corpus, tant il semble aller de soi que le texte linéaire constitue la norme, qu'il soit structuré ou non à l'aide de balises XML ou autres. Compte tenu des objectifs poursuivis dans ce travail, il a semblé utile de conserver ces contextes particuliers. En effet, ils constituent une bonne illustration de la manière souvent très laconique qui est adoptée pour diffuser l'information dans les organisations, et en particulier sur leurs intranets. On dispose ainsi d'échantillons d'un contexte assez particulier, mais représentatif, pour observer le fonctionnement de certaines unités et, entre autres, les incidences d'un changement de dispositif à l'intérieur d'un même *medium*.

2. Contextes à dominante syntactique et non syntactique

Dès lors que l'on a opté pour la conservation de ces sortes de contextes, il a fallu décider de la manière de les intégrer aux zones de texte suivi. Or, ce que l'on

peut constater c'est que, globalement et de manière « évidente » pour les premières parties dont le corpus est composé, le changement de contexte suit le découpage des textes en rubriques. Il convenait alors de se donner les moyens de caractériser les deux formes de contextes conservés, de manière à différencier ceux qui relevaient plutôt du texte suivi, et les autres, composés de tableaux, schémas et autres listes. Si ce découpage se vérifie dans les grandes lignes, il reste que la mise en page adoptée pour les rapports d'activité est complexe, de telle sorte que l'on ne peut pas dire qu'il existe des contextes « purs » dans le sens où ils relèveraient soit du texte suivi, soit des graphiques. On donne ci-dessous un exemple de mise en page complexe observée surtout dans la dernière période couverte par le corpus de rapports d'activité.

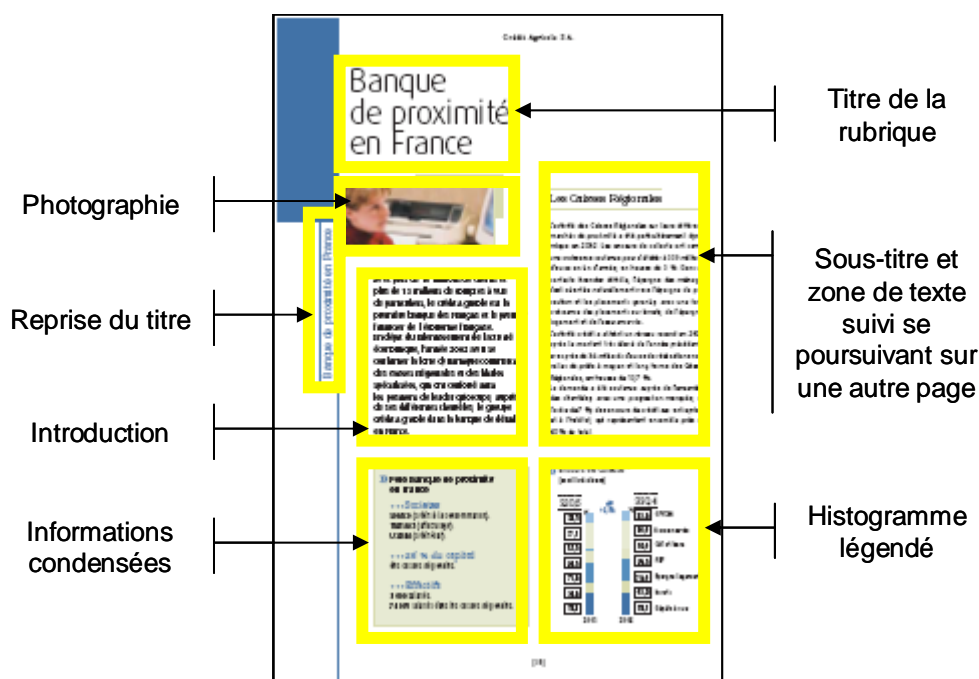


Figure 8.

Découpage en zones de la page 18 du rapport d'activité du Crédit Agricole de 2002

Dans ces conditions, la caractérisation des rubriques doit être conçue en termes de tendance dominante, chaque page pouvant être plus ou moins représentative. Pour caractériser les rubriques du corpus, on a choisi d'emprunter à

Jack Goody¹ les notions de contexte à dominante syntactique ou non syntactique. Dans le domaine de l'anthropologie, les travaux de cet auteur ont mis en lumière les relations existant entre certaines formes de pensée et l'usage de l'écriture. Les listes (livres de comptes, inventaires, listes lexicales, etc.), les tableaux et les matrices constituent des formes d'écriture non syntactiques : « *Cette sorte d'agencements spatiaux où les termes linguistiques sont abstraits de la phrase (sortis de leur contexte) se rencontre de bonne heure dans l'histoire de l'écriture, et en fait domine sa production. En sorte que ce qui est significatif dans l'emploi de la langue dans les premiers systèmes d'écriture est qu'une grande partie présente une structure syntaxique très différente du discours parlé*². » La partition caractéristique des textes des rapports d'activité entre ces deux types de dispositifs textuels a été reprise lorsqu'il s'est agi de baliser le corpus. Celui-ci a été découpé selon des contextes de type documentaire, c'est-à-dire « (...) *définis à partir d'une unité textuelle (paragraphe, partie, article, chapitre, document...)*³», avec une répartition à grosses mailles des rubriques, entre contextes à dominante syntactique et non syntactique. On donne ci-dessous un exemple de page de rubrique caractérisée comme étant non syntactique :

¹ J. Goody (1986 : 65).

² J. Goody (1994 : 280).

³ B. Habert, A. Nazarenko, A. Salem (1997 : 179).

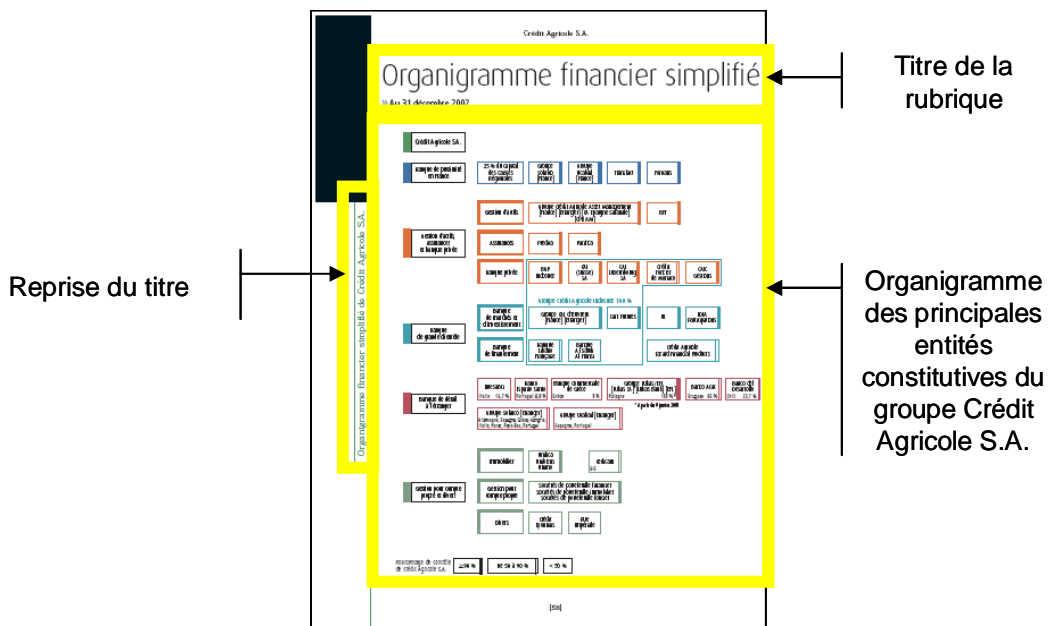


Figure 9.
 Découpage en zones de la page 58 du rapport d'activité du Crédit Agricole 2002

On peut compléter cette présentation à l'aide d'autres exemples de contextes caractéristiques rencontrés dans des rubriques à dominante non syntactique :

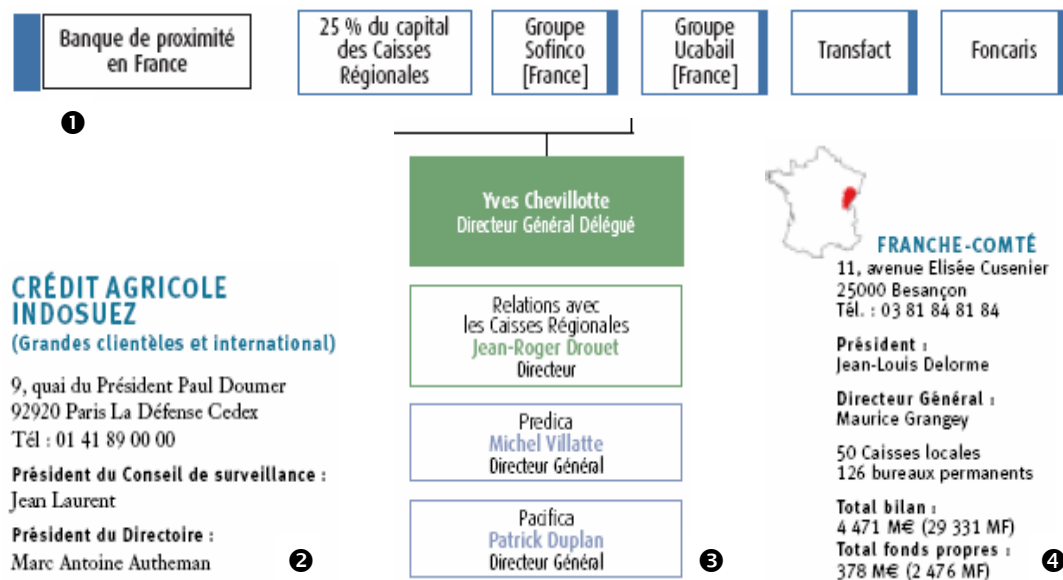


Figure 10.

Exemples d'éléments caractéristiques des rubriques non syntactiques :
1) organigramme financier ; 2) coordonnées de filiale ; 3) organigramme de direction ; 4) notice de caisse régionale¹

On a vu précédemment qu'il est nécessaire de distinguer les contextes de façon relative, c'est pourquoi on a préféré parler de dominante syntactique ou non syntactique. La portée de la distinction n'en est pas réduite pour autant, car elle se double d'une répartition des rôles entre contextes. On reviendra plus en détail sur ce point, mais on peut d'ores et déjà noter que le texte suivi des CDS (contextes à dominante syntactique) est plutôt utilisé pour la présentation détaillée de l'activité, alors que les CDN (contextes à dominante non syntactique) offrent majoritairement des données relatives aux structures de l'organisation et aux différents postes en fonction desquels cette activité est quantifiée. Dans le premier cas, il s'agit d'organigrammes hiérarchiques et financiers ou de listes d'entités avec leurs

¹ Les illustrations 1 et 3 sont extraites du *Rapport annuel du Crédit Agricole 2002* (pp. 12 et 58), les illustrations 2 et 4 sont tirées du *Rapport annuel du Crédit Agricole 2000* (p. 22 et 17).

coordonnées. Ces types de contextes favorisent la sélection des éléments présentés : ils ne doivent pas être en trop grand nombre, sinon cela nuirait à la clarté recherchée, mais ils doivent entretenir un rapport essentiel avec l'activité décrite par ailleurs, sinon leur présence perdrait en pertinence.

La nature du rapport que les données rassemblées dans les CDN entretiennent avec l'activité présentée dans les CDS est double : d'une part, il s'agit de quantifier ou de récapituler des chiffres fournis au fil du texte. Ces derniers sont rapportés à l'axe temporel (il s'agit toujours de chiffres pour une période donnée, et l'année du rapport d'activité y est mise en avant). D'autre part, il s'agit d'enraciner les entités mentionnées dans un référentiel spatiotemporel censé être partagé avec le destinataire : coordonnées des entreprises, noms des personnes physiques participant aux décisions prises dans le cadre de l'activité de l'entreprise, et/ou faisant partie d'instances nécessaires à l'organisation de cette dernière¹. La distinction des deux contextes sera reprise dans les deuxième et troisième parties, lorsque l'on étudiera plus finement le fonctionnement du vocabulaire des rapports d'activité.

B. Le découpage d'un monde

1. Les rubriques de sommaire : entre histoire du genre et histoire de l'institution

Si, comme dans la première section, on remonte une cinquantaine d'années en arrière, de manière à mieux cerner la sorte d'évolution dans laquelle se trouvent pris les rapports d'activité dont le corpus est composé, plusieurs constats peuvent être dégagés. Le premier porte sur l'augmentation progressive du nombre de rubriques² : on passe ainsi de sept rubriques mentionnées dans le sommaire de 1944 (deux pour le « rapport d'activité » et trois pour le « rapport financier », qui ne sont pas ainsi dénommés à cette époque), à une trentaine de rubriques figurant dans les sommaires du rapport d'activité 2003 (24 rubriques) et du rapport financier 2003 (5

¹ Il va sans dire que la plupart de ces noms sont essentiellement connus dans le contexte de d'activité du Crédit Agricole, et qu'il ne s'agit pas de personnes célèbres ou dont les noms auraient été popularisés par l'actualité médiatique du moment.

² À ce stade de l'exposé, on reprend la notion de rubrique dans son acception ordinaire. Celle-ci fonctionne comme titre, dont on supposera pour le moment qu'il correspond à un titre repris dans le corps du texte, et comme texte correspondant à un sujet déterminé repris régulièrement dans les rapports annuels.

rubriques). En second lieu, on devine que cette inflation de rubriques a deux causes principales. En effet, tout se passe comme si le nombre de rubriques avait augmenté parallèlement à l'extension des compétences bancaires du Crédit agricole. D'autre part, l'apparition de certaines rubriques ne pouvant pas s'expliquer de manière satisfaisante par cette première cause, il est nécessaire de trouver une autre explication qui pourrait provenir des usages communicatifs qui sont faits de ce document.

Le sommaire du *Rapport* de 1944 comprend les rubriques suivantes :

« <i>Exposé général</i> »	p. 1-7
<i>Ressources du CAM</i>	p. 8-12
<i>Opérations</i>	p. 13-25
1° <i>Opérations de la Caisse nationale de CAM</i> ¹	p.13-19
2° <i>Opérations des Caisses régionales de CAM</i>	p. 20-23

Bilan de la Caisse nationale de crédit agricole

Bilan résumé des Caisses régionales de CAM »

Quatre ensembles peuvent être dégagés dans ce sommaire : le premier correspond à l'« exposé général », le second aux « ressources » servant à financer le troisième, les « opérations » (ou crédits) ; enfin, le quatrième ensemble permet de rassembler les bilans. L'activité du Crédit agricole se cantonne alors au financement de l'activité agricole. Ce noyau est complété dès 1946 par des annexes présentant les noms du directeur général et du contrôleur général, et surtout, les coordonnées de la Caisse nationale et des Caisses régionales de Crédit Agricole Mutuel. Or, depuis 1945, le Crédit Agricole ne bénéficie plus de la redevance de la Banque de France ni des avances de la Caisse des Dépôts et Consignations². Afin de financer les crédits qu'il accorde, il doit s'organiser pour collecter de l'épargne dans le milieu rural. Par ailleurs, est créée à la même époque une association³ destinée à défendre les intérêts du Crédit Agricole auprès des pouvoirs publics⁴. L'apparition de ces annexes n'est

¹ Crédit agricole mutuel.

² J.-P. Henry, M. Régulier (1986 : 66 et suiv.).

³ Il s'agit de la création de la Fédération Nationale de Crédit Agricole, le 17 mai 1945.

⁴ A. Gueslin (1985 : 21 et suiv.).

donc probablement pas étrangère aux besoins de contacts nés des activités de lobbying et de collecte de l'épargne dans l'immédiat après-guerre.

Au cours des années 1960 à 1980, le Crédit Agricole acquiert progressivement auprès des pouvoirs publics le droit de prêter à des clientèles de plus en plus diversifiées. Par exemple, en 1971, il obtient de pouvoir prêter aux collectivités publiques, aux particuliers et aux professionnels établis en milieu rural, en plus des agriculteurs. Cette ouverture sur l'ensemble du monde rural a des répercussions importantes sur la présentation des rapports annuels et du rapport d'activité de 1972. C'est à cette époque que la photographie est mise au service de l'exposition de cette ruralité, du moins telle que la conçoit le Crédit Agricole de l'époque. Le sommaire connaît un bouleversement sensible¹ :

<i>« Les conditions de l'activité du Crédit Agricole</i>	<i>11</i>
<i>L'origine et la destination des fonds du CA</i>	<i>21</i>
<i>Le CA au service de l'agriculture</i>	<i>25</i>
<i>Le CA au service des ménages</i>	<i>31</i>
<i>Le CA au service du développement régional</i>	<i>35</i>
<i>Le CA au service des épargnants</i>	<i>39</i>
<i>Les moyens d'action</i>	<i>45</i>
<i>Bilans</i>	<i>51</i>
<i>Les Filiales du CA</i>	<i>59</i>
<i>Les Caisses Régionales de CAM</i>	<i>63 »</i>

Les « conditions de l'activité » reprennent « l'exposé général » de 1946 ; les ressources et les opérations sont devenues « L'origine et la destination des fonds du CA ». Or, ce sommaire comporte des transformations plus importantes qui reflètent l'évolution de l'institution et de la société de l'époque. Les « moyens d'actions » sont apparus comme entrée de sommaire en 1968, sous la dénomination « La vie de l'institution ». Les filiales, qui supportent le développement de l'activité du Crédit Agricole sur de nouveaux métiers, ont fait leur apparition au début des années 1970. L'ouverture, encore timide, à de nouvelles clientèles, est néanmoins présente dans le

¹ Les annexes apparues en 1946 se sont étoffées et figurent en début de rapport ; elles comprennent la liste des membres de la Commission plénière du Crédit Agricole, du conseil d'administration et de la direction générale.

sommaire (agriculteurs, ménages, développement régional – pour les professionnels et les collectivités publiques –, ainsi que les épargnants). Cependant, ce n'est qu'en 1991 qu'un décret supprime toute limitation à l'exercice de l'activité bancaire. Le Crédit Agricole devient alors une banque universelle, c'est-à-dire qu'il peut accorder des crédits aussi bien aux agriculteurs qu'aux particuliers, aux commerçants, aux professionnels, aux collectivités locales ou aux entreprises. La présentation de l'activité, qui était jusqu'alors plutôt organisée autour de la collecte des ressources et de la distribution des crédits, va évoluer vers une présentation des différents marchés sur lesquels le Crédit Agricole intervient. On adoptera donc le sommaire de l'année 1991 pour référence, lorsqu'il s'agira de commenter les évolutions les plus récentes dont les sommaires des rapports d'activité du corpus pourraient se faire l'écho :

SOMMAIRE			
Chiffres clés	1	<i>Les professionnels :</i>	
Message du Président et du Directeur général	2	maîtrise et innovation	18
<i>Analyse financière des résultats :</i>		<i>Les capitaux :</i>	
la régularité dans la performance commerciale et financière	4	un positionnement solide	20
<i>Les particuliers :</i>		<i>L'international :</i>	
une écoute attentive, des solutions adaptées	8	direction l'Europe	22
<i>Les agriculteurs :</i>		La mobilisation des ressources humaines	24
une relation naturelle	12	Les regroupements de Caisses régionales	26
<i>Les collectivités locales :</i>		La présentation des Caisses régionales	28
un marché en devenir	14	Le Conseil d'administration de la CNCA	38
<i>Les entreprises :</i>		L'organigramme de la CNCA	39
la montée en puissance	15	Rapports financiers consolidés du Crédit agricole	41

*Figure 11.
Reproduction du sommaire du rapport annuel du Crédit Agricole de 1991*

Si ce bref rappel a permis de mettre en lumière l'existence de correspondances entre les évolutions de l'institution et les rubriques des rapports d'activité, il ne s'agit pas d'une relation mécanique. En effet, le sommaire doit être également considéré comme une rubrique à part entière du support de communication institutionnelle qu'est le rapport d'activité. À ce titre, il hérite de certaines propriétés qu'il convient de présenter maintenant.

2. Le sommaire : une rubrique comme les autres

a. Les présentoirs de la vitrine

Dans le sommaire du *Rapport annuel* de 1991 que l'on a reproduit au-dessus, le rapport d'activité s'étend de la première (Chiffres clés) à la quarantième page, après l'organigramme de la CNCA. On mesure mieux le développement pris par cette présentation de l'activité si on la rapporte à ce qui en constitue l'embryon : les sept pages de l'« Exposé général » du rapport de 1944. Mais cette évolution impressionnante appelle d'autres remarques.

Alors que l'ouverture de la banque à toutes les sortes de clientèles semble permettre de fixer la structure principale du sommaire autour de l'énumération des différents marchés, on constate l'apparition de deux nouveautés dans ce secteur. En effet, la rubrique « L'international » s'est ajoutée à la liste, de même que « Les capitaux ». Ces rubriques ont fait leur première apparition dans les sommaires, pour la première, en 1977, et pour la seconde, en 1986. On a vu précédemment que les sommaires sont susceptibles de capter avec relativement peu de décalage ce type d'évolutions. Car, même si le réseau international du Crédit Agricole est alors balbutiant (la première succursale est implantée à Chicago en 1979), le souci de montrer que l'institution est également présente dans ce domaine suffit à constituer une entrée de sommaire. Celui-ci fonctionne donc bien comme un reflet, non pas tant du fonctionnement réel de l'institution, que de la manière dont l'institution souhaite parler d'elle. On dispose là d'une indication précieuse sur le fonctionnement particulier du discours de communication institutionnelle comme « vitrine » de l'activité. On peut considérer que l'entrée intitulée « Les regroupements de Caisses régionales » obéit à la même intention. En effet, le projet de Groupe de 1988 prévoit que ces caisses régionales, alors au nombre de quatre-vingt-quatorze, doivent progressivement se regrouper. La rubrique apparaît en 1990, alors que les premiers regroupements peuvent être signalés. Il s'agit alors d'insister auprès des lecteurs sur le fait que le Crédit Agricole est en train de moderniser ses structures. Le sommaire est alors utilisé comme un présentoir destiné à mettre en exergue ce changement majeur aux yeux de l'émetteur.

b. Effets de genre

La présence du « Message du Président et du Directeur général » vient pour sa part mettre indirectement l'accent sur le changement de statut de l'émetteur. En effet, la Caisse Nationale de Crédit Agricole (CNCA), qui était autrefois un

établissement public soumis à la tutelle de l'Etat, est devenue en 1988 une société anonyme dont le capital est détenu par les Caisses régionales de Crédit Agricole. Ce message inaugural correspond à une déclaration solennelle adressée aux actionnaires par le président du conseil d'administration et le directeur de la CNCA. Ces derniers y expriment de façon synthétique les caractéristiques de l'année écoulée et esquissent, en des termes généralement vagues, les orientations retenues pour l'avenir. Le sommaire permet donc d'enregistrer également des éléments qui semblent plutôt relever de ce que l'on pourrait appeler le bon usage du genre : en vertu de son nouveau statut, le Crédit Agricole se doit de faire figurer dans son rapport d'activité ce type de message.

C'est la même raison qui préside très probablement à l'apparition dans le sommaire de 1988 des « Chiffres clés ». En effet, cette rubrique reprend sous la forme de tableaux ou d'histogrammes des indicateurs généraux tels que le produit net bancaire, le total du bilan, les encours de collecte et les crédits, etc. Il s'agit alors de donner aux publics du rapport d'activité un aperçu rapide sur les principaux indicateurs permettant de caractériser l'activité bancaire de l'année écoulée. Là aussi, le genre « rapport d'activité de société anonyme » semble avoir fonctionné comme modèle, au moins en ce qui concerne la mention dans le sommaire de l'une de ses rubriques « attendues ». Mais, contrairement à l'exemple précédent, les chiffres clés existaient déjà dans les rapports antérieurs à 1988, où ils étaient en général présentés dès les premières pages, avant même le sommaire. C'est là une dernière caractéristique du fonctionnement des rapports d'activité sur laquelle on voudrait attirer l'attention : le fait que leur structure modulaire permet à l'émetteur de modifier la mise en valeur et l'ordonnancement de matériaux, somme toute relativement stables.

c. Effets kaléidoscopiques

On aura peut-être noté que le sommaire du rapport de 1972 mentionne une rubrique consacrée aux filiales qui n'est pas reprise dans le sommaire de 1991. Est-ce à dire que les filiales, dont on a dit plus haut que leur rôle consistait à assurer le développement de l'institution dans de nouveaux secteurs d'activité, ont disparu ? En réalité, la présentation de leur activité a été répartie en fonction des différents marchés sur lesquels elles sont amenées à intervenir. Car, comme cela a été indiqué

plus haut, ce qui importe à ce moment-là, c'est de mettre en avant le fait qu'en matière de distribution de crédit, plus rien ne distingue le Crédit Agricole des autres grandes banques françaises à réseau de l'époque (BNP, Société Générale et Crédit Lyonnais). Le matériau « activité des filiales » n'a donc pas disparu des rapports d'activité, mais sa présentation a fait l'objet d'une organisation nouvelle. Dans la période couverte par le corpus (1995 à 2003), on assiste à l'apparition de nouvelles entrées de sommaire consacrées aux filiales à partir de 1999.

3. Les rubriques : un premier niveau d'interaction entre la restitution de l'activité, l'évolution d'un référentiel et les textes du corpus

a. Principales évolutions des rubriques de sommaire des textes du corpus

En tenant compte des remarques présentées dans les deux sections précédentes, on peut tirer de l'étude des rubriques de sommaire un premier enseignement relatif à la manière dont s'articulent restitution de l'activité et organisation des textes rassemblés dans le corpus. On a vu précédemment que le rapport d'activité est découpé en rubriques dont le sommaire propose une vue d'ensemble. Cette présentation est sensible aux évolutions majeures qui affectent l'institution, mais on a vu qu'elle est également influencée par des phénomènes qui tiennent plus aux caractéristiques communicationnelles du genre : mise en valeur temporaire de certains aspects du fonctionnement de l'institution, prise en compte de ce que l'on a appelé les effets de genre (le fait de présenter les informations attendues dans ce type de document dès le sommaire), enfin, utilisation du sommaire comme dispositif de surface, destiné à mettre en avant de nouveaux agencements concernant des sortes d'informations déjà présentes dans les rapports précédents.

Afin de rendre compte de ces évolutions, on a repris des intitulés génériques de rubriques, en ce sens qu'ils ne tiennent pas compte des variantes lorsqu'elles concernent le titre sans modifier notablement le contenu de la rubrique, que l'on a ventilé sur les neuf parties que compte le corpus. On a commencé par la présentation des rubriques les plus récurrentes à l'échelle du corpus, pour présenter dans un deuxième temps celles qui sont apparues progressivement.

*Tableau 4.
Rubriques les plus partagées par les sommaires des rapports d'activité du corpus
(chaque rapport est signalé par une année)*

Rubriques	Années								
	95	96	97	98	99	00	01	02	03
Sommaire	*	*	*	*	*	*	*	*	*
Profil	*	*	*	*	*	*	*	*	*
Message du Président et du Directeur Général	*	*	*	*	*	*	*	*	*
Un Groupe uni et décentralisé	*	*	*	*	*	*	*		
Chiffres clés	*	*	*	*	*	*	*	*	*
Le Crédit agricole en [telle année]	*	*	*	*	*	*			
Les particuliers	*	*	*	*	*	*			
L'agriculture	*	*	*	*	*	*			
Les collectivités locales et des professionnels	*	*	*	*	*	*			
Le marché des entreprises	*	*	*	*	*	*			
Les activités internationales	*	*	*	*	*	*			
Les capitaux	*	*							
Carte des Caisses régionales de Crédit agricole	*	*	*	*	*	*			
Les Caisses régionales	*	*	*	*	*	*	*		*
Le Conseil d'administration de la CNCA [puis de CA s.a.]	*	*	*	*	*	*	*	*	*
L'organigramme de la CNCA [puis de CA s.a.]	*	*	*	*	*	*	*	*	*

Le tableau présenté ci-dessus a été constitué de la manière suivante : la colonne de gauche comporte le nom de la rubrique avec la dénomination correspondant à sa première mention dans le sommaire de l'un des rapports d'activité du corpus (la période couverte allant de 1995 à 2003). Chaque fois que la rubrique apparaît dans un autre rapport d'activité, sa présence est indiquée à l'aide d'un astérisque. Comme cela a été indiqué plus haut, on n'a pas tenu compte dans ce tableau simplifié des variations de dénomination ; pour une rubrique similaire retrouvée dans plusieurs rapports d'activité, on a conservé la dénomination de la première mention. Enfin, les rubriques de ce tableau correspondent à des pages dans le corps du texte, y compris lorsque celles-ci se limitent à un titre. En revanche, les méta-rubriques qui ont une fonction de regroupement au niveau du sommaire, mais

qui ne correspondent pas à une page titrée dans le corps du texte, n'ont pas été reprises. Par exemple, dans le sommaire du rapport de 1991, les principaux marchés sont mis sur le même plan que les autres rubriques, alors que, dans le sommaire de 1995, ils sont chapeautés par le titre « Les marchés ». Dans la mesure où il ne correspond pas à une page, ce « titre de sommaire » n'est pas repris dans les tableaux.

En dehors du « Profil » et de la rubrique « Un groupe uni et décentralisé¹ », on retrouve dans le sommaire de 1995 le « Message », les « Chiffres clés », l'organigramme et la présentation du conseil d'administration, déjà présents dans le sommaire de 1991. La présentation générale de l'activité (ici « Le Crédit Agricole en telle année »), les cinq principaux marchés et l'international sont présents jusqu'en 2000. Les notices descriptives des Caisses régionales connaissent une éclipse en 2002, et à partir de cette même année, la présentation de l'organisation du Crédit Agricole disparaît du sommaire. Enfin, on remarque que la rubrique « Les capitaux » disparaît très tôt du sommaire, soit dès 1997.

On recherchera les raisons de ces éclipses soudaines en les mettant en relation avec l'apparition concomitante de nouvelles rubriques. L'intitulé de ces dernières permettra en effet d'établir d'éventuelles correspondances avec l'évolution de l'activité de l'institution. Dans le cas contraire, c'est le fonctionnement du sommaire en tant que dispositif au service d'une visée communicationnelle particulière qui sera interrogé. Dans la mesure où l'on dispose d'une série chronologique, ces modifications seront observées dans leur ordre d'apparition.

La première innovation remarquable intervient en 1997, avec la disparition de l'entrée concernant les « capitaux ». L'évolution des rubriques de sommaire du rapport de 1997 se trouve éclairée si l'on tient compte du fait que l'année 1996 a été marquée par l'acquisition de la banque Indosuez (il s'agit d'une banque d'affaires française disposant d'un réseau dans le monde entier).

¹ Ces deux rubriques sont néanmoins présentes dans le rapport d'activité de 1991, mais elles sont données hors pagination, dans les toutes premières pages précédant le sommaire. La présentation de l'organisation du Crédit Agricole est mentionnée en 1992 et le « Profil » apparaît dans le sommaire de 1993.

*Tableau 5.
Rubriques apparues en 1997 et leur reprise dans les sommaires postérieurs*

Rubriques	95	96	97	98	99	00	01	02	03
Organigramme du Crédit Agricole			*	*	*	*	*		
La banque de proximité			*	*	*	*	*	*	*
Le Crédit agricole dans le monde			*	*	*	*			
La gestion d'actifs			*	*	*	*	*	*	*
Les implantations du Crédit agricole dans le monde			*				*		

L'apparition de l'entrée « banque de proximité » – cette notion remonte au moins au début des années 1990 –, ne se comprend bien que si on l'oppose à l'apparition concomitante de rubriques mettant en avant le renforcement de la dimension internationale du Crédit Agricole : le « Crédit agricole dans le monde » et ses différentes implantations. En effet, cette rubrique vient chapeauter les cinq marchés qui structuraient le traitement de l'activité dans les précédents sommaires (en 1995 et 1996 une rubrique ne correspondant à aucune page dans le texte regroupait les « Marchés »). Depuis 1991 les « capitaux » (pour les activités ayant pour cadre les marchés de capitaux), deviennent la « gestion d'actifs », traduction française approximative¹ d'« *asset management* ». Et ce sont les changements d'organisation occasionnés par le rachat de la banque d'affaires qui, semble-t-il, sont mis en avant par le signalement de l'organigramme du Crédit Agricole, alors que ce type de présentation existait déjà dans le corps du texte des rapports antérieurs. Les rubriques qui apparaissent durant la période couverte par les années 1998 à 2000 viennent en quelque sorte étayer le dispositif destiné à mettre en valeur la nouvelle dimension internationale de la banque.

¹ Ce point est repris dans les chapitres 5 et 6, consacrés à la terminologie bancaire.

*Tableau 6.
Rubriques apparues dans les sommaires entre 1998 et 2000*

Rubriques	95	96	97	98	99	00	01	02	03
Dates clés				*	*	*			
Une banque proche de ses clients				*					
Les participations dans les banques étrangères				*	*	*			
Le Groupe Crédit agricole					*		*	*	*
Organigramme financier simplifié du Groupe Crédit agricole / Principales filiales et participations					*	*	*	*	*
Les principales filiales du Crédit agricole					*	*			
L'activité en [telle année]					*	*			
La banque privée					*	*			

La recherche d'une meilleure lisibilité de l'évolution du Crédit Agricole peut expliquer l'apparition de rubriques telles que, les « dates clés » (commentaire succinct de quelques dates jalonnant le parcours du banquier de l'agriculture française devenu banque internationale), l'organigramme financier simplifié (présentation des entités dont se compose le groupe ainsi que des relations capitalistiques qui les unissent), le récapitulatif des « principales filiales » ou encore les deux méta-rubriques « Le Groupe Crédit Agricole » et « L'activité en [telle année] », qui fonctionnent en opposition en regroupant ce qui concerne, d'une part, l'institution, et d'autre part, son activité. L'apparition des entrées « participations dans les banques étrangères » et « banque privée » peuvent en revanche être attribuées au développement de ces activités. On notera enfin que l'année 2000 n'apporte aucune modification notable, comme si ce *statu quo* avait eu pour fonction de conjurer des craintes millénaristes (on pense, dans le même registre, au *bug* de l'an 2000). Comme on va le voir, la période suivante introduit à nouveau d'importants changements.

*Tableau 7.
Rubriques apparues dans le sommaire de l'année 2001*

Rubriques	95	96	97	98	99	00	01	02	03
Crédit Agricole SA : une banque universelle							*		
Banque de grande clientèle							*	*	
Banque de détail à l'étranger							*	*	*
L'action Crédit Agricole SA							*	*	
Un acteur majeur en France et en Europe							*		
Gestion pour compte propre et divers							*	*	
Bilans consolidés et simplifiés du groupe Crédit Agricole							*	*	*
Résultats consolidés et simplifiés du groupe Crédit Agricole							*	*	*

L'année 2001 se caractérise tout d'abord par la disparition des marchés traditionnellement présentés dans le sommaire. Celui-ci est refondu en prenant en compte, non plus les différents marchés, mais les différents métiers exercés par l'entité Crédit Agricole SA. Aux cinq marchés traditionnels ont été substitués des métiers, dont certains sont apparus dès 1996 mais qui, jusqu'alors, étaient venus compléter le canevas de 1991. Il s'agit de la « banque de proximité » (qui regroupe la banque dite « de détail », opérant sur les marchés des particuliers, des professionnels, des agriculteurs, des entreprises et des collectivités publiques), de la « gestion d'actifs, de l'assurance et de la banque privée » (gestion financière effectuée pour des particuliers sur les marchés boursiers), de la banque de grandes clientèles (financements internationaux et gestion financière pour le compte de sociétés), de la « banque de détail à l'étranger » (participations dans des établissements faisant le même métier que le Crédit Agricole, mais à l'étranger), et de la « gestion pour compte propre et divers » (gestion du patrimoine du Crédit Agricole).

Cette réorganisation peut être directement attribuée au changement de statut du destinataire du rapport d'activité. En effet, l'organe central du Crédit Agricole est devenu en 2001 une société cotée en bourse. Cependant, comme cela a déjà été présenté dans le premier chapitre, les actions cotées ne représentent pas la totalité du

capital du Crédit Agricole, mais seulement une partie comprenant l'organe central, toutes les filiales nationales du groupe et 25% du capital des Caisses régionales de Crédit Agricole. Le fait que la banque de détail en France ne pèse plus qu'à hauteur de 25% dans les résultats de la nouvelle entité permet un rééquilibrage au profit des autres activités du Groupe Crédit Agricole S.A.

Le rapport d'activité devient alors un double instrument de communication : il est toujours au service du Crédit Agricole, mais surtout de Crédit Agricole S.A., la partie du groupe cotée en bourse. La rubrique « Crédit Agricole SA : une banque universelle » offre une illustration éloquente de ce changement, en attribuant à la nouvelle entité une formule qui était jusqu'alors traditionnellement réservée au groupe Crédit Agricole dans son ensemble (voir plus haut, « une introduction à l'histoire de l'institution »). La présentation du groupe Crédit Agricole n'est pas éliminée, mais elle est reléguée au deuxième plan. Les rubriques telles que « le groupe Crédit Agricole », « l'organisation du groupe », « les implantations du groupe », « les Caisses régionales », etc. sont préservées – car il s'agit bien de capitaliser sur l'image du groupe –, mais elles sont présentées après tout ce qui concerne « l'autre groupe », le groupe Crédit Agricole S.A.

L'apparition de la rubrique « L'action de Crédit Agricole S.A. » vient mettre l'accent sur une donnée nouvelle directement liée à la cotation en bourse. Quant à la présence des bilans et résultats simplifiés, elle marque une nouvelle étape dans l'extension prise par cette partie du rapport annuel. Ce qui fait toujours l'objet de commentaires dans le rapport financier a été remonté dans cette partie, probablement dans le but de fournir au public des non spécialistes (actionnaires individuels, entre autres), les éléments chiffrés fondamentaux présentés lors de l'assemblée générale. Après ce deuxième changement majeur, l'année 2002 se caractérise plutôt par des compléments (comme cela avait déjà été remarqué pour la période 1998-2000).

*Tableau 8.
Rubriques apparues dans le sommaire 2002*

Rubriques	95	96	97	98	99	00	01	02	03
Le groupe Crédit Agricole SA								*	
Les ressources humaines								*	*
Le développement durable								*	*

La logique qui préside à l'apparition de la méta-rubrique « Le groupe Crédit Agricole S.A. », qui fonctionne en opposition avec « Le groupe Crédit Agricole », a déjà été exposée plus haut : le rapport d'activité est devenu un instrument de communication institutionnelle et financière principalement au service de la nouvelle entité. Cela se traduit par des opérations de « calque » ou d'écho, qui consistent à reprendre ce qui caractérisait exclusivement la présentation du Crédit Agricole dans son ensemble, pour l'attribuer également à Crédit Agricole S.A. Quant à l'apparition des rubriques « Les ressources humaines » et « Le groupe Crédit Agricole et le développement durable », il semble qu'il faille plutôt les attribuer à ce que l'on a proposé d'appeler plus haut les « effets de genre ». En effet, les ressources humaines font partie du corps du texte des rapports d'activité depuis la fin des années 1960, avec des développements variables selon les périodes (d'une mention dans un organigramme à une rubrique à part entière, comme dans le rapport d'activité de 1991). Leur apparition concomitamment avec celle du développement durable¹ laisse penser qu'il s'agit d'un effet de mode en relation avec d'autres thématiques telles que la protection de l'environnement, le commerce équitable ou les fonds éthiques. Il n'est probablement pas erroné de penser que parmi plusieurs millions d'actionnaires, certains sont sensibles à ce genre d'arguments. Ils font en

¹ Le rapport Brundtland a introduit la notion de développement durable : « *Humanity has the ability to make development sustainable to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs. (...) sustainable development is not a fixed state of harmony, but rather a process of change in which the exploitation of resources, the direction of investments, the orientation of technological development, and institutional change are made consistent with future as well as present needs.* » Report of the World commission on environment and Development – “Our Common Future” A/42/427 - English p.24-25. United Nations, 1987.

tout cas partie d'une tendance générale relevée dans les rapports annuels produits ces dernières années¹.

Contrairement à 2002, l'année 2003 apporte un nombre conséquent de nouveautés au sommaire. Celles-ci peuvent être regroupées en deux ensembles selon la cause qui a motivé leur apparition. Le premier concerne l'acquisition du Crédit Lyonnais et rassemble « Un groupe puissant. 2003 : la concrétisation d'un projet ambitieux », « Six pôles métiers », « Services financiers spécialisés » et « Banque de financement et d'investissement ». L'acquisition du Crédit Lyonnais a permis à Crédit Agricole S.A. de posséder un réseau d'agences (la banque de détail du Crédit Lyonnais) et de se renforcer dans les métiers déjà traités par ses filiales. Les activités de banque de financement et d'investissement ont été rapprochées de celles de l'ex-banque d'Indosuez, devenue entre-temps Crédit Agricole Indosuez. Les métiers de crédit à la consommation, d'affacturage et de crédit-bail, regroupés sous le nom de services financiers spécialisés, ont été renforcés par l'arrivée d'entités autrefois concurrentes. Ainsi, la refonte du sommaire introduite en 2001 se trouve-t-elle complétée : la banque de proximité se dédouble en Crédit Agricole (à hauteur de 25%) et Crédit Lyonnais (100%) ; la gestion d'actifs, l'assurance la banque privée, ainsi que la banque de détail à l'étranger sont reconduites à l'identique ; la banque de financement et d'investissement remplace la banque de grande clientèle. L'apparition des services financiers spécialisés vient compléter la série de six pôles autour desquels s'organise l'activité de Crédit Agricole S.A.

¹ A. Guimard (2001 : 61-62).

*Tableau 9.
Rubriques apparues dans le sommaire 2003*

Rubriques	95	96	97	98	99	00	01	02	03
Un groupe puissant. 2003 : la concrétisation d'un projet ambitieux									*
Gouvernance de l'entreprise									*
Six pôles métiers									*
Services financiers spécialisés									*
Banque de financement et d'investissement									*
Solidarité et engagement auprès de la société civile									*
Relations avec les clients et les fournisseurs									*
L'environnement : des formes d'action variées pour une implication forte									*
Crédit Agricole SA et ses actionnaires									*

Un deuxième ensemble de nouvelles rubriques est à rattacher au développement de thématiques apparues en 2001 ou en 2002. Les informations participant des relations entretenues avec l'actionnariat sont développées, et la nouvelle rubrique inclut « l'action Crédit Agricole S.A. ». D'autre part, le développement durable s'est également étoffé et a absorbé les ressources humaines (rubriques « Solidarité et engagement auprès de la société civile », « Relations avec les clients et les fournisseur », et « L'environnement... »). L'effet de genre semble ainsi contribuer, autant que les transformations de l'institution, à l'évolution du découpage en rubriques des textes des rapports d'activité. On dispose d'un indice, qui devra être contrôlé par d'autres moyens, relatif à un infléchissement de l'usage communicationnel des rapports d'activité et allant dans le sens d'une plus grande sensibilité aux attentes d'un public d'investisseurs, individuels mais aussi institutionnels. C'est pourquoi il a semblé nécessaire de prendre en compte ce découpage documentaire secondaire pour la mise en place d'un corpus au format électronique. Cette étape nécessite néanmoins que l'on procède au préalable à une définition critique de la notion de rubrique.

b. Définition de la notion de rubrique

De même que, dans la première section, il s'est avéré nécessaire de définir la notion de contexte à dominante non syntactique, afin de prendre toute la mesure de ce que ce type de contexte a de particulier, on a pensé devoir revenir sur la notion de rubrique. Dans la mesure où l'on souhaite observer le fonctionnement en contexte de certaines unités linguistiques, il convient de ne pas se limiter aux seules rubriques proposées par les sommaires, afin de prendre en compte tout l'appareil d'intertitres présent dans les textes du corpus¹. En effet, si les rubriques de sommaire fournissent des indications précieuses sur les principales évolutions affectant l'activité de l'institution, elles sont néanmoins sujettes à un certain nombre d'effets, plus directement liés au fonctionnement du rapport d'activité comme support de communication. Comme on a pu le constater, il peut s'agir de valorisations temporaires, d'apparitions et de disparitions de rubriques alors que leur matière fait toujours l'objet d'un traitement dans le corps du texte, mais aussi de phénomènes de mode visant à présenter des réalités anciennes sous un jour nouveau (développement durable) ou encore, de la présence de méta-rubriques au statut mal défini. Ce dernier découpage, qui a pour fonction première de structurer les entrées de sommaire, correspond aussi, mais dans certains cas seulement, à des pages présentant un titre isolé ou associé à des introductions plus ou moins développées.

Cette variabilité du sommaire dissimule souvent un découpage plus stable à l'aide d'intertitres qui structurent directement le corps du texte. Pour prendre en compte cette structuration des rapports d'activité, on propose de définir la rubrique comme un intervalle de texte, suivi ou non, compris entre deux intertitres. En procédant de la sorte, on accroît considérablement le nombre de contextes à répertorier, puisque l'on passe d'une cinquantaine d'entrées (présentées par ordre

¹ À propos d'un genre différent (celui de l'œuvre littéraire), G. Genette fait les remarques suivantes : « *La table des matières n'est, elle non plus [il s'agit des titres courants], en principe, rien d'autre qu'un instrument de rappel de l'appareil titulaire – ou d'annonce, lorsqu'elle se trouve en tête, comme souvent autrefois, et encore aujourd'hui dans les livres allemands ou anglo-saxons. (...) Mais la table n'est pas toujours le relevé fidèle de l'appareil intertitulaire. Elle peut le trahir, par réduction, comme dans certaines éditions économiques ou négligentes où les chapitres numérotés sans titres ne font tout simplement pas l'objet d'une table ; ou par amplification, en attribuant des titres à des chapitres qui n'en comportaient pas in situ : c'est, je le rappelle, le parti adopté par Proust pour les Jeunes Filles ; ou par variation désinvolte, comme il arrive chez Stendhal (...)* ». in *Seuils*, (1987 : 292).

d'apparition dans les tableaux de cette section) à près de huit cents¹. C'est au niveau de ce découpage, beaucoup plus fin que le précédent, qu'il conviendra donc de caractériser les rubriques, selon qu'elles relèvent contexte à dominante syntactique ou non syntactique.

L'examen des rubriques de sommaire des neuf rapports d'activité qui composent le corpus principal de cette étude a permis d'identifier, d'une part, deux sortes de contextes caractéristiques (à dominante syntactique et non syntactique), et d'autre part, le rôle structurant de ce découpage dans la restitution de l'activité bancaire. La mise en forme *machine readable* du corpus, qui est destinée à le rendre exploitable par un programme informatique, devra par conséquent chercher à restituer ces propriétés. En effet, celles-ci organisent le contexte immédiat du vocabulaire que l'on se propose d'étudier par la suite.

IV. CONSTITUTION DU CORPUS

A. Le passage au format « texte »

L'exploitation d'un corpus de textes au format électronique repose sur une transformation préalable du support d'origine. Celui-ci peut être dans un format papier ou électronique. La première sorte de support nécessite une transcription manuelle, ou assistée par la production d'une image électronique (scanner) à laquelle une reconnaissance automatique de caractères est appliquée. La seconde sorte de format, bien qu'électronique, peut elle aussi nécessiter certains traitements, entre autres, lorsqu'elle se présente sous la forme d'un fichier de PAO ou d'un fichier PDF². Cette transformation, bien que restant fidèle au texte, entraîne la disparition d'éléments graphiques. Ainsi, les polices de caractères perdent-elles leur diversité, leurs couleurs et leurs différentes casses, au profit d'un noir et blanc

¹ Ce découpage est donné en annexe.

² *Portable Document Format* ; format de fichier électronique, ayant pour extension « .pdf ». Ce format est très répandu sur le Web et sur les intranets institutionnels, dans la mesure où il permet de diffuser des supports de communication élaborés (plaquettes, revues, etc.), avec leur mise en page originale. Ces fichiers, conçus à l'aide de logiciels de PAO plutôt rares et coûteux, peuvent alors être lus à l'aide du logiciel « Acrobat » diffusé gratuitement. De plus, ce format conserve les polices de caractères, ce qui permet d'indexer le texte des documents ou de le récupérer au format texte, et compresse les images, ce qui diminue le poids des fichiers et facilite leur diffusion.

homogène et d'une police unique. Les images disparaissent, de même que les schémas. Les tableaux, blocs de textes et autres mises en pages sur plusieurs colonnes, sont ramenés à des suites linéaires. Paradoxalement, c'est cet appauvrissement qui conditionne la richesse de l'exploration des données textuelles au format texte simple¹. D'autre part, on verra que le texte ainsi transformé va s'accroître d'un certain nombre de balises dont la fonction est de restituer en partie les caractéristiques de l'original.

Les rapports d'activité proviennent de la sélection de la première partie des rapports annuels des années 1995 à 2003 disponibles au format PDF. L'unicité du format dissimule une différence de traitement appliquée à la source. Les textes correspondant aux deux premières années (1995 et 1996) ont été obtenus à partir d'un scannage de l'original papier. L'image numérique obtenue a ensuite été soumise au traitement d'un logiciel de reconnaissance optique de caractères. Le résultat est globalement satisfaisant, mais il comporte un certain nombre d'erreurs. Par exemple, une page a pu être oubliée, et surtout, certains mots ont pu être mal reconnus. Dans les cas repérés, on a procédé à la réparation des oublis ainsi qu'à la correction de l'orthographe. Lorsque ces erreurs n'ont pas été détectées lors de la phase de traitement, elles sont apparues au moment de la segmentation des formes par l'outil textométrique. Cependant, l'examen du dictionnaire des formes du corpus a montré que ces scories ont en général une fréquence faible (une occurrence) et sont dispersées, ce qui n'a pas semblé devoir imposer une correction systématique. Par ailleurs, on remarquera que dans le cadre opérationnel d'une extraction d'information, il n'est jamais possible de compter sur la parfaite qualité d'un texte numérique récupéré sur un intranet.

La deuxième série de fichiers numériques au format PDF (années 1997 à 2003), a été obtenue à partir des fichiers d'impression originaux. Dans ce cas, les problèmes de reconnaissance de caractères sont supprimés. Cependant, des

¹ Le format XML permet de conserver une organisation de l'information avec ses différentes composantes, mais il nécessite le développement de feuilles XSL de manière à produire un résultat affichable, par exemple, à l'aide d'un navigateur. Le format HTML, du fait qu'il combine éléments nécessaires à la mise en forme et éléments textuels ou graphiques, peut être difficile à exploiter par les programmes textométriques du fait du nombre important de scories générées par les balises en de mise en forme.

difficultés peuvent surgir au moment de la récupération du texte¹. En effet, si le format PDF permet d'incorporer les polices ayant servi pour l'original numérique, cette possibilité n'est pas toujours utilisée. Dans ce cas, le texte recopié peut comporter des caractères mal reconnus ou mal séparés (chaînes entre lesquelles les espaces ont disparu), qu'il convient de traiter. Une autre difficulté peut provenir des modifications de mise en page provoquées automatiquement par le passage du format PDF au format texte : c'est par exemple le cas lorsqu'un texte en colonnes est coupé par une image ou un diagramme. L'ordre des paragraphes peut alors être interverti ou complètement modifié. Ces transformations non souhaitées ont été corrigées manuellement.

Lors d'une première étape, le texte brut récupéré depuis les fichiers PDF a subi un certain nombre de corrections élémentaires qui ont eu pour but de retranscrire le plus fidèlement possible le texte original. C'est lors d'une deuxième étape que celui-ci retrouve également l'essentiel de sa segmentation en paragraphes et rubriques.

B. Traitements appliqués au texte brut

1. Balisages documentaires : rapports d'activité et rubriques

Cette section décrit les principaux traitements appliqués au texte brut copié à partir des fichiers PDF des rapports d'activité. Ces corrections ont été automatisées dans la mesure du possible, ce qui n'a pas dispensé de recourir à des corrections manuelles. Elles ont porté sur la restitution des chaînes de caractères, des lignes et des paragraphes d'origine. La pagination a été conservée à titre indicatif, ne serait-ce que pour faciliter un éventuel recours à la présentation originale. Les mots coupés en fin de ligne ont été rétablis dans leur forme normale.

Dans la mesure où le corpus constitué est ouvert, il est nécessaire de formaliser les traitements qui sont appliqués aux textes qui le composent. Il devient

¹ Le format PDF permet d'obtenir une image plus ou moins dégradée de l'image numérique ainsi qu'une copie des polices de caractères utilisées pour le texte. Le logiciel Acrobat permet de sélectionner l'intégralité ou une partie du texte du document. Cette sélection peut ensuite être collée dans le presse-papiers du système d'exploitation de l'ordinateur, dans un éditeur de texte ou un logiciel de traitement de texte.

ainsi possible de répéter ces opérations autant de fois que nécessaire, sans nuire à l'homogénéité des données. Les principaux traitements sont décrits dans les lignes qui suivent. Ils ont été réalisés à l'aide d'un éditeur de texte¹.

Des balises² ont été déclarées dans le corpus pour que les textes et leurs sous parties puissent être identifiés par les programmes informatiques. L'unité principale du corpus est le rapport d'activité tiré d'un rapport annuel. Elle est identifiée à l'aide de la balise « doct=aaaaRA », ou simplement « doct=aaaa » si une telle notation n'est pas ambiguë, pour un document correspondant au rapport d'activité d'une année particulière. Par exemple, le rapport d'activité de l'année 2001 sera déclaré ainsi : <doct=2001RA> ou <doct=2001>. Cette balise permet ensuite de comparer de façon statistique les parties ainsi délimitées (ces traitements ont été réalisés à l'aide du logiciel Lexico, présenté plus loin).

La deuxième unité textuelle prise en compte pour le balisage est la rubrique, telle qu'elle a été définie plus haut : un intervalle de texte, suivi ou non, compris entre deux intertitres. La balise est positionnée avant le titre de la rubrique et détermine un intervalle de texte allant jusqu'à la balise suivante. La convention adoptée a consisté à qualifier la rubrique de « rub » et à lui adjoindre une année, celle du rapport d'activité auquel elle appartient, suivie d'un numéro correspondant à son ordre d'apparition dans le texte. Par exemple, la rubrique numéro un du rapport d'activité de l'année 2001 est notée <rub=2001-01>. Cette balise est redondée par un signe typographique, en l'occurrence « \$ ». De cette manière, la rubrique peut être considérée à la fois comme une partie du corpus à part entière, au

¹ A titre indicatif, le temps de traitement d'une page comportant huit éléments de natures différentes (accroche, graphique, titre spécial, texte court spécial avec mise en forme particulière), 220 mots, 1250 caractères nécessite environ cinq minutes. L'éditeur de texte pour Windows « TextPad » a été utilisé.

² Le balisage utilisé pour ce corpus ne respecte pas l'encodage proposé dans le cadre de la TEI (Text encoding initiative). Il convient de préciser les raisons qui ont motivé notre choix. L'esprit de la TEI est animé par deux préoccupations principales : la première est de mettre en place un cadre logique, une norme, pour faciliter la diffusion internationale des données textuelles ; la seconde, est de documenter au mieux ces données à l'aide d'informations fiables et exhaustives. Dans notre perspective, les échanges internationaux ne sont pas une priorité, et le choix des formats SGML ou XML ne s'imposent pas, d'autant plus que les principaux logiciels que nous utilisons travaillent encore avec de simples fichiers textes. En revanche, la documentation du corpus correspond dans notre perspective à un impératif méthodologique. Cela nous a parfois conduit sur ce point à dépasser les exigences des recommandations de la TEI. On renvoie à l'article de B. Habert (2000 : 42-46), pour une discussion de l'usage de la TEI.

même titre que chaque rapport d'activité, ce qui autorise des traitements statistiques de comparaison de partie à partie, et comme une simple délimitation graphique, comme les paragraphes présentés ci-dessous.

2. Balisage des contextes CDN et CDS et paragraphes

La troisième unité textuelle déclarée sous la forme d'une balise correspond à la nature du contexte auquel chaque rubrique peut être rattachée. On a vu plus haut qu'une distinction a été opérée entre les contextes à dominante syntactique ou non syntactique. Cette balise se présente sous la forme suivante : **<contexte=CDS>**, pour une rubrique à dominante syntactique, et **<contexte=CDN>** pour une rubrique de nature non syntactique. Par exemple, la première rubrique du rapport d'activité de l'année 2001 (il s'agit du sommaire), est considérée comme correspondant à un contexte à dominante non syntactique ; elle est donc balisée de cette manière : **<contexte=CDN> <rub=2001-01>**.

Par ailleurs, la pagination d'origine des rapports d'activité a été globalement conservée, une exception ayant été faite pour les pages blanches et les pages de titre. Il en résulte donc un léger décalage qui a peu d'incidences sur la concordance entre pages du corpus et pages du document original. La balise adoptée reprend le signe « paragraphe » suivi du numéro de page correspondant. Par exemple, la page numéro 3 de l'un quelconque des rapports d'activité est notée « §3 ».

Ces balises encadrent la sélection manuelle et la copie des pages ou des parties de page du fichier PDF source. Le logiciel Acrobat permet de copier l'intégralité du texte d'un fichier PDF dans le presse papier de Windows, mais cette option n'a pas été retenue pour la constitution de notre corpus. En effet, le gain de temps est alors contrebalancé par la perte de certains repères figurant dans le texte original. Lorsque la délimitation des pages du fichier Acrobat n'est pas reprise systématiquement, la correspondance entre le format PDF et le format texte est perdue, ou nécessite un fastidieux repérage pour être rétablie. D'autre part, la copie de mises en page complexes fait souvent perdre le fil d'un paragraphe coupé par une illustration ou un graphique. Enfin, les titres et chapeaux en tête de page sont systématiquement collés en fin de page. Si ce phénomène se produit également lors d'une sélection manuelle page à page, il est plus aisé à maîtriser. En conséquence,

on a estimé que pour les documents composant le corpus principal, la sélection intégrale du texte du fichier PDF ne présente pas d'intérêt. Cependant, on se réserve la possibilité de l'utiliser pour des documents aux mises en pages moins élaborées et dont le contenu pourra alimenter des corpus *ad hoc* « jetables », pour lesquels la rapidité de récupération du texte l'emportera sur la qualité de la restitution.

Enfin, les paragraphes d'origine ont été restaurés, pour la plus grande partie, de façon automatique. Cependant, dans certains cas, le recours à une correction manuelle s'est avéré nécessaire. En effet, la conservation de la mise en forme opérée par le format PDF est essentiellement destinée à l'affichage. Le texte extrait du fichier est organisé, non pas comme un texte suivi, mais comme une suite de lignes empilées les unes sur les autres, sans liaison. Transposée au format texte, chaque ligne se termine par un signe de fin de ligne, que la phrase soit achevée ou non. Il est alors impossible de départager les fins de lignes créés artificiellement pour respecter la largeur de la colonne dans le texte d'origine, de celles qui correspondent à une fin de phrase réelle ou à un retour à la ligne marquant un changement de paragraphe.

Deux traitements ont donc été appliqués : l'un consiste en la restauration de la continuité des phrases coupées artificiellement, et l'autre est destiné à rétablir les paragraphes d'origine, tels qu'ils avaient été voulus par les auteurs des textes. Ces opérations ont été réalisées à l'aide d'un éditeur de texte permettant de saisir des expressions régulières élémentaires¹. En définitive, le paragraphe repéré automatiquement est indiqué à l'aide du caractère « # ». Il faut remarquer que

¹ Le traitement se décompose selon les étapes suivantes, qui peuvent être regroupées dans une macro-commande :

- pour les fins de ligne se terminant par un point et le signe de fin de ligne, ces deux signes sont remplacés arbitrairement par une chaîne sans occurrence dans le texte d'origine : l'expression régulière « `\\.n` » permet de repérer ces fins de ligne particulières qui sont remplacées par la chaîne « `@@` » ;

- pour les fins de ligne ne se terminant pas par un point mais comportant le signe de fin de ligne, ce dernier signe est remplacé par un espace : les phrases coupées artificiellement sont alors rétablies ;

- les chaînes « `@@` » sont ensuite remplacées par un point, suivi du caractère « # », destiné à repérer les paragraphes, et du signe de fin de ligne qui vient terminer l'expression régulière : « `\\.#n` » : les paragraphes d'origine sont alors rétablis ;

- il reste alors à traiter le cas peu fréquent des phrases se terminant en fin de ligne mais qui ne correspondent pas à un véritable paragraphe : le traitement automatique les a transformées par erreur en paragraphes. Dans ce cas, la continuité du texte d'origine a été rétablie manuellement.

Pour les expressions régulières, on renvoie à B. Desgraupes (2001).

l'utilisation systématique d'une expression régulière peut générer des paragraphes artificiels. Dans ce cas, un traitement correctif a été réalisé manuellement¹.

Ces différents balisages permettent d'enrichir le texte brut en restituant les découpages originaux (document, rubriques, pages et paragraphes), ou en lui adjoignant une caractérisation spécifique (contextes à dominante syntactique ou non syntactique). On a vu plus haut que ce dernier type de contexte comporte des tableaux, des schémas, des histogrammes et d'autres sortes de graphiques dont les données textuelles ont été conservées. Leur transcription n'est cependant pas aussi naturelle que pour le texte suivi. Il a donc été nécessaire d'adopter une règle : l'unité la plus petite, par exemple un cartouche d'organigramme, a été considérée comme étant l'équivalent d'un paragraphe. Un histogramme comportant peu d'intitulés de colonnes a été considéré comme un seul paragraphe, décomposé en éléments séparés par des virgules. Cette transposition a visé à ménager autant que possible l'organisation initiale des informations. Elle ne constitue pas un équivalent tout à fait satisfaisant, dans la mesure où l'on perd la possibilité de la lecture synoptique offerte par l'original, mais elle permet néanmoins de conserver certaines dénominations apparaissant dans ces contextes particuliers.

On donne ci-dessous un exemple de transformation du texte dans son format électronique d'origine en un texte formaté pour qu'il puisse être exploité à l'aide d'un programme de traitement textométrique. La perte de la mise en forme est en quelque sorte compensée par l'apparition d'un balisage qui n'existe pas en tant que tel dans le format d'origine, ou qui nécessiterait une transposition.

¹ La mise en forme du corpus présentée ici correspond aux possibilités offertes par le logiciel Lexico. L'autre logiciel utilisé dans ce travail, Hyperbase, nécessite un balisage légèrement différent pour les différentes parties et la pagination des textes. Quant à la délimitation des paragraphes, elle est effectuée automatiquement dans ce logiciel en adoptant des critères prédéfinis. Le remplacement d'un balisage par un autre se fait très rapidement et de manière fiable à l'aide d'un éditeur de texte de type Tepad.

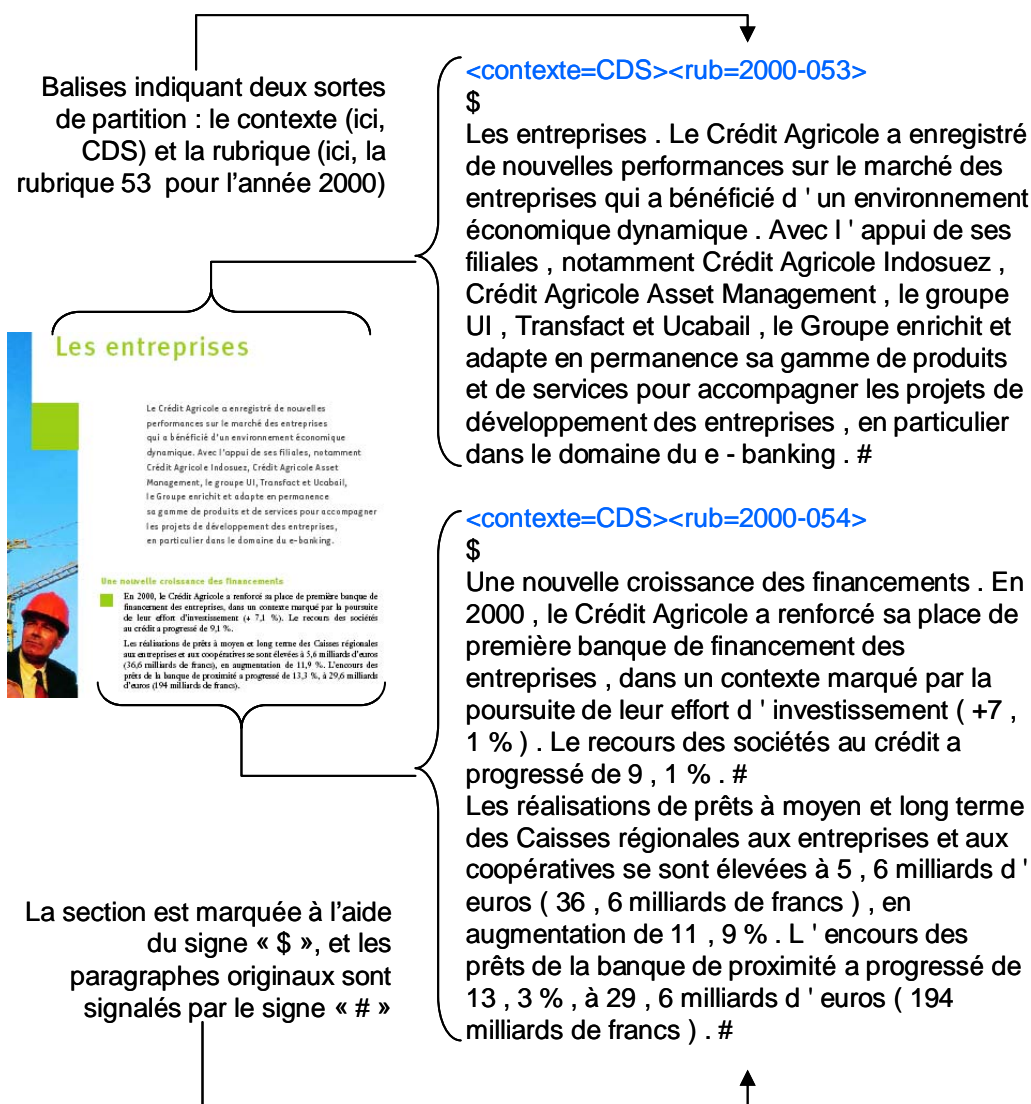


Figure 12.

Page extraite du rapport d'activité du Crédit Agricole pour l'année 2000 (côté gauche) et sa transposition en texte balisé exploitable par un logiciel de textométrie¹ (côté droit)

Le fichier comprenant les textes du corpus, balisés comme cela vient d'être exposé, peut ensuite être traité par les programmes de textométrie. Ce traitement comporte généralement plusieurs étapes qui sont rappelées dans la section suivante.

¹ Le texte brut présenté ici a déjà subi une opération de segmentation repérable aux espaces qui encadrent les signes de ponctuation, les parenthèses et les caractères spéciaux. En effet, cette opération vise à ramener le texte à une série de formes simples séparées les unes des autres par deux espaces (ou blancs).

3. Préparation des textes en vue de leur exploitation textométrique

a. Traitements automatiques de base

Les principaux logiciels de textométrie utilisés pour ce travail, Lexico 3¹ et Hyperbase 5² opèrent automatiquement différents traitements sur les textes formatés qu'on leur soumet. Mais alors que le second a pris l'option de rassembler le résultat des traitements sous la forme d'une base autonome, constituée automatiquement à l'aide de paramétrages classiques, le premier laisse en permanence toute latitude au chercheur pour intervenir sur les différents paramètres utilisés par le programme. C'est donc Lexico 3 qui servira à illustrer les étapes de la préparation des textes à leur exploitation textométrique.

On a indiqué plus haut que les traitements appliqués au texte brut ne concernent en définitive que son organisation ou les éventuelles scories laissées par une reconnaissance optique de caractères. On a pris le parti de ne pas procéder à de plus amples traitements dans la mesure où ces derniers n'ont pas semblés indispensables. C'est ainsi que l'on n'a pas cherché à aligner les formes graphiques identiques distinguées seulement par la présence d'une majuscule initiale. L'inconvénient prévisible, par exemple que « Il » et « il » soient comptabilisés comme deux formes distinctes, n'a pas paru rédhibitoire, dans la mesure où les logiciels proposent des facilités de manipulation. Hyperbase procède automatiquement à la conversion des formes avec majuscules lorsqu'elles existent également dans le corpus sans majuscule initiale³. Lexico permet, après traitement, de réunir les deux formes graphiques dans un groupe de formes. Il devient alors possible d'avoir une vision consolidée de la fréquence du pronom pris comme exemple, ou de n'importe qu'elle autre forme. En revanche, l'intérêt de conserver cette distinction permet à ce stade de ne pas trancher certains points délicats qui ne se limitent pas aux seuls débuts de phrase, mais qui touchent également aux usages

¹ Lexico 3 : A. Salem *et alii*, SYLED – ILPGA, Université de la Sorbonne nouvelle, Paris III. Le logiciel est présenté et téléchargeable à l'adresse suivante : <http://www.cavi.univ-paris3.fr/ilpga/ilpga/tal/lexicoWWW/>.

² Hyperbase 5 : É. Brunet, UPRESA « Bases, corpus et langage », Université de Nice. Le logiciel est présenté et peut être commandé à l'adresse suivante : <http://ancilla.unice.fr>.

³ En fait l'algorithme est plus complexe. On revient sur ce point dans la troisième partie à propos de l'identification des dénominations propres présentes dans un corpus.

de la majuscule dans certains dispositifs non syntactiques, ou encore, à des phénomènes liés à l'utilisation de cet artifice graphique pour valoriser les réalités désignées par certains mots. Ces différents points sont exposés dans les parties suivantes qui traitent de façon détaillée du vocabulaire du corpus.

Les traitements de base appliqués aux textes du corpus consistent en une vérification des règles de formatage du corpus (respect de la grammaire utilisée pour les balises), ainsi qu'en une segmentation automatique : « *ensemble des opérations réalisées au moyen de procédures informatisées qui aboutissent à découper, selon des règles pré-définies, un texte stocké sur un support lisible par un ordinateur en unités distinctes que l'on appelle des unités minimales*¹. » La règle principale qui doit être adoptée concerne le choix des délimiteurs, c'est-à-dire les ponctuations faibles et fortes qui, en plus des espaces ou « blancs », vont permettre de délimiter chaque occurrence présente dans les textes : « *suite de caractères non-délimiteurs bornée à ses extrémités par deux caractères délimiteurs de forme*². » Par exemple, dans le cadre de ce travail, l'exploitation des corpus repose sur la liste suivante de caractères délimiteurs : « `_,.,:;!/?/_-\'"()[]{}$$-.*#&+><` ». La segmentation rend ensuite possible le comptage des occurrences et la constitution d'un index. Celui-ci reprend toutes les formes graphiques isolées par la segmentation et leur associe une fréquence globale.

b. Principales caractéristiques textométriques du corpus

À ce stade, le logiciel fournit les principales caractéristiques du corpus : il se compose de 201 902 occurrences pour 11 257 formes. Pour ces dernières, le nombre de hapax (fréquence égale à une occurrence) s'élève à 4 166, et la fréquence la plus élevée est de 12 744 occurrences pour le mot « de », ce qui est un trait caractéristique des corpus de langue française possédant au moins quelques milliers d'occurrences.

Lorsqu'un corpus est composé de plusieurs parties, comme c'est le cas pour celui que l'on utilise, le programme propose un découpage des textes selon les

¹ L. Lebart, A. Salem (1994 : 317).

² *Ibid.* (315).

partitions définies lors de la déclaration des balises. On donne ci-dessous les caractéristiques du corpus lorsqu'il est découpé en fonction de ses parties principales, correspondant au texte de chacun des neuf rapports d'activité qui ont été rassemblés.

Tableau 10.
Principales caractéristiques textométriques du corpus de rapports d'activité
« RA9503 » en fonction de sa partition principale

Partie	Nb occurrences	Nb formes	Nb hapax	Fréq. Max	Forme
1995RA	18274	3818	2323	1140	de
1996RA	19306	3726	2231	1304	de
1997RA	21965	3874	2203	1417	de
1998RA	18478	3382	1911	1182	de
1999RA	21756	3720	2039	1347	de
2000RA	22032	3783	2095	1321	de
2001RA	21162	3515	1823	1338	de
2002RA	19870	3418	1800	1260	de
2003RA	39059	5582	2916	2435	de

Lecture du tableau 10 : les six colonnes du tableau reprennent respectivement les informations suivantes pour chacune des parties dont se compose le corpus : « Partie » indique le nom de la partie qui figure dans la balise « doct » présentée plus haut ; « Nb occurrences » donne le nombre d'occurrences ; « Nb formes » donne le nombre de formes graphiques différentes présentes dans chaque partie du corpus ; « Nb hapax » donne le nombre de formes hapax présentes, et « Fréq. Max » indique la fréquence la plus élevée observée pour une partie donnée ; enfin, « Forme » indique quelle est la forme graphique qui possède la fréquence la plus élevée dans chacune des parties du corpus.

Sauf indication explicite, c'est ce corpus qui sera utilisé dans les développements qui suivent. Le découpage en partie va permettre de suivre la ventilation des occurrences sur une période d'une dizaine d'années, mais aussi la répartition des formes en fonction des rubriques ou des types de contextes qui ont été identifiés plus haut. Les logiciels utilisés permettent d'effectuer d'autres opérations, allant de la localisation d'une occurrence dans les paragraphes du corpus à la comparaison d'ensembles de vocabulaire équivalant aux contextes à dominante syntactique et non syntactique, aux parties principales ou encore aux rubriques. Ces autres facettes de l'approche textométrique seront présentées au fil des chapitres, lorsque cela sera nécessaire.

La constitution d'un corpus, en vue de servir de ressource pour l'élaboration d'un référentiel terminologique adaptable au contexte, dépend d'un certain nombre de conditions, qui l'enracinent dans les différentes dimensions contextuelles de son utilisation.

Il s'agit tout d'abord des objectifs assignés au référentiel à élaborer. On a vu plus haut qu'il s'agit de documenter des activités d'information mettant en jeu la présentation d'un référentiel spécifique à une population de visiteurs néophytes consultant les sites intranets d'un groupe bancaire. Cette situation précise, on aurait également pu ajouter l'indication de la période, l'année 2003, est renforcée par un choix d'une autre nature. En effet, la démarche propose de sélectionner le corpus parmi les traces laissées par de précédents échanges langagiers ayant eu lieu dans des situations comparables à celle où se trouvent les sites intranets évoqués plus haut, mais pour lesquelles on a des raisons de penser que la prise en charge des destinataires a été réalisée dans de meilleures conditions. La nécessité de trouver dans le contexte des ressources adéquates découle de cette option. Mais en réalité, il n'y a à ce stade aucune certitude sur l'existence, et *a fortiori*, l'adéquation de telles ressources. On alors cherché des valeurs approchées dans le stock local de discours. Parmi les productions de la communication d'entreprise, on a identifié que les rapports annuels, et plus précisément une partie de ces derniers, les rapports d'activité, étaient susceptibles de convenir. Ces derniers présentent à échéance régulière l'état le plus complet sur les activités de l'entreprise émettrice. Enfin, l'adéquation avec le référentiel d'entreprise mobilisable dans les activités d'information que l'on cherche à documenter a constitué un quatrième critère pour la sélection des textes du corpus. Compte tenu, d'une part, de l'actualité récente du groupe bancaire adopté comme contexte opérationnel, et d'autre part, du caractère nécessairement actuel du référentiel terminologique à construire, la période 1995-2003 a été retenue.

Afin de valider la pertinence de ces choix, il a été nécessaire de documenter finement le type de source retenue à partir de son usage réel et de ses caractéristiques textuelles. Pour atteindre ce but, il a fallu rompre avec une lecture naïve des rapports d'activité, ce qui a permis de situer les enjeux associés par

l'entreprise à l'usage d'un tel outil de communication. En effet, la visée communicationnelle influe nécessairement sur les caractéristiques de toute production discursive. L'étude de celle-ci a permis, d'une part, de vérifier la proximité des finalités existant entre les rapports d'activité et les sites intranets d'un organe central diffusant vers les entreprises entrant dans son giron, et d'autre part, de commencer à cerner les limites d'une telle ressource en ce qui concerne la restitution d'un référentiel d'entreprise, et partant, la constitution d'un référentiel terminologique adaptable au contexte.

L'examen des caractéristiques textuelles des rapports d'activité a permis d'identifier des propriétés dont on doit tenir compte pour le formatage d'un corpus exploitable par des programmes informatiques. C'est ainsi que l'on a pu identifier la présence de deux sortes de contextes aux propriétés textuelles différentes, les contextes à dominante syntactique et non syntactique. Par ailleurs, le découpage de ces textes en rubriques va au-delà de la présence d'un système intertitulaire assez élaboré. Cette structure reflète la nature de ces documents, qui fonctionnent un peu à la manière de meubles à tiroirs : l'information sur l'activité du groupe bancaire y est compartimentée, au point que certaines parties de la version électronique sont données à télécharger séparément sur les sites mettant en lignes les rapports annuels.

L'ensemble de ces remarques apporte un double éclairage : le premier, relatif à l'usage de ces documents dans le contexte de l'entreprise, fournira des clés d'interprétation lorsque l'on étudiera le fonctionnement du vocabulaire des textes du corpus ; le second a permis de justifier le découpage du corpus en différentes parties qui ne sont pas seulement documentaires : découpage en rubriques et caractérisation de ces dernières en fonction de leur appartenance à l'un ou l'autre des deux types de contexte identifiés. Par ailleurs, le suivi de l'évolution des sommaires a été l'occasion d'observer que ce type de texte est assez sensible aux évolutions de l'actualité qu'il sert à restituer. Il conviendrait maintenant de savoir si cette réceptivité des rapports d'activité aux changements qui affectent l'activité de leur émetteur est susceptible d'aller au-delà du remaniement des seules rubriques. En effet, au moment de choisir la période à retenir pour borner le corpus, on a pu noter que le groupe bancaire a connu en une dizaine d'années un certain nombre d'événements majeurs : achat de la banque Indosuez en 1996, cotation de son organe

central en 2001 et, en 2003, rachat du Crédit Lyonnais, groupe bancaire jusqu'alors concurrent du Crédit Agricole. Jusqu'à quel point ces événements ont-ils pu avoir des répercussions sur la situation d'énonciation, l'organisation textuelle des rapports d'activité, ou encore leur vocabulaire ? Ce sont autant de questions auxquelles on tentera de répondre dans le chapitre suivant.

CHAPITRE 3

PREMIÈRE APPROCHE DIACHRONIQUE

Les traitements appliqués à un corpus de textes de manière à le rendre exploitable par un programme textométrique, constitue un préalable nécessaire aux exploitations que l'on va chercher à réaliser dans le cadre de la construction d'un référentiel terminologique. Ce qui est moins souligné ordinairement, et qui pourtant ne constitue pas le moindre des apports de ce type d'approche, c'est que l'on dispose ainsi de la possibilité d'opérer un certain nombre de contrôles de nature à permettre l'identification des traces laissées dans les discours par d'éventuelles modifications de la situation d'énonciation. Certes, il ne s'agit pas de chercher à contrôler tous les paramètres énonciatifs, si tant est que cela soit réaliste ou même souhaitable : « Celle-ci [la comparaison des textes d'un corpus] confronte des énoncés dont les invariants constitutifs doivent de loin l'emporter sur les variables interprétatives. Car que dire sur une comparaison où fluctueraient à la fois les locuteurs, les destinataires, les thèmes, les enjeux, les projets, les genres, les canaux, les dates, les référents, les lieux, les circonstances, etc., bref toutes les causes et conditions de l'énonciation¹ ? » Il reste que, lorsque le corpus est ouvert, et qu'il est par conséquent destiné à être augmenté par l'ajout de nouvelles parties au fil du temps, il paraît utile de contrôler la manière dont les nouveaux textes s'insèrent dans la série textuelle chronologique.

Certes, on pourrait faire l'hypothèse que la situation d'énonciation ne varie pas, au moins pour ce qui est de ses principaux paramètres, durant une période de dix années ; on pourrait également envisager que d'une année sur l'autre les rapports d'activité se contentent de mettre à jour des données chiffrées, l'activité et la manière de la présenter étant stables par ailleurs (c'est du reste le cas, par exemple, pour les années 1950) ; les changements seraient purement cumulatifs, sans répercussion sur les autres aspects demeurés stables ; ils seraient présentés avec une organisation et un vocabulaire identiques à ceux qui étaient utilisés auparavant, moyennant l'apparition probable de quelques néologismes isolés. En définitive, cela reviendrait à considérer que, dans la mesure où le corpus est composé d'un ensemble homogène de discours produits par le même émetteur en direction des mêmes

¹ M. Tournier in P. Charaudeau et D. Maingueneau (2002 : 343).

destinataires sur une période relativement courte, ce sont les traits permanents qui doivent l'emporter. Formulée ainsi, cette hypothèse heurte l'impression qui se dégage des premières observations réalisées dans le précédent chapitre au sujet de l'évolution du découpage des rapports d'activité.

C'est pourquoi on a pris le parti de réaliser des contrôles portant sur des points caractéristiques des textes du corpus : l'organisation textuelle et le vocabulaire pris dans leur globalité. Ces tests consistent à s'assurer que, pour ces deux propriétés essentielles du corpus, la configuration des premières parties ne s'écarte pas de façon trop importante de celle que l'on observe pour les dernières. Si des différences sont mises en évidence, il conviendra alors d'en rechercher la cause, y compris parmi les composantes de base de la situation d'énonciation que sont les interlocuteurs, le message ou le canal. Un tel contrôle suppose que l'on puisse caractériser la norme valant pour le corpus étudié.

On a donc cherché à travailler à l'échelle de grands ensembles textuels. En effet, c'est le meilleur moyen d'avoir une vision globale sur des évolutions qui, sans cela, soit passeraient inaperçues, soit seraient difficiles à évaluer et, partant, à contrôler. C'est cet aspect de l'exploitation d'un corpus que l'on abordera dans les développements qui suivent. On demande, ainsi, aux logiciels de textométrie de fournir une assistance afin de disposer d'une vue d'ensemble sur les données rassemblées dans le corpus. On a vu plus haut que les rapports d'activité possèdent deux sortes de contextes qui ont été pris en compte au moment du balisage. La mise en place d'un corpus de suivi rend nécessaire la surveillance de cette bipartition : la répartition des deux sortes de contextes est-elle stable, chaque type de contexte reste-t-il homogène, ou bien ces deux éléments varient-ils dans le temps ? D'autre part, on vérifiera que les différentes parties du corpus s'inscrivent dans le comportement attendu pour une série textuelle chronologique, dans laquelle le vocabulaire se renouvelle de façon régulière. Enfin, à l'occasion de cette étude globale du vocabulaire du corpus, on cherchera à identifier la position occupée par les noms propres, afin de repérer la présence d'éventuels indices de leur participation à la restitution des évolutions affectant un référentiel.

I LE CONTRÔLE DES CONTEXTES OU LA TECTONIQUE DES DISPOSITIFS TEXTUELS

A. Mise en place d'un test textométrique

1. Rappel des traitements textométriques fondamentaux

Le premier de ces contrôles porte sur la stabilité des deux sortes de contextes identifiés dans le chapitre précédent et marqués à l'aide de balises spécifiques lors de la préparation du corpus. On a vu plus haut que la qualification d'une rubrique en contexte à dominante syntactique ou non syntactique a été réalisée manuellement. Dans la plupart des cas, cette caractérisation a pu se faire sans ambiguïté, dans la mesure où la rubrique correspond fréquemment, soit à du texte suivi, soit aux sortes de contextes que l'on a tenté de décrire plus haut (organigrammes, histogrammes, tableaux et listes de divers genres). Néanmoins, le corpus s'enrichissant chaque année d'une nouvelle partie, on a pu constater que l'arbitrage pouvait devenir de plus en plus difficile. En effet, on a pu relever que certains contextes syntactiques ont tendance à admettre de plus en plus de graphiques ou de résumés donnés dans une sorte de style télégraphique. Il est donc apparu nécessaire de mettre en place un test de manière, d'une part, à pouvoir fiabiliser dans la durée la qualification des rubriques, et d'autre part, à prendre la mesure de cette évolution afin d'en esquisser une interprétation.

Dans le précédent chapitre, on a présenté les opérations essentielles réalisées par les logiciels de textométrie, qui sont le contrôle de la validité du formatage du corpus et la segmentation des textes en unités simples, les occurrences des formes graphiques. Le résultat immédiatement visible de ce traitement est la production d'un dictionnaire des formes du corpus, chaque forme étant documentée par sa fréquence observée (nombre d'occurrences) et donc son rang textométrique.

Ces étapes préliminaires rendent possible de nouvelles observations s'appuyant sur le découpage du corpus en différentes parties. On a vu dans le chapitre précédent que les rapports d'activité constituent l'unité documentaire de base du corpus. Cette unité, tout comme les rubriques et les types de contexte, a fait

l'objet d'un balisage spécifique. Celui-ci est ordinairement destiné à observer les variations qui affectent la ventilation des occurrences selon la partition utilisée. C'est, du reste, ce que l'on fera dans les chapitres suivants. Mais cela ne signifie pas pour autant que ces découpages soient tout à fait neutres, au sens où, pour la période couverte par le corpus, on pourrait considérer qu'il s'agit d'une variable négligeable. Il a donc paru nécessaire d'observer également la manière dont ces éléments ont évolué entre le début et la fin du corpus. On donne ci-dessous, à titre d'illustration, une présentation des formes les plus fréquentes du corpus principal de rapports d'activité dans un index et dans le tableau lexical entier.

*Tableau 11.
Premières formes du corpus « Rapports d'activité » classées par ordre
textométrique de fréquence*

Forme	Fréquence
de	12744
et	4486
la	4244
des	3887
en	3641
d	3438
à	3119
le	3024

Lecture du tableau 11 : dans la colonne de droite figurent les formes graphiques les plus fréquentes du corpus « Rapports d'activité », triées par ordre textométrique, c'est-à-dire par ordre décroissant de fréquences. Par exemple, on a pour la forme « de » une fréquence de 12744 occurrences attestées dans le corpus.

Le tableau lexical entier (TLE) rassemble les principales données qui vont être utilisées pour la suite des traitements, que ceux-ci soient orientés dans une perspective textuelle ou lexicale. Il s'agit d'un « *tableau à double entrée dont les lignes sont constituées par les ventilations des différentes formes dans les parties du corpus. Le terme générique $k(i,j)$ du TLE est égal au nombre de fois que la forme i est attestée dans la partie j du corpus. Les lignes du TLE sont triées selon l'ordre*

*lexicométrique des formes correspondantes*¹. » La fréquence globale, c'est-à-dire l'addition de toutes les occurrences d'une forme graphique à l'échelle du corpus pris dans son ensemble, est ainsi complétée par celle qui peut être détaillée en fonction de la partition retenue pour le corpus.

Tableau 12.

Extrait du tableau lexical entier présentant les formes les plus fréquentes dans le corpus « Rapports d'activité », ventilées sur les neuf parties principales

Forme	P1	P2	P3	P4	P5	P6	P7	P8	P9	Total
de	1140	1304	1417	1182	1347	1321	1338	1260	2435	12744
et	336	396	458	337	452	420	555	562	970	4486
la	382	473	457	417	445	436	418	394	822	4244
des	334	397	467	347	385	343	369	433	812	3887
en	272	341	368	342	335	338	400	442	802	3640
d	230	247	308	225	339	391	431	419	848	3438
à	308	311	376	326	319	317	254	298	610	3119
le	296	322	328	285	340	367	250	268	568	3024

Lecture du tableau 12 : dans la première colonne figurent les premières formes du corpus « Rapports d'activité », classées par ordre décroissant de fréquence. Les colonnes nommées P1 à P9 correspondent aux neuf parties principales du corpus, soit les neuf rapports d'activité rangés de gauche à droite dans l'ordre chronologique, de 1995 à 2003. Dans chaque case du tableau, les chiffres correspondent à la fréquence observée pour une forme dans telle partie. La colonne « total » présente la somme des fréquences d'apparition d'une forme dans chaque partie du corpus. On retrouve la fréquence totale donnée pour chaque forme dans la liste d'index. Par exemple, on a pour la forme « de » une fréquence de 1140 occurrences dans la partie P1 et de 1338 occurrences dans la partie P7.

C'est à partir des données du TLE qu'il est possible de disposer d'une première vision synthétique du corpus en fonction de la partition retenue, comme cela a déjà été indiqué dans le précédent chapitre, où les principales caractéristiques textométriques du corpus ont été données. Les données statistiques obtenues à l'aide de ces traitements fondamentaux vont permettre de réaliser les contrôles évoqués plus haut.

¹ L. lebart et A. Salem (1994 : 318) :

2. Principales caractéristiques textométriques de la partition CDN - CDS

Dans la mesure où elle est réalisée manuellement et à des moments espacés dans le temps, compte tenu de la périodicité annuelle des mises à jour du corpus, l'opération manuelle de balisage doit pouvoir être guidée par une vue d'ensemble des contextes à l'échelle du corpus. La mise en place d'indices textométriques a donc été retenue. Ces derniers, s'appuyant dans un premier temps sur les caractéristiques constatées en général pour les deux sortes de contextes, doivent ensuite permettre de contrôler, pour chaque rubrique, si elle a été correctement qualifiée de CDS ou de CDN.

Les caractéristiques textométriques de cette partition permettent d'opposer les deux sortes de contexte sur des plans différents.

Tableaux 13.

Principales caractéristiques textométriques du corpus « Rapports d'activité » en fonction de la partition par type de contexte

Parties	Nb occurrences	Nb formes	Nb hapax	Fréq. Max	Forme
CDN	37087	3776	1288	1171	de
CDS	164817	9180	3667	11573	de

Lecture du tableau 13 : dans la première colonne figurent les deux parties principales du corpus partitionné en fonction des deux types de contextes, CDN et CDS. Chacun de ces contextes regroupe toutes les rubriques du corpus qui ont été caractérisées, soit comme relevant d'un contexte à dominante non syntactique (CDN), soit d'un contexte à dominante syntactique (CDS). Pour chaque partie, les colonnes suivantes présentent, par ordre d'apparition : le nombre d'occurrences attestées, le nombre de formes graphiques, le nombre d'hapax, la fréquence la plus élevée et la forme qui possède cette fréquence dans la partie du corpus correspondante.

Les principales caractéristiques textométriques de la partition entre CDN et CDS révèlent des différences importantes entre les deux sortes de contextes. Il s'agit tout d'abord de simples différences quantitatives indiquant que les deux parties sont de tailles inégales. L'ensemble des CDS possède en effet de quatre à cinq fois plus d'occurrences que l'ensemble des CDN ; ce rapport passe d'un à dix si l'on considère la fréquence la plus importante constatée dans chaque contexte. Mais si l'on rapporte la fréquence de la forme « de » au nombre d'occurrences présentes

dans chaque contexte, on identifie, outre la différence quantitative, une différence de rapports. Alors que dans les CDN ce rapport donne, pour une occurrence de la forme « de », plus de trente-deux occurrences d'unités diverses, il tombe à quatorze occurrences dans les CDS, soit deux fois moins. Bien que la forme « de » soit la plus fréquente dans chaque partie, sa plus faible présence relative dans les CDN vient confirmer la caractéristique principale de ces contextes, à savoir le regroupement d'unités linguistiques sorties du contexte de la phrase. On sait en effet qu'à contrario, les morphèmes grammaticaux sont « *des signes d'opérations linguistiques servant à construire des relations entre unités linguistiques*¹. » L'opposition des deux sortes de contextes autour de la forme « de » est confirmée par le calcul des spécificités dont le résultat est présenté dans la figure ci-dessous.

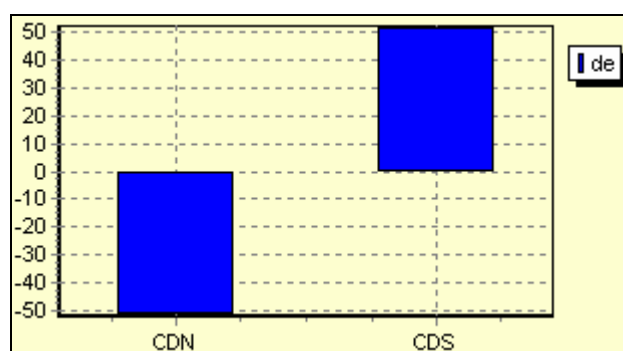


Figure 13.

Spécificités de la préposition « de » en fonction de la partition du corpus entre CDN et CDS

Lecture de la figure 13 : le calcul des spécificités permet de mettre en évidence la sous-représentation de la forme « de » dans les contextes de type non syntactique et sa surreprésentation dans les contextes à dominante syntactique. Si ce coefficient avait été égal à ou proche de zéro, cela aurait en revanche indiqué la banalité de la forme, autrement dit, une répartition comparable dans chaque type de contexte.

L'examen du tableau des principales caractéristiques textométriques de cette partition du corpus suggère l'existence d'une autre opposition entre les deux contextes. En effet, tout semble se passer comme si la pauvreté relative des CDN en

¹ I. Tamba (2005 : 79).

morphèmes grammaticaux était compensée par une relative richesse en unités d'une autre sorte. Si l'on rapporte pour chaque contexte le nombre d'occurrences à celui des hapax, on constate que, pour une forme présente dans les CDN, ce rapport est de un à trente, alors qu'il s'élève à un contre quarante-cinq en CDS. Les CDN seraient donc également riches en formes rares. On peut alors chercher à caractériser de façon plus détaillée le vocabulaire propre à chaque type de contexte.

3. Caractéristiques des vocabulaires des CDN et des CDS

a. Formes spécifiques des deux sortes de contextes

À partir du balisage déjà réalisé pour le corpus, on a donc cherché à établir quelle sorte de vocabulaire est le plus caractéristique de chaque sorte de contexte pris dans son ensemble. Pour cela, on a appliqué le calcul des spécificités¹ aux deux parties du corpus regroupant toutes les rubriques, selon qu'elles ont été caractérisées comme relevant d'un CDS ou d'un CDN. Ce calcul permet de comparer les stocks de vocabulaire des deux ensembles CDN et CDS qui correspondent, dans ce cas, aux deux parties du corpus. Le calcul des spécificités révèle la présence d'unités caractéristiques nettement différentes dans chacune des deux parties, indiquant par là que les différences de présentation observées lors du repérage manuel sont doublées d'une différence de vocabulaire.

¹ Ce calcul permet de déterminer les spécificités négatives (formes particulièrement peu fréquentes) ou positives (formes particulièrement fréquentes) pour une partie du corpus. Seules les fréquences positives ont été retenues pour ce test, car les deux parties étant prises en compte, les formes les plus caractéristiques de l'une sont également celles qui sont les moins présentes dans l'autre, et réciproquement. On reprend la définition des spécificités positives données dans L. Lebart et A. Salem (1994 : 318) : « pour un seuil de spécificité fixé, une forme *i* et une partie *j* données, la forme *i* est dite spécifique positive pour la partie *j* (ou forme caractéristique de cette partie) si sa sous-fréquence est « anormalement élevée » dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ. » Le seuil de probabilité retenu ici est de 5 et la fréquence minimale est de 10 pour le nombre d'occurrences.

*Tableaux 14 et 15.
Les dix premières formes les plus spécifiques des deux parties CDN et CDS du
corpus de rapports d'activité*

Formes spécifiques CDN			
Forme	Frq. Tot.	Fréquence	Coeff.
avenue	79	79	***
fonds	567	339	***
MF	663	662	***
propres	453	361	***
locales	484	358	***
Général	485	467	***
Indosuez	827	404	***
Paul	86	85	***
5	86	83	***
Total	688	675	***

Formes spécifiques CDS			
Forme	Frq. Tot.	Fréquence	Coeff.
un	1269	1261	***
par	1219	1176	***
sur	1209	1173	***
%	1500	1429	***
une	1443	1433	***
dans	1330	1301	***
avec	1004	987	***
qui	643	640	***
francs	625	624	***
son	601	600	***

Lecture des tableaux 14 et 15 : chaque tableau présente les dix premières formes des parties possédant le coefficient de spécificité positif le plus élevé (toutes ces formes ont un coefficient supérieur à 50, ce qui est indiqué ici par les « *** »). Pour chaque tableau, et pour chaque élément de la colonne « Forme », les deux colonnes suivantes donnent, dans l'ordre, la fréquence totale observée dans l'ensemble du corpus (Frq. Tot.), et celle qui est observée dans la partie concernée (Fréquence). Par exemple, dans l'ensemble des formes spécifiques de la partie CDS, la forme « un », qui est dotée d'un coefficient de spécificité très élevé puisque supérieur à 50, possède 1269 occurrences dans l'ensemble du corpus et 1261 occurrences pour la seule partie CDS.

Le test textométrique que l'on cherche à établir aurait pu s'appuyer sur ces seules données, mais on a préféré renforcer la force du contraste qu'il doit permettre de révéler entre les deux types de contextes. Dans ce but, on a retenu, parmi les formes caractéristiques de chaque contexte, celles qui possèdent deux propriétés supplémentaires : être présentes dans l'ensemble du corpus et être tout particulièrement représentatives de chacune des parties. En effet, la présence de la forme dans l'ensemble du corpus permet de neutraliser, en partie, la variable diachronique. D'autre part, le rassemblement de formes qui, à défaut d'être

originales¹, peuvent être qualifiées de typiques, vise à accentuer l'opposition entre les deux sortes de contextes, de manière à la rendre plus lisible et plus facilement interprétable.

La ventilation des formes peut être mise en évidence de façon graphique. Les données du TLE qui sont alors utilisées permettent de repérer rapidement les formes spécifiques de chaque contexte qui répondent, en outre, à l'exigence d'être présentes dans toutes les parties du corpus :

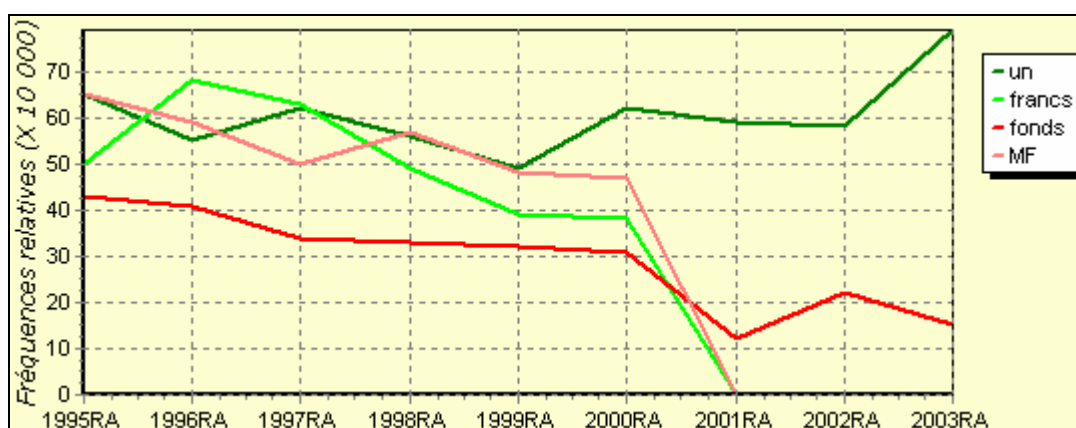


Figure 14.
Fréquences relatives de quatre formes des contextes CDS et CDN

Lecture de la figure 14 : présentation de la ventilation de deux des formes les plus caractéristiques des CDS « un » et « francs » et deux des formes les plus caractéristiques des CDN « fonds » et « MF » sur toutes les parties principales du corpus (notées 1995RA à 2003RA et correspondant chacune à un rapport d'activité). La fréquence relative correspond à la fréquence de chaque forme dans chaque partie du corpus rapportée à la taille du corpus.

Ce graphique permet de mettre en évidence le fait que des formes rassemblées en fonction de leur spécificité ne possèdent pas pour autant la même ventilation dans le corpus. Alors que la forme « un », particulièrement caractéristique des CDS, traverse l'ensemble de la période couverte par le corpus, la forme « francs » disparaît brutalement en 2001. On constate d'autre part que le

¹ Forme originale : « (pour une partie du corpus) forme trouvant toutes ses occurrences dans cette seule partie. » L. Lebart et A. Salem (1994 : 313).

même type de variations affecte également les contextes à dominante non syntactique. Ce premier repérage permet donc d'éliminer toutes les formes caractéristiques de deux sortes de contextes qui ne sont pas réparties sur l'ensemble du corpus.

En examinant de plus près le tableau précédent, on remarque également que certaines formes ayant un coefficient de spécificité équivalent ne sont pas pour autant présentes dans les mêmes proportions dans une partie donnée. Le tableau ci-dessous met en évidence des disparités importantes entre formes spécifiques, en rapportant la présence dans la partie à la présence dans l'ensemble du corpus.

*Tableau 16.
Deux formes caractéristiques des CDN possédant un rapport fréquence CDN sur fréquence totale très différent*

Forme	Frq. Tot.	Fréquence CDN	Fréquence / Frq. Tot
fonds	567	339	60%
avenue	79	79	100%

Lecture du tableau 16 : ce tableau est un extrait du tableau « Formes spécifiques des CDN ». La fréquence totale observée pour une forme dans le corpus et la fréquence attestée pour la partie CDN ont été complétées par le rapport de la seconde sur la première, exprimé en pourcentage. Par exemple, la forme « avenue » est une forme originale de la partie CDN dans la mesure où sa fréquence locale CDN est égale à sa fréquence totale dans l'ensemble du corpus, soit 79 sur 79 (soit un rapport égal à 1 ou à 100%).

Le tableau 16 permet de mettre en évidence une autre différence pouvant exister entre formes dont le coefficient de spécificité est proche ou identique : alors que dans certains cas une forte spécificité correspond également à l'originalité de la forme, ou tout au moins à une forte tendance à l'originalité, dans d'autres, en revanche, l'écart peut s'avérer substantiel, puisque la partie peut ne posséder que les trois cinquième des occurrences de la forme (c'est le cas de « fonds »). En fait, ce constat ne fait que formuler différemment ce qui est mis en évidence par le calcul des spécificités : rendre compte de l'importance relative d'une forme dans une partie donnée du corpus, compte tenu de la taille du corpus. Autrement dit, le calcul des spécificités révèle que, par exemple, compte tenu de la taille de la partie CDN, la

présence des 339 occurrences de la forme « fonds » y est particulièrement remarquable, alors qu'une quantité voisine d'occurrences est beaucoup plus « diluée » dans la partie CDS, et donc moins caractéristique. Cependant, on a vu que le résultat recherché est, à partir du dégrossissement fourni par le calcul des spécificités, la constitution d'un ensemble de formes « typiques » des deux sortes de contextes, syntactiques et non syntactiques.

b. Formes typiques des deux sortes de contextes

On a donc retenu parmi les formes les plus caractéristiques de chaque contexte celles qui, en plus d'une présence avérée sur l'ensemble du corpus, présentent un ratio fréquence locale sur fréquence totale très proche de cent pour cent. En suivant cette procédure, on obtient les deux listes suivantes qui regroupent, pour chaque contexte, des formes ayant un coefficient de spécificité très élevé, une présence sur l'ensemble du corpus et un ratio de fréquence également très élevés :

Tableaux 17 et 18.

Les dix formes les plus typiques des deux contextes CDS et CDN du corpus des rapports d'activité

Formes typiques CDN					Formes typiques CDS				
Forme	Frq. Tot.	Frq.	Coeff.	Ori.	Forme	Frq. Tot.	Frq.	Coeff.	Ori.
Paul	86	85	***	99%	un	1269	1261	***	99%
Pierre	169	168	***	99%	une	1443	1433	***	99%
Tél	449	444	***	99%	qui	643	640	***	100%
François	73	73	***	100%	son	601	600	***	100%
Jacques	74	74	***	100%	est	934	920	***	99%
Bernard	89	88	***	99%	milliards	828	819	***	99%
Michel	125	124	***	99%	ont	726	723	***	100%
Alain	60	60	45	100%	a	2275	2256	***	99%
Henri	52	52	39	100%	marché	588	586	49	100%
Louis	37	37	28	100%	sa	585	583	49	100%

Lecture des tableaux 17 et 18 : les tableaux 17 et 18 présentent les dix formes les plus spécifiques des contextes CDN et CDS (coefficient de spécificité positif égal ou supérieur à 28 pour les CDN et égal ou supérieur à 49 pour les CDS), caractérisées, en outre, par une très forte originalité (la colonne « Ori » pour « originalité » présente le résultat arrondi de la fréquence de la forme dans la partie étudiée rapportée à sa fréquence totale dans le corpus), et une présence dans toutes les parties du corpus, lorsque celui-ci est divisé selon les neuf parties de la

série textuelle chronologique (voir figure 14). On a proposé de qualifier ces formes de « typiques » du fait de leurs propriétés particulières.

Les deux filtres complémentaires du calcul des spécificités (appartenance au vocabulaire commun et typicité), ont eu pour effet de produire deux ensembles assez homogènes possédant chacun un vocabulaire très particulier. Avant d'en entreprendre le commentaire, on peut comparer les listes ainsi obtenues avec celles qui figurent dans les tableaux 14-15 des formes spécifiques obtenues sans filtrage complémentaire. On constate alors une différence dans la manière dont les filtres ont opéré. Pour ce qui est des CDS, le filtrage a eu pour effet d'éliminer certaines formes spécifiques, tout en conservant à la liste un même air de famille (place importante des morphèmes grammaticaux). En revanche, la comparaison des deux listes CDN, avant et après filtrage, fait ressortir que les critères de sélection complémentaires ont eu pour effet d'homogénéiser une liste qui paraissait hétéroclite.

Cette différence entre les deux listes semble pouvoir être attribuée à deux facteurs principaux. Le premier correspond à une plus grande sensibilité des formes typiques des CDN aux phénomènes extérieurs. C'est ainsi que l'on voit, par exemple, qu'à partir d'un certain moment, les sommes présentées dans les rapports d'activité ne sont plus notées à l'aide de l'abréviation « MF » pour « millions de francs », ou bien encore, pour une partie du corpus, les coordonnées des Caisses régionales ne sont plus mentionnées, ce qui a pour conséquence de faire disparaître momentanément la forme « avenue ». L'autre facteur explicatif réside dans le fait que les formes spécifiques des CDN sont fréquemment présentes dans l'autre type de contexte. La forme « Indosuez » est à cet égard exemplaire : elle est à la fois très caractéristique des CDN et mieux représentée en termes d'occurrences dans les contextes CDS. Les résultats des filtres éclairent donc sous un angle nouveau la subdivision des parties principales du corpus (les rapports d'activité), en deux types de contextes : les CDN possèdent bien un vocabulaire propre, mais non exclusif et qui plus est, particulièrement exposé aux divers remaniements que les données peuvent subir dans cet environnement.

Les filtres complémentaires ont eu pour conséquence d'accentuer certaines caractéristiques moins visibles à partir des seules spécificités. Pour ce qui est des

unités dont se compose l'ensemble des formes caractéristiques des CDS, elles sont à rattacher au groupe des constituants de la phrase de base¹ et de ses extensions (les déterminants indéfinis « un » et « une », les déterminants possessifs « son » et « sa », les trois particules verbales « est », « ont » et « a » ainsi que les deux substantifs (« milliards » et « marché ») et le pronom relatif « qui ». Le groupe des formes typiques des contextes à dominante non syntactique présente un caractère tout opposé : neuf prénoms et l'abréviation « Tél. », pour « téléphone » ont été réunies. Cette surreprésentation des noms propres met en évidence une propriété partagée par tous les substantifs. En effet, ceux-ci peuvent être utilisés en dehors de la phrase du fait que, même sortis de ce contexte, ils continuent à désigner une classe d'objets (nom commun) ou un particulier (nom propre). Qu'ils puissent être utilisés sans déterminants dans des environnements textuels les prédispose à occuper une position centrale dans les CDN. On trouve là une confirmation de l'intérêt qu'il y a à prendre en compte ce type de contexte pour en extraire des données susceptibles d'alimenter un référentiel terminologique. Quant à la forme « Tél », abréviation de « téléphone », elle ne peut se trouver que dans des contextes particuliers, où il s'agit de donner les coordonnées téléphoniques de personnes physiques ou morales, selon un schéma du type « nom + adresse + coordonnées téléphoniques ». Sa présence suppose donc qu'une entité soit nommée dans son environnement cotextuel immédiat, ce qui en fait un auxiliaire possible de la catégorie des noms propres dont il vient d'être question. Une fois constituées, les deux listes de formes typiques peuvent être utilisées comme critères afin de tester la stabilité de la partition du corpus entre contextes à dominante syntactique et non syntactique.

B. Mise en œuvre du test textométrique de contrôle des contextes

1. Contrôle du découpage sur la carte des sections de Lexico

Pour procéder au contrôle de la stabilité et de l'homogénéité de la partition du corpus en rubriques de type CDN ou CDS, ou pour décider de l'appartenance de nouvelles rubriques à l'un des deux types de contexte, on dispose maintenant

¹ La phrase de base correspond à une structure canonique, susceptible de rendre compte de la diversité des phrases observables. Sur ce point, on renvoie à M. Riegel *et alii* (1998 : 108 et suiv.).

d'indices facilement exploitables. En effet, il suffit de sélectionner les formes les plus caractéristiques de l'un des deux contextes, et de vérifier leur présence dans les paragraphes¹ dont sont constituées les rubriques des différentes parties du corpus. Dans un premier temps, on forme deux groupes de formes rassemblant les dix formes les plus typiques des contextes à dominante syntactique et non syntactique, déjà présentées dans les tableaux 17 et 18 ci-dessus. Il s'agit pour les CDS de « un », « une », « qui », « son », « est », « milliards », « ont », « a », « marché » et « sa », et de « Paul », « Pierre », « Tél », « François », « Jacques », « Bernard », « Michel », « Alain », « Henri », « Louis », pour les CDN. Ces groupes de formes sont ensuite projetés sur la carte des sections de Lexico 3. Un premier regroupement permet de voir la présence des formes « test » dans les paragraphes des neuf parties principales du corpus. On compare ensuite cette projection avec celle obtenue pour la même année, mais avec la partition tenant compte de l'appartenance des rubriques à l'une ou l'autre des deux sortes de contextes. On dispose de la sorte d'une présentation graphique du corpus dans laquelle il devient facile d'identifier les paragraphes répondant positivement au test et ceux qui, bien qu'étant rattachés à un contexte, comportent au moins l'une des formes caractéristiques de l'autre contexte.

¹ On aurait pu retenir la phrase comme unité de base. Cependant, cette unité n'est pas pertinente dès lors que l'on s'attache également à l'étude des contextes non syntactiques. On a vu précédemment que pour ces contextes, l'unité de sens la plus fine semble plutôt correspondre au paragraphe. Ce dernier découpage étant également présent dans les contextes syntactiques, il a donc été retenu. On ajoutera que le repérage des phrases à l'aide du point final, même avec une analyse morpho-syntaxique, est source d'erreur, dans la mesure où le point est utilisé dans bien d'autres cas que pour la fermeture des phrases (on trouve dans le corpus un nombre important d'utilisations du point, environ 10% des cas, pour séparer les chiffres d'un numéro de téléphone ou les lettres d'un sigle, comme par exemple dans « Indosuez W.I. Carr Securities (Pakistan) PTE Ltd » ou « Crédit Agricole S.A. »).

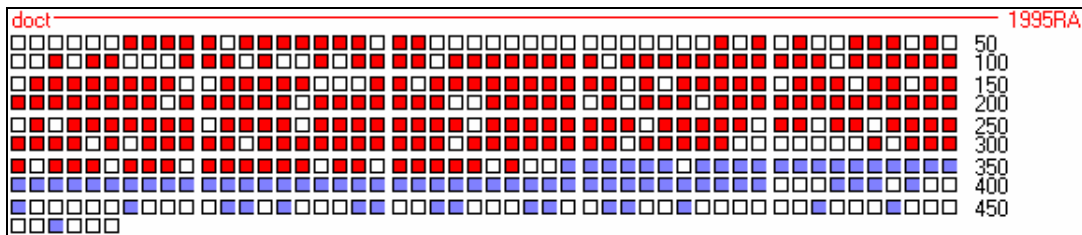


Figure 15.

Carte des sections (paragrapes) de l'année 1995 avec projection des formes des CDS (rouge) et des CDN (bleu)

Lecture de la figure 15 : la carte des sections de Lexico 3 permet de représenter à l'aide de carrés un type de section particulier présent dans le corpus. Dans cette carte, chaque carré représente un paragraphe positionné dans l'ordre exact dans lequel il apparaît dans le texte d'origine. Lorsqu'un paragraphe est coloré en rouge, c'est qu'il contient l'une des formes caractéristiques des contextes de type syntactique. Un paragraphe est coloré en bleu lorsqu'il contient au moins l'une des formes caractéristiques des contextes non syntactiques. Les paragraphes apparaissant en blanc ne contiennent aucune des formes les plus caractéristiques des deux sortes de contextes. La mention « 1995RA » indique la partie du corpus à laquelle les paragraphes appartiennent. On distingue un premier bloc d'environ 325 paragraphes à dominante rouge, suivi d'un deuxième bloc de 125 paragraphes environ où domine le bleu.

Pour commenter cette figure, on peut partir des carrés vides qui semblent infliger un démenti à la pertinence du test, du moins en ce qui concerne son caractère discriminant. En réalité, une première catégorie de sections vides correspond à des paragraphes techniques destinés à séparer correctement le texte d'avec les balises utilisées pour le repérage d'autres partitions (rubriques et contextes). Ces paragraphes artificiels se remarquent bien dans la première partie de la carte (carrés colorés en rouge), où l'on voit des carrés blancs isolés à intervalles assez réguliers. Il reste que des plages de carrés beaucoup plus importantes doivent être expliquées.

Dans la zone majoritairement rouge, on remarque au moins trois plages importantes (comportant plus de trois carrés vides entre deux carrés rouges), que le test invite à examiner. En fait, ces trois plages correspondent respectivement au sommaire, à l'organigramme des entités du groupe Crédit agricole (ligne « 50 ») et aux coordonnées des succursales et bureaux de représentation du même groupe

(ligne « 300 »). Autrement dit, ces plages correspondent en fait à des contextes de type non syntactique. De ce point de vue, le test a fonctionné par défaut, mais ces plages auraient pu être colorées en bleu, au moins partiellement.

La zone à dominante bleue donne l'image d'une répartition de nature un peu différente. Elle offre soit une succession ininterrompue de carrés colorés (ce qui indique une rubrique de taille importante, puisque la présence d'autres rubriques serait signalée par des carrés blancs réservés aux balises), soit de grands espacements entre carrés bleus, au point que ceux-ci sont en minorité parmi les carrés blancs. La série ininterrompue qui court sur les lignes « 350 » et « 400 » correspond à une succession de présentations signalétiques donnant les coordonnées de diverses entités du Groupe Crédit Agricole. C'est la forme « Tél » qui est responsable d'un repérage que la diversité des dénominations aurait sinon rendu impossible (celles-ci excèdent largement la taille de l'échantillon de test). En revanche, la dispersion qui caractérise la ligne « 450 » est bien à imputer aux caractéristiques de la liste utilisée pour le test. Étant donné que les formes représentatives des CDN sont en général moins fréquentes que celles qui caractérisent les CDS, l'échantillon retenu pour la liste est voué de ce fait à être moins représentatif. On peut en donner une illustration à l'aide des dix derniers paragraphes de la ligne « 450 ». Ceux-ci sont répartis de la façon suivante : deux paragraphes colorés en bleu pour huit paragraphes vides. Tous les paragraphes vides comportent la même nature d'information que ceux qui sont colorés en bleu : dans tous les cas, il s'agit de la mention de l'un des membres de la direction de l'entreprise, accompagné de son titre. La différence de coloration peut donc être imputée au fait que le prénom de la personnalité qui est présentée dans le paragraphe non coloré est moins fréquent que ceux qui figurent dans la liste utilisée pour le test. Ce seul exemple suffit à montrer que le test comporte une limite : il est, d'une manière générale, mieux adapté aux contextes syntactiques qu'aux contextes non syntactiques, ces derniers comportant proportionnellement un nombre plus important de formes relativement peu fréquentes pour lesquelles un échantillon représentatif peut difficilement être mis en place. Cette faiblesse doit néanmoins être relativisée : le test permet au moins d'identifier par défaut les CDN dont les

propriétés lexicales contrastent fortement avec les CDS. C'est ce que met en évidence la partition du corpus en CDN et CDS répartis par année.

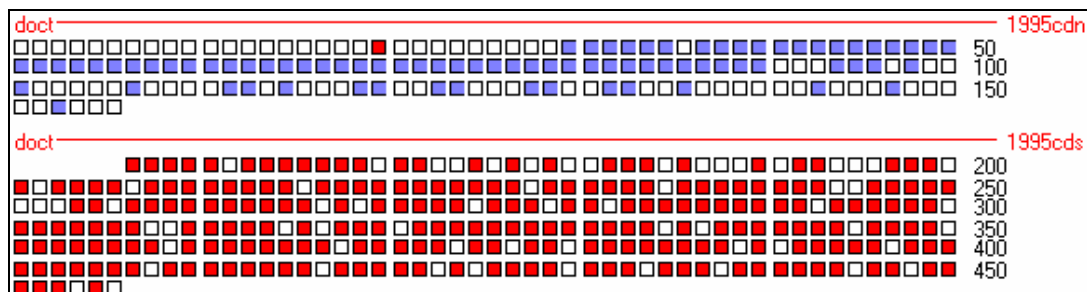


Figure 16.

Carte des sections (paragraphes) de l'année 1995 partitionnée en fonction des contextes CDN et CDS et sur laquelle les formes les plus typiques des deux contextes ont été projetées

Lecture de la figure 16 : comme pour la précédente figure, chaque carré représente un paragraphe. Par ailleurs, les rubriques ont été organisées cette fois selon deux critères : d'une part, leur ordre d'apparition et, d'autre part, leur appartenance à l'une ou l'autre des deux sortes de contextes. Lorsqu'un paragraphe est coloré en rouge, c'est qu'il contient l'une des formes caractéristiques des contextes de type syntactique. Un paragraphe est coloré en bleu lorsqu'il contient au moins l'une des formes caractéristiques des contextes non syntactiques. Les paragraphes apparaissant en blanc ne contiennent aucune des formes les plus caractéristiques des deux sortes de contextes. Les mentions « 1995cdn » et « 1995cds » indiquent que les rubriques de la partie 1995RA du corpus ont été regroupées selon qu'elles appartiennent à un CDN ou à un CDS. Par rapport à la figure précédente, les colorations sont inversées : cela est dû au fait que les regroupements en CDN et CDS se font en fonction du type du premier contexte. La partie 1995RA débutant par un CDN, ce sont les paragraphes appartenant aux rubriques de type CDN qui apparaissent les premiers (colorés en bleu).

Le regroupement des rubriques de type CDN et CDS permet de clarifier la répartition des paragraphes. En effet, on remarque que la coloration de ces derniers s'accorde globalement avec la caractérisation manuelle des contextes réalisée au moment du formatage du corpus. On note une exception (le seul paragraphe coloré en rouge dans le regroupement CDN), qui est due à la présence du mot « milliards » destiné à caractériser les montants affichés dans un histogramme (« en milliards de francs »). Cette exception n'est pas la seule. Dans la partie CDS, dont l'homogénéité n'est plus perturbée par l'inclusion de rubriques de type CDN, il demeure de petites plages, peu nombreuses (on dénombre dix-sept paragraphes sur trois cents), de un à

trois paragraphes vides (les autres paragraphes vides, une cinquantaine, correspondent à l'encadrement des balises). Ce phénomène peut être expliqué par la brièveté de la liste utilisée pour le test ainsi que par le caractère typique des formes qu'elle recense. On admet donc que le contraste obtenu de cette manière puisse produire artificiellement quelques exceptions.

2. Ventilation des formes typiques des deux sortes de contextes

Il reste maintenant à savoir si la configuration que l'on observe pour cette première partie du corpus se maintient chez les autres, et si le test permet en outre de signaler d'éventuelles évolutions du traitement discursif. Une manière de parvenir à cette information consiste à observer, sur toute la durée du corpus, les variations affectant la spécificité des formes du test en fonction du type de contexte. Si cette variation est nulle ou faible, on pourra en déduire que, de la première partie du corpus à la dernière, la subdivision entre CDS et CDN est stable dans la durée, et que les propriétés de chaque sous-partie demeurent homogènes. En revanche, si l'on constate des variations dans les spécificités, il sera nécessaire de pousser plus loin les investigations de manière à identifier la cause de ce phénomène. Il peut s'agir, par exemple, d'une erreur de balisage – ce test a précisément été construit pour permettre la détection de telles erreurs – mais aussi d'une évolution du traitement discursif dont il conviendra alors de rendre compte.

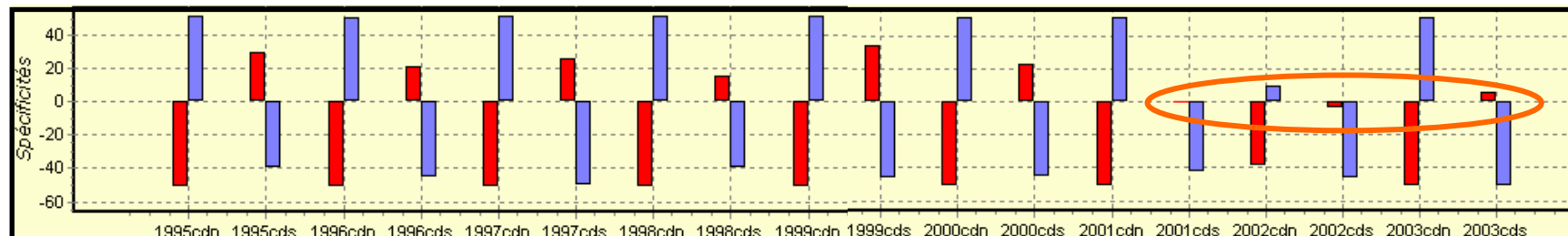


Figure 17.
Spécificités des formes appartenant aux deux groupes de formes typiques des CDS (en rouge) et CDN (en bleu) ventilées sur l'ensemble du corpus partitionné en années-types de contextes.

Lecture de la figure 17 : ce graphique présente la ventilation des spécificités des deux groupes de formes typiques sur toute l'étendue du corpus. Le groupe de formes CDS est représenté en rouge et le groupe CDN en bleu. Chaque partie (ou année) du corpus comporte un élément CDN (regroupement des rubriques de type CDN) et un élément CDS (regroupement des rubriques de type CDS).

Si l'on prend l'année 2003 comme exemple, le graphique se lit de cette manière : pour la partie 2003cdn on note une surreprésentation des formes typiques des CDN et une sous-représentation des formes typiques des CDS. De même, pour la partie 2003cds, on note une faible représentation des formes typiques des CDS et une sous-représentation des formes typiques des CDN. On constate que l'opposition n'est pas symétrique : alors que l'on aurait pu s'attendre à ce que les formes typiques des CDS soient nettement surreprésentées dans les contextes syntactiques de l'année 2003, on constate seulement une faible représentation. Les trois autres indications fournies sur la spécificité des formes des deux groupes correspondent en revanche aux attentes en termes de résultats produits par le test (dans la partie 2003cdn, formes typiques CDN surreprésentées et formes typiques CDS sous-représentées ; dans la partie 2003cds, formes typiques CDN sous-représentées).

Ce graphique permet d'obtenir une vision d'ensemble de la manière dont les formes typiques de chaque contexte se répartissent dans la durée, en fonction des contextes établis par balisage pour chaque partie du corpus. Compte tenu du fait que ces formes sont issues d'un calcul ayant porté sur l'ensemble du corpus, on aurait pu s'attendre à ce que leur répartition soit à peu près identique dans chaque partie. L'impression générale fournie par la lecture du tableau ne semble pas aller tout à fait dans ce sens. Il convient cependant de distinguer les résultats obtenus pour les formes typiques de chaque contexte.

Pour les formes typiques des CDN, la ventilation des spécificités indique une répartition relativement constante pour l'ensemble du corpus : ces formes sont toujours surreprésentées dans les contextes CDN et sous-représentées en CDS. Ce résultat attendu est en quelque sorte renforcé par l'indication selon laquelle cette répartition se fait pratiquement toujours dans les mêmes proportions dans les différentes parties du corpus, à l'exception de l'année 2002. Pour 2002, ces formes sont moins bien représentées dans la partie CDN. On verra plus loin que l'explication principale de ce résultat peut être attribuée à une diminution exceptionnelle de la fréquence des formes test pour cette année (une rubrique dédiée à la présentation détaillée des Caisses régionales de Crédit Agricole n'a pas été reprise cette année-là¹).

Le groupe des formes typiques des CDS présente une ventilation de ses spécificités qui varie nettement en fonction du contexte. En effet, on peut constater que d'une manière générale, ces formes sont sous-représentées dans les contextes CDN, ce qui est conforme aux résultats attendus pour le test. On note une légère différence pour l'année 2002, qui peut être attribuée au fait que la sous-représentation des formes typiques des CDS y est moins importante à cause la brièveté de la sous partie. En revanche, pour ce qui est des contextes de type CDS, il convient de distinguer plusieurs périodes. On peut en effet opposer les séries 1995-1997 et 1999-2000 avec les années 1998 puis 2001-2003, ces dernières étant caractérisées par une baisse plus ou moins marquée de la surreprésentation des

¹ On renvoie à la section du chapitre 2 où il est question du dispositif textuel des rapports d'activité.

formes censées être typiques des CDS. L'infléchissement que l'on note pour 1998 semble trouver un écho dans la banalité constatée pour 2001, la sous-représentation de 2002 et la légère augmentation de 2003. Ces résultats surprenants méritent que l'on pousse plus loin les investigations. Mais au préalable, on peut les confronter avec la ventilation des formes typiques des deux sortes de contextes sur l'ensemble du corpus.

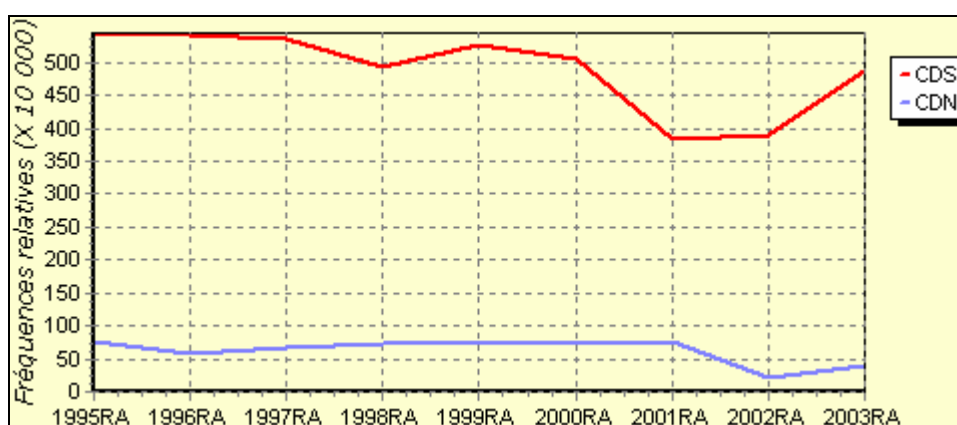


Figure 18.

Ventilation des deux groupes de formes typiques CDS et CDN sur l'ensemble du corpus partitionné en années

Lecture de la figure 18 : la courbe rouge représente la ventilation du groupe des dix formes les plus typiques des CDS et la courbe bleue, la ventilation du groupe des dix formes les plus typiques des CDN. La fréquence relative offre une lecture de la fréquence absolue de chaque forme dans chaque partie rapportée à la taille du corpus pris dans son ensemble. Sur l'axe des abscisses figurent les parties du corpus notées sous la forme « 1995RA ».

Ce graphique permet de mettre en lumière un point de l'interprétation fournie plus haut, au sujet de la moins bonne surreprésentation des formes du groupe CDN dans la partie CDN de l'année 2002 : on voit en effet que la fréquence des formes du test connaît un infléchissement net en 2002. Mais ce graphique vaut surtout parce qu'il éclaire les remarques formulées plus haut concernant la ventilation des spécificités des formes appartenant au groupe CDS. En effet, il permet de distinguer trois paliers que l'on peut rapprocher des périodes identifiées plus haut : 1995-1997, 1999-2000, puis 2001-2002. L'année 1998 correspond bien à un « creux » auquel répond l'infléchissement net marqué par la courbe de fréquence en 2001 et 2002,

l'année 2003 étant caractérisée par une nouvelle augmentation¹. Ces indications concordantes ayant été réunies, il convient alors de proposer une interprétation du phénomène.

3. Un contexte hybride pour un nouveau lectorat ?

Afin de d'expliquer la régression des formes typiques des contextes de type CDS, on peut émettre l'hypothèse selon laquelle ce phénomène enregistrerait une tendance de l'émetteur à utiliser différemment les deux dispositifs textuels identifiés dans le chapitre précédent. Un dispositif mixte émergerait dans le dernier tiers du corpus, consistant à introduire, plus fréquemment et en des proportions nouvelles, des graphiques, schémas et autres tableaux dans les contextes CDS. Afin d'illustrer ces deux aspects du phénomène, on donne dans la figure qui suit la carte des sections de la partie 2002, que l'on pourra comparer avec l'année 1995 (présentée plus haut, dans la section consacrée à la traduction textométrique des deux types de contextes).

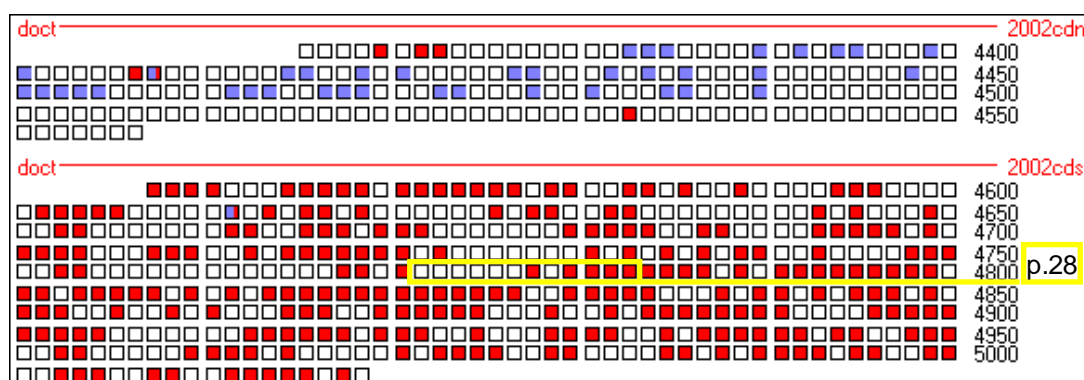


Figure 19.

Carte des sections (paragraphes) pour l'année 2002, partitionnée en fonction des contextes CDN et CDS et sur laquelle les formes typiques des deux contextes ont été projetées

¹ Compte tenu des formes retenues pour le test, mots grammaticaux et particules verbales, il semble que l'augmentation des fréquences relatives pour la partie RA2003 puisse être attribuée à une augmentation mécanique de la présence des mots grammaticaux du fait de la taille plus élevée de cette partie. En effet, alors que la taille moyenne des parties précédentes est d'un peu plus de 20 000 occurrences, la partie 2003 en compte plus de 39 000, soit près du double.

Lecture de la figure 19 : la carte des sections de Lexico permet de représenter à l'aide de carrés un type de section particulier présent dans le corpus. Dans cette carte, chaque carré représente un paragraphe. Lorsqu'un paragraphe est coloré en rouge, c'est qu'il contient l'une des formes typiques des contextes de type syntactique. Un paragraphe est coloré en bleu lorsqu'il contient au moins l'une des formes typiques des contextes non syntactiques. Les paragraphes apparaissant en blanc ne contiennent aucune des formes les plus caractéristiques des deux sortes de contextes. Les mentions « 2002cdn » et « 2002cds » indiquent que les rubriques de la partie 2002RA du corpus ont été regroupées selon qu'elles relèvent plutôt d'un CDN ou d'un CDS. La zone encadrée en jaune correspond à la page 28 du rapport d'activité 2002 ; elle est commentée plus bas.

Si l'on compare les deux années 1995 et 2002, les points suivants méritent d'être relevés : pour la partie CDN, la densité des carrés bleus semble devenue plus lâche (voir section précédente), et des carrés rouges y sont présents en plus grand nombre. Quant à la partie CDS, elle présente une érosion considérable de la présence des formes typiques des CDS.

La présence de certaines formes typiques des CDS en contextes CDN a deux raisons : tout d'abord, la multiplication des notes et autres mentions explicatives (quatre cas sur six), et la présence de l'unité de mesure « milliards », qui était la seule exception recensée pour l'année 1995. Il s'agit là de phénomènes que l'on peut considérer comme marginaux, surtout si on les compare avec l'érosion caractérisant les contextes de type CDS. On commencera par illustrer ce point à l'aide d'une comparaison entre une page tirée du texte dans son format original et sa transposition dans la carte des sections (il s'agit de la page 28 délimitée dans la figure 20, présentée ci-dessous).

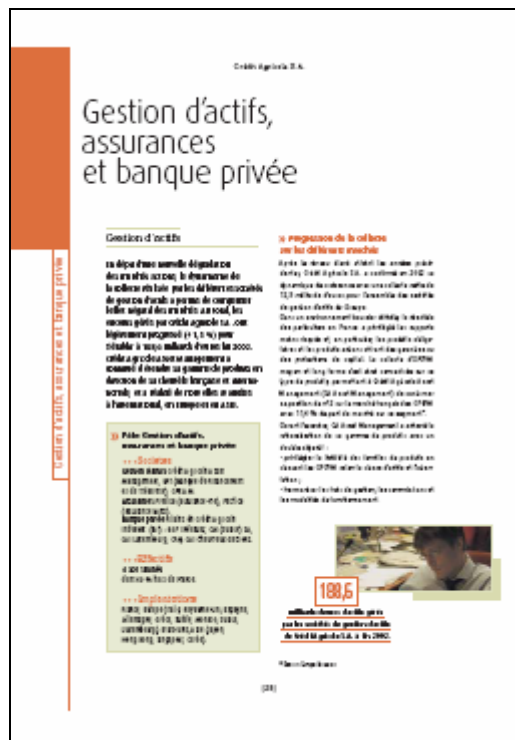


Figure 20.
Page 28 du rapport d'activité de l'année 2002, correspondant à la zone délimitée dans la figure 19

La reproduction de cette page du rapport d'activité dans son format original permet de mettre en évidence un découpage de l'espace résultant d'une mise en page complexe. On dénombre en effet sur cette seule page, un titre principal, isolé par une police de grande taille et reprenant le titrage de la page précédente, un titre de deuxième niveau permettant d'isoler, à l'intérieur de « Gestion d'actifs, assurances et banque privée », le sous-chapitre dédié à la gestion d'actifs, puis, sous ce dernier, un titre de troisième niveau destiné à annoncer une section plus détaillée. Par ailleurs, on distingue trois blocs de texte : l'un d'entre eux sert d'introduction au sous-chapitre, l'autre correspond au début de la section consacrée à l'un des aspects de la gestion d'actifs, enfin, un encart isole un texte qui vient dans le prolongement de l'introduction. Une photographie et un chiffre mis en exergue viennent parachever la construction de cette page. Ce sont donc six zones différentes dont il convient maintenant de commenter la transposition.

L'encart de texte se décompose en trois parties intitulées « Sociétés », « Effectifs » et « Implantations », la première d'entre elles étant à son tour découpée

en « Gestion d'actifs », « Assurances » et « Banque privée ». Cet encart est donc à rattacher au titre principal et non au sous-chapitre « Gestion d'actifs ». Il est décomposé en cinq paragraphes, selon la règle adoptée pour l'ensemble du corpus. L'introduction et la première section ont été comptées comme deux rubriques distinctes, la première se composant d'un unique paragraphe et la seconde de trois. Le chiffre mis en exergue a été considéré comme un paragraphe à part entière. La zone isolée sur la carte des sections comporte douze paragraphes qui, à la suite des commentaires qui viennent d'être donnés, se répartissent en dix paragraphes regroupant l'information de la page et deux paragraphes servant à délimiter les deux rubriques (comme cela a été indiqué dans le chapitre précédent, les photographies n'ont pas fait l'objet d'un balisage).

Si l'on considère maintenant les paragraphes qui ont été sensibles au test, on voit qu'ils correspondent aux textes de l'introduction et de la première section. On donne à titre d'illustration le texte de l'introduction correspondant au premier carré rouge en partant de la gauche à l'intérieur de la zone isolée dans la figure 19. Les mots en rouge sont ceux qui appartiennent à la liste des formes typiques des CDS :

En dépit d' **une** nouvelle dégradation des marchés actions , le dynamisme de la collecte réalisée par les différentes sociétés de gestion d' actifs **a** permis de compenser l' effet négatif des marchés . Au total , les encours gérés par Crédit Agricole S . A . **ont** légèrement progressé (+ 1 , 5 %) pour s' établir à 188 , 6 **milliards** d' euros fin 2002 . Crédit Agricole Asset Management **a** continué d' étendre **sa** gamme de produits en direction de **sa** clientèle française et internationale , et **a** réalisé de nouvelles avancées à l' international , en Europe et en Asie . #

Figure 21.

Texte correspondant au premier paragraphe coloré en rouge de la zone isolée dans la figure 19

Si, en revanche, on observe le contenu des paragraphes non identifiés par le test, alors qu'ils sont porteurs d'information, on voit qu'ils correspondent à une série de données livrées dans un contexte phrastique très appauvri, dans lequel dominent l'ellipse et la juxtaposition. On donne ci-dessous un exemple d'un tel contexte :

Banque privée Filiales de Crédit Agricole Indosuez (CAI) : BGP Indosuez , CAI (Suisse) SA , CAI Luxembourg , CFM , CAI Cheuvreux Gestions . #

Figure 22.

Texte correspondant au quatrième paragraphe non coloré de la zone isolée dans la figure 19

L'ensemble des indications réunies dans cette section semble converger vers la caractérisation d'un nouveau type de contexte. Ce dernier est constitué de la réunion de trois éléments : le recours à une mise en page complexe, un découpage du texte très poussé à l'aide de rubriques de faible dimension et une multiplication des paliers intermédiaires et enfin, l'insertion dans les CDS de paragraphes, voire de rubriques, destinés à fournir les informations sous une forme condensée, si ce n'est télégraphique. On constate donc que la période qui débute en 2001 correspond sur le plan du dispositif textuel à une évolution notable. Tout ce passe comme si ce changement était destiné à fournir plusieurs niveaux de lecture pour les parties textuelles. La place accordée au texte principal est rognée au profit d'informations condensées (introduction, encart) ou proposées sous la forme de « flashes », comme la mise en exergue d'un « chiffre clé ». Sur la même page sont donc proposés trois niveaux de lecture, une attention particulière étant accordée au plus superficiel des trois, le feuilletage.

À défaut de disposer de données précises permettant de caractériser la position du destinataire dans la situation de communication, on peut néanmoins formuler l'hypothèse selon laquelle l'apparition de ce nouveau type de contexte renseigne sur les dispositions du destinataire à son égard. Celui-ci paraît, en effet, vouloir s'adapter ainsi à un nouveau public, ou tout au moins, à l'image qu'il s'en fait. On ne peut alors manquer de rapprocher cette évolution de l'ouverture du capital du Crédit Agricole à des millions d'actionnaires individuels, à la suite de l'introduction en bourse qui a eu lieu en 2001. Capter également l'attention du grand public passerait ainsi par un appauvrissement du texte au profit des données brutes, susceptibles d'être interprétées par l'*homo-oeconomicus* dont il était question dans le chapitre précédent. Ce traitement se rapproche, et probablement pour les mêmes raisons, des manières de diffuser l'information que l'on rencontre sur Internet. Cela confirme au moins le caractère représentatif des rapports d'activité, dès lors que l'on s'intéresse aux problématiques nouvelles de diffusion de l'information apparues avec ce médium, et plus particulièrement lorsqu'il a pénétré dans les organisations sous la forme d'un intranet.

Pour ce qui est de l'exploitation directe du corpus, la mise en place de ce test a permis de rectifier une dizaine de balisages de rubriques erronés, soit un pour cent

du total. Cela peut paraître négligeable. Cependant, toutes les erreurs provenant des dernières parties du corpus, celles où émerge précisément le nouveau type de contexte, le test a démontré son utilité. Il a, en effet, permis d'identifier la principale cause d'erreur liée à l'apparition d'un nouveau type de contexte ; le corpus étant ouvert, on dispose ainsi d'un moyen de contrôle efficace. On peut, par ailleurs, se demander si la division opérée entre deux types principaux, CDS et CDN, n'a pas à être remise en cause, afin de laisser la place à ce nouveau contexte. Il semble encore trop tôt pour procéder à cette prise en compte, d'autant que la tendance paraît s'infléchir dans la dernière partie du corpus¹.

De façon plus générale, la mise en place de ce test permet de montrer que la prise en compte des contextes non syntactiques est non seulement envisageable, mais aussi souhaitable dans une perspective d'extraction de l'information. Située en amont des études portant sur le profilage de textes ou sur l'identification de genres qui reposent en général sur des contextes syntactiques, la distinction des deux sortes de contextes semble constituer un préalable utile. À condition de rester dans le cadre d'une approche textométrique traitant des données nues (non étiquetées), on dispose, en effet, du moyen de contrôler un type de contexte particulièrement résistant aux traitements automatiques de type étiquetage et analyse morpho-syntaxique. Le test textométrique, en fournissant à l'interprétation des données quantifiées et reproductibles, permet de suivre l'évolution de contextes aux caractéristiques très contrastées. Modifications intervenant dans la répartition entre contextes, perturbation de l'homogénéité d'une sorte de contexte, émergence d'un nouveau type de contexte : tous ces phénomènes peuvent être identifiés de manière fiable et fournir des pistes d'investigation nouvelles, comme les répercussions des modifications de la situation de communication sur les dispositifs textuels d'une même série de documents. Surtout, la prise en compte des contextes de type non

¹ L'étude du vocabulaire spécifique des paragraphes réagissant négativement au test CDS pour les parties 2001 à 2003 montre 1°/ une majorité de phrases nominales, 2°/ une répartition thématique différente pour 2001-2002 et 2003 : pour 2001-2002 le vocabulaire de ces contextes est dominé par les noms d'entités, les unités relatives à l'organisation du nouveau Groupe et le classement de ce dernier parmi les autres établissements bancaires ; pour 2003 domine un vocabulaire relatif aux ressources humaines, 3°/ ce nouveau contexte n'est stable ni du point de vue de son vocabulaire ni du point de vue de son dosage dans les rapports d'activité. Il peut donc provisoirement être tenu pour relativement instable et non fixé.

syntactiques offre la possibilité de pousser l'extraction d'information sur un terrain peu exploré alors qu'il s'avère riche en entités nommées. On dispose alors d'un cadre pour l'observation du fonctionnement de ces dénominations qui s'apparente à celui que l'on rencontre sur les pages de l'Internet, mais aussi sur celles des intranets dans les organisations. Il devient alors possible d'étudier l'influence de ces sortes de contextes sur les dénominations elles-mêmes, mais aussi sur les signalétiques particulières dans lesquelles elles sont insérées.

II SÉRIE TEXTUELLE CHRONOLOGIQUE ET ÉVOLUTION DU VOCABULAIRE : FAIRE DU NEUF AVEC DU VIEUX ?

A. Le renouvellement du vocabulaire des rapports d'activité

1. Le renouvellement théorique du vocabulaire

La précédente section a permis de montrer que l'organisation textuelle des textes du corpus est relativement perméable aux changements introduits dans la situation de communication. En effet, c'est de cette manière que l'on a proposé d'interpréter l'apparition concomitante de contextes hybrides, offrant des possibilités de lecture simplifiée, et d'un public élargi pour les rapports d'activité, devenus instruments de communication d'une société cotée en bourse. Il paraît maintenant légitime de se demander si un tel changement a également eu des répercussions sur le vocabulaire des rapports d'activité. Il s'agit à ce stade de raisonner sur des grandes masses et non sur des vocables particuliers, ce point étant traité dans les chapitres suivants. La vision globale que l'on cherche à obtenir sur les principales variations diachroniques affectant le vocabulaire du corpus est utile pour éclairer le travail de collecte qui sera à réaliser plus loin. Cette « profondeur de champ » offerte par les séries textuelles chronologiques est l'un des bénéfices indéniables que l'on peut attendre de leur utilisation.

En effet, le corpus utilisé dans cette étude rassemble une série de rapports d'activité d'une même entreprise dont les publications se sont succédées entre 1996 à 2004, ce qui correspond aux rapports d'activité des années 1995 à 2003. Le corpus utilisé est une série textuelle chronologique : « *Nous appelons séries textuelles chronologiques ces corpus homogènes constitués par des textes produits dans des*

situations d'énonciation similaires, si possible par un même locuteur, individuel ou collectif, et présentant des caractéristiques lexicométriques comparables¹. »

Lorsque l'on dispose d'une telle série chronologique, il devient possible d'observer la manière dont le vocabulaire étudié varie dans le temps, à conditions d'énonciations similaires. Le même type de textes émis régulièrement par la même organisation à un an d'intervalle possède-t-il toujours le même vocabulaire ou, au contraire, ce dernier se renouvelle-t-il, dans quelles proportions, et en particulier, à l'aide de quelles unités ? Ces questions qui valent pour les nouveautés s'appliquent également aux disparitions : certaines unités disparaissent-elles, à quel rythme, et pour quelles raisons ? Enfin, existe-t-il un vocabulaire stable, ou soumis à de très faibles variations que l'on retrouve d'une année sur l'autre ? Pour aussi évidentes qu'elles soient, ces questions méritent toujours d'être posées. Car si un tel corpus offre un cadre qui relève de la synchronie à l'échelle de la langue, il permet, entre autres dans la perspective de l'étude d'un référentiel, de montrer des évolutions significatives du vocabulaire à l'intérieur d'intervalles temporels très rapprochés.

L'analyse factorielle des correspondances est capable de produire une comparaison globale des différents stocks de vocabulaire correspondant aux différentes parties dont le corpus est composé. Elle permet d'indiquer dans quelle mesure les stocks ainsi comparés se renouvellent d'une année sur l'autre². Si pour les cinq parties A, B, C, D et E d'une même série chronologique, on observe, d'une part, qu'une partie du vocabulaire de A ne se retrouve pas dans B, qu'une autre partie se retrouve effectivement dans B, et que B apporte son lot de nouveautés qui ne figurent pas dans A, et que d'autre part, le phénomène se reproduit dans des proportions similaires d'une année sur l'autre jusqu'à E, l'AFC produit une parabole caractéristique, appelée également « effet Guttman » :

¹ L. Lebart et A. Salem (1994 : 217).

² A. Salem (1988).

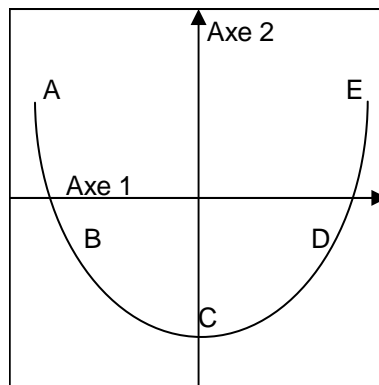


Figure 23.

Répartition caractéristique d'une série chronologique homogène produite par l'AFC¹

Lecture de la figure 23 : la figure 23 représente la répartition théorique d'une série textuelle chronologique A, B, C, D, E dans laquelle le vocabulaire se renouvelle de façon régulière d'une partie sur l'autre. Les axes 1 et 2 représentent les deux premiers facteurs de l'analyse factorielle des correspondances, c'est-à-dire ceux qui résument les deux quantités d'informations les plus importantes parmi les autres facteurs qui ont été calculés. Les facteurs sont orientés et se croisent en leur point d'inertie.

L'analyse factorielle des correspondances met ainsi en évidence, à travers cette distribution particulière des parties d'un corpus, ce qu'A. Salem a proposé de nommer le temps lexical : « (...) les vocabulaires des parties correspondant à des périodes consécutives dans le temps présentent en général plus de similitudes entre eux que les vocabulaires correspondant à des périodes séparées par un intervalle de temps plus long. C'est ce phénomène que nous appelons « temps lexical »². » Si un renouvellement régulier du vocabulaire d'une année sur l'autre est susceptible de produire la courbe théorique présentée dans le schéma ci-dessus, cette dernière permet également de mettre en évidence des irrégularités qu'il aurait difficile d'identifier autrement.

¹ Pour une présentation détaillée, on renvoie à A. Salem (1988) et (1993).

² A. Salem (1988 : 107).

2. Une rupture introduite par l'année 2001

La figure donnée ci-dessous met en évidence une courbe assez proche du modèle si l'on fait porter l'analyse seulement sur les six premières parties, soit les années 1995 à 2000.

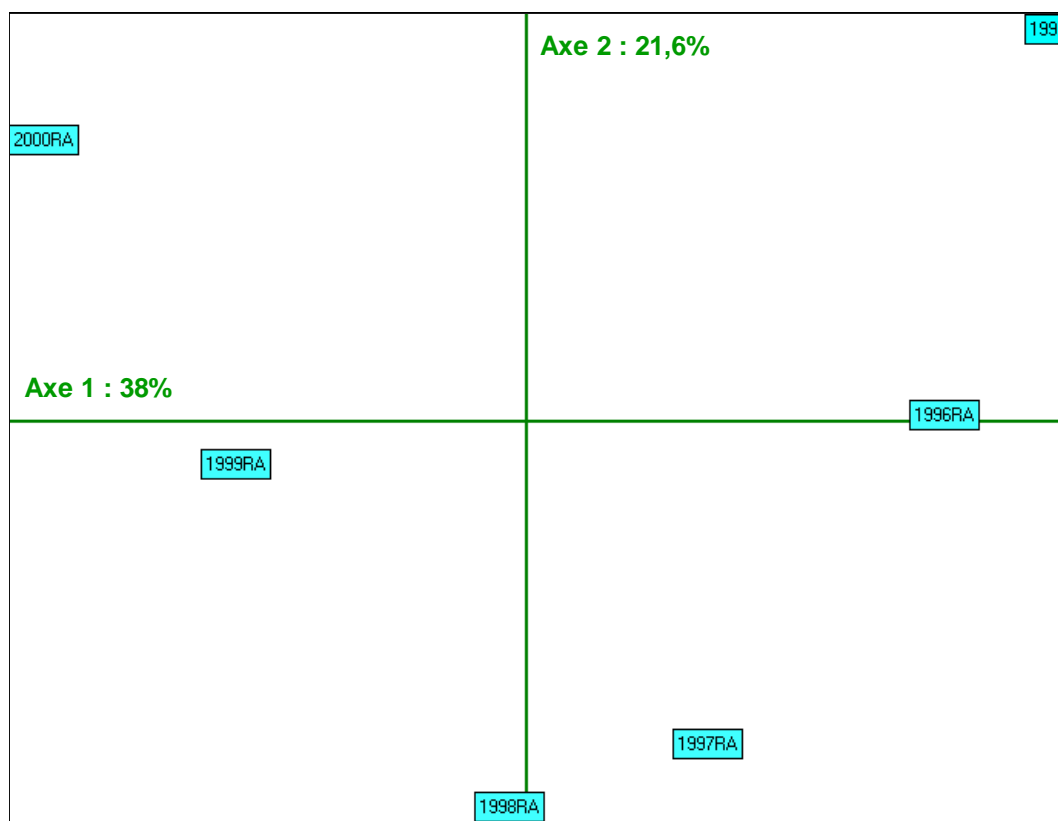


Figure 24.

AFC sur les parties 1995RA à 2000RA du corpus « rapports d'activité »

Lecture de la figure 24 : la figure est composée d'étiquettes correspondant aux différentes parties du corpus sur lesquelles l'analyse factorielle des correspondances a porté. La distribution des parties de gauche à droite ou de droite à gauche n'a pas de signification particulière, de même que l'aspect concave ou convexe de la courbe : il s'agit d'effets graphiques produits pour l'affichage. Seule importe ici l'allure générale de la courbe. Les caractéristiques de ces parties sont données dans le tableau 10 du précédent chapitre. L'analyse a pris en compte les formes dont la fréquence est supérieure à 2, soit, pour les six premières parties, 63,6% des formes et 97,5% des occurrences. Les étiquettes se distribuent sur le plan par rapport à deux axes correspondant chacun aux deux premiers facteurs, c'est-à-dire à ceux qui représentent, par ordre décroissant, 38% de l'information pour l'axe 1, et 21,6% de l'information pour l'axe 2. La valeur associée à un axe correspond à la quantité d'informations du TLE que le facteur résume.

On observe une répartition assez régulière des six premières parties dont se compose le corpus « rapports d'activité ». On remarque, en effet, que leur succession sur le plan dessine une parabole inversée assez proche de la répartition théorique fournie par l'AFC lorsque le renouvellement du vocabulaire se produit régulièrement d'une partie à l'autre. Le corpus étant un corpus de suivi ayant vocation à s'accroître d'un nouveau stock de vocabulaire chaque année, l'AFC permet de mettre en évidence une modification sensible de la courbe à partir de 2001 :

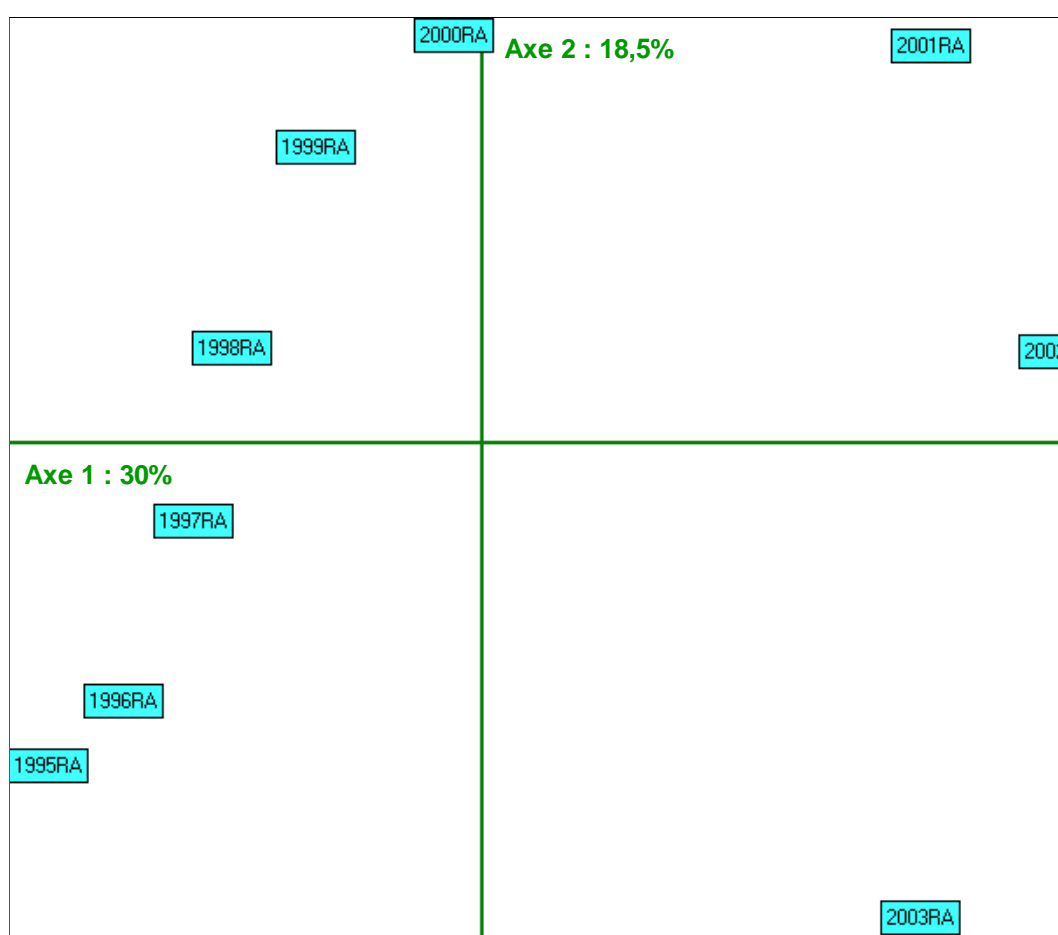
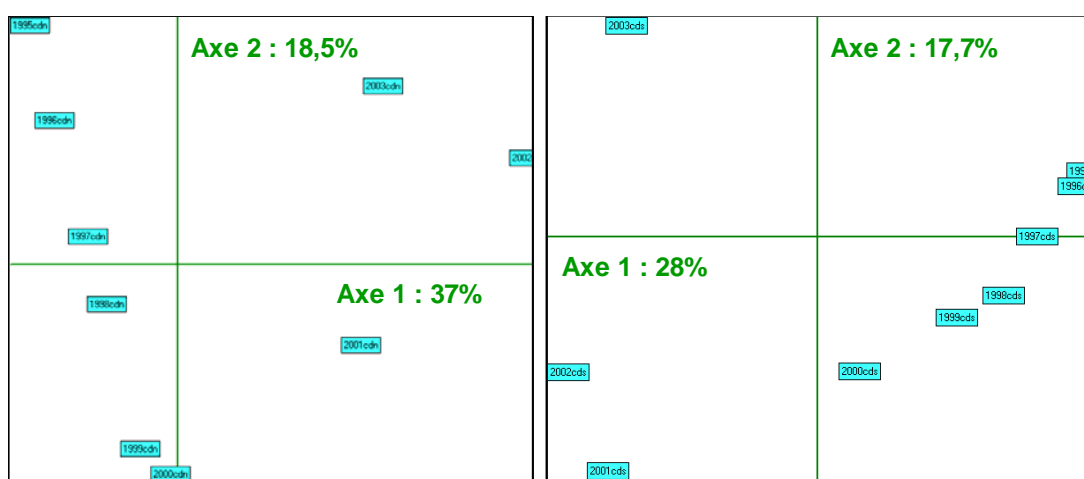


Figure 25.
AFC sur les parties 1995RA à 2003RA du corpus « rapports d'activité »

Lecture de la figure 25 : la figure 25 est composée d'étiquettes correspondant aux différentes parties du corpus sur lesquelles l'analyse a porté. Les caractéristiques de ces parties sont données dans le tableau 10. L'analyse a pris en compte les formes dont la fréquence est supérieure à 2, soit, pour les neuf parties du corpus, 63% des formes et 98% des occurrences. Les étiquettes se distribuent sur le plan

par rapport à deux axes correspondant chacun à un facteur. La valeur associée à un axe correspond à la quantité d'informations du TLE que le facteur résume. Les facteurs retenus sont les deux premiers, c'est-à-dire ceux qui possèdent la valeur propre la plus élevée, et qui représentent respectivement 30% (axe 1) et 18,5% (axe 2) de l'information prise en compte.

La présence des années 2001 à 2003 a eu pour effet de repousser les parties précédentes à une extrémité de l'axe principal. Cela indique l'apparition de différences importantes entre les éléments de la série – qui jusqu'alors s'organisaient de manière relativement homogène –, et les nouvelles parties. On note, par ailleurs, que l'année 2002 a accentué encore cet écart entre les vocabulaires, alors que l'année 2003, qui clôt le corpus, semble amorcer une stabilisation, puisqu'elle s'aligne à peu près sur la position de 2001. Dans la précédente section, il a été question d'une rupture intervenant également au moment de la cotation de Crédit Agricole S.A. Il s'agissait alors d'une rupture au niveau de l'organisation textuelle, les contextes à dominante non syntactique tendant à gagner les contextes à dominante syntactique. Ce qu'indique l'AFC dans les deux figures ci-dessus correspond à un phénomène d'une autre nature. En effet, ce ne sont plus les types de contextes qui sont pris en compte, mais les stocks de vocabulaire correspondant à chaque partie du corpus. On peut donc se demander si, en reprenant cette autre division du corpus, on obtient un profil de série chronologique similaire.



Figures 26 et 27.

Répartition des parties 1995cdn à 2003cdn – Figure 26 – et 1995cds à 2003cds – Figure 27 – du corpus « rapport d'activité », produite par une AFC

Lecture des figures 26 et 27 : Chaque figure est composée d'étiquettes correspondant aux différentes parties du corpus sur lesquelles l'analyse a porté. L'analyse a pris en compte les formes dont la fréquence est supérieure à 2, soit, pour les CDN des neuf parties, 66% des formes et 96,5% des occurrences, et pour les CDS des neuf parties, 60% des formes et 97,8% des occurrences. Les étiquettes se distribuent sur le plan par rapport à deux axes correspondant chacun à un facteur. La valeur associée à un axe correspond à la quantité d'informations du TLE que le facteur résume. Les facteurs retenus sont ceux qui possèdent la valeur propre la plus élevée.

Lorsque l'on compare les types de contextes CDN et CDS des différentes parties, l'AFC permet d'établir que la rupture constatée pour les parties prises globalement se répète pour chaque type de contexte et dans des proportions similaires : 2001 « repousse » à l'autre extrémité de l'axe principal les autres parties, 2002 confirme ce mouvement alors que 2003 indique un recul de cette tendance. Comme dans la figure précédente, l'écart de 2003, moins prononcé sur l'axe principal est en revanche extrême sur l'axe 2. En comparant les trois figures, toutes les parties du corpus, puis tous les contextes CDN et CDS, on constate que ce sont surtout ces derniers qui semblent responsables de l'écart établi entre cette dernière partie et les deux années 2001 et 2002.

Sans entrer ici dans une analyse détaillée du vocabulaire – celle-ci constitue la matière des chapitres suivants –, on peut néanmoins chercher à cerner de plus près cette évolution. Tout d'abord, on peut se demander dans quelle mesure l'apport de formes nouvelles est responsable de cette rupture, pour ensuite tenter d'identifier les unités qui sont particulièrement responsables de ces écarts, et voir dans quelle mesure leurs caractéristiques ancrent dans le vocabulaire un certain nombre de faits déjà notés plus haut : cotation de l'entreprise en 2001, acquisition du Crédit lyonnais en 2003, et durant cette période, modification du dispositif textuel des rapports d'activité. Pour cela, les indications fournies l'AFC peuvent, d'une part, être comparées à la courbe d'accroissement du vocabulaire du corpus, et d'autre part, être éclairées à l'aide d'une application particulière du calcul des spécificités.

B. Changement de destinataire et vocabulaire du corpus

1. Vérification de l'accroissement du vocabulaire du corpus

Alors que les observations réalisées dans la première section, à propos de l'évolution du dispositif textuel, paraissent interprétables en termes d'ajustement des rapports d'activité à un nouveau lectorat – ce point reste malgré tout soumis à des vérifications complémentaires –, on fera l'hypothèse dans cette section que les irrégularités constatées dans le renouvellement du vocabulaire du corpus tiennent pour une part non négligeable au changement de statut de l'émetteur.

Pour cela, on commencera par vérifier que ces perturbations ne sont pas imputables à un changement dans rythme du seul accroissement du vocabulaire. On donne ci-dessous la courbe de cet accroissement pour le corpus pris dans son ensemble.

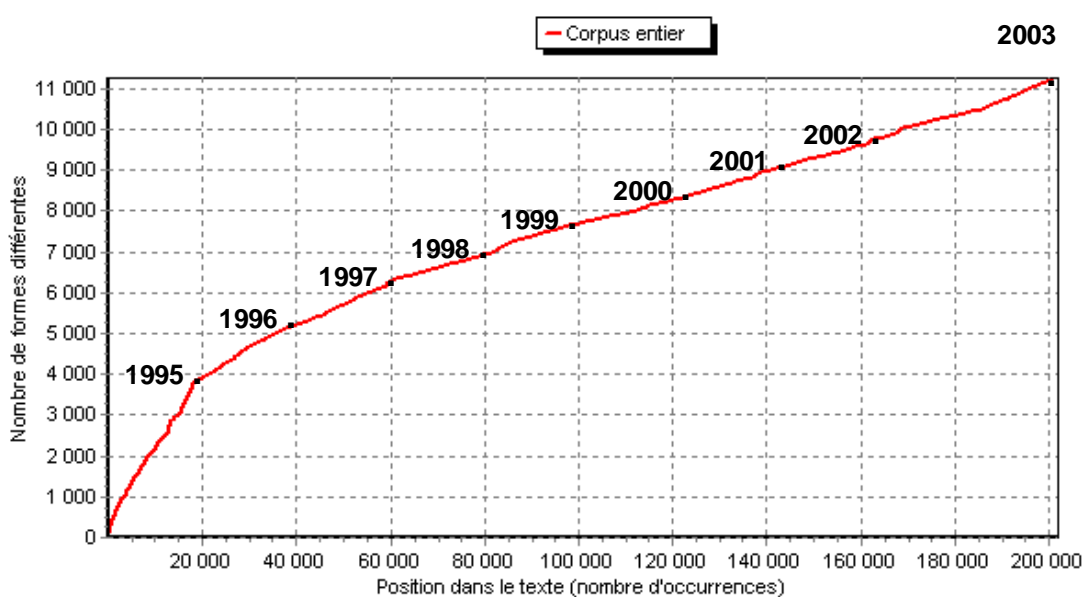


Figure 28.

Accroissement du vocabulaire pour l'ensemble des parties dont se compose le corpus « rapports d'activité »

Lecture de la figure 28 : Sur l'axe des ordonnées l'accroissement du nombre de formes graphiques différentes est indiqué en milliers de formes. Sur l'axe des abscisses, l'accroissement du nombre d'occurrences est indiqué par séquences de vingt mille. L'apport du vocabulaire de chaque partie est indiqué par la mention de l'année correspondante. Le graphique se lit de cette manière : la première partie 1995 comporte 3818 formes réparties en 18274 occurrences (voir les principales caractéristiques textométriques du corpus données à la fin du

précédent chapitre) ; l'année 1996 ajoute à ce stock initial plus d'un millier de formes nouvelles réparties sur 19306 occurrences, ce qui situe alors le stock d'occurrences du vocabulaire à 18274 + 19306 occurrences, soit 37580 occurrences, et ainsi de suite pour les années suivantes.

Ce graphique livre une image synthétique du volume de l'accroissement du vocabulaire. Le stock de formes de chaque partie ne se cumule pas intégralement avec celui des parties précédentes. Une portion de ce vocabulaire disparaît, une autre est commune entre la nouvelle partie et les précédentes, enfin, une dernière portion de ce vocabulaire est nouvelle et vient par conséquent enrichir le stock des formes du corpus. Ces formes nouvelles enrichissent le fonds commun du corpus dans des proportions oscillant entre environ 750 et 1500 formes d'une partie à l'autre, soit approximativement un quart des formes graphiques de chaque partie. Pour d'autres corpus on a pu observer que la courbe d'accroissement du vocabulaire a tendance s'infléchir au fur et à mesure que le nombre d'occurrences prises en compte augmente¹. Il n'en va pas de même pour la courbe du vocabulaire du corpus « rapports d'activité », qui est linéaire pour la période centrale 1997-2002 et a tendance à augmenter légèrement pour la dernière partie. On déduit de ce résultat que la rupture entre les périodes 1995-2000 et 2001-2003 mise en évidence par l'analyse factorielle des correspondances ne peut pas s'expliquer par un accroissement brutal du nombre de formes nouvelles. On fait alors l'hypothèse que la rupture qui apparaît à partir de 2001 reposerait principalement sur des variations de fréquence importantes affectant des formes qui étaient déjà présentes dans le vocabulaire du corpus avant 2001. Afin de vérifier ce point, on aura recours au calcul des accroissements spécifiques.

2. Hypothèse d'une nouvelle configuration du vocabulaire : accroissements spécifiques des parties 2001 et 2003

a. Reconfiguration du groupe Crédit agricole

Il s'agit moins ici de repérer le vocabulaire spécifique d'une partie par rapport à l'ensemble du corpus, que d'identifier les unités qui la distinguent des

¹ L. Lebart et A. Salem (1994 : 55-56 et 77).

parties précédentes seulement. On étudiera particulièrement les années 2001 et 2003, la première par rapport aux années 1995 à 2000, et la seconde, par rapport à toutes les autres parties du corpus, 2003 clôturant provisoirement ce dernier.

Le calcul des spécificités¹ utilisé dans la première section de ce chapitre permet de porter un diagnostic sur la présence d'une forme dans une partie du corpus. Ce calcul prend en compte la taille du corpus, la fréquence de la forme dans le corpus, la fréquence de la forme dans la partie et la taille de la partie, et il compare ces données avec celles que l'on observe pour les autres formes du corpus dans des échantillons de même taille que la partie retenue. Il permet alors de dire en probabilité si la sous-fréquence de la forme étudiée est relativement faible dans cette partie (la forme est spécifique négative), relativement élevée (la forme est spécifique positive), ou banale. Pour ce type de spécificités, on voit que la norme retenue est celle du corpus. Il n'en va pas de même pour les spécificités évolutives qui ne reprennent pas tout le corpus comme norme, mais seulement le vocabulaire des parties qui précèdent chronologiquement la partie étudiée. C'est ainsi que, pour le calcul des spécificités évolutives de la partie 2001, le corpus sera réduit à la période 1995-2001, et le stock de vocabulaire de la partie 2001 sera comparé à celui des parties 1995 à 2000. Le but est alors d'isoler les formes qui possèdent une fréquence anormalement élevée ou basse dans la partie 2001 par rapport aux parties précédentes. Les cas limites d'une telle configuration correspondent aux formes présentes dans les parties précédant 2001 et absentes de cette dernière, ainsi que les formes originales apparues seulement en 2001.

Afin de prendre la mesure des caractéristiques du vocabulaire de l'année 2001, on examine les formes affectées des plus forts coefficients de spécificité positifs et négatifs, ces derniers signalant des sur ou des sous-emplois significatifs. Les accroissements seront observés à l'échelle des parties principales (rapports d'activité).

¹ *Ibid.* (p. 172 *sq.*).

*Tableaux 19 et 20.
Principaux accroissements spécifiques positifs et négatifs de la partie RA2001*

Forme	Frq. Tot.	Fréq.	Coeff. +	Forme	Frq. Tot.	Fréq.	Coeff. -
S	166	150	***	agricole	1097	5	***
Agricole	1139	408	***	Total	670	3	-42
2001	207	181	***	régionales	394	1	-27
A	215	161	***	locales	453	4	-26
Régionales	96	96	***	bilan	375	2	-24
2002	43	37	25	1996	220	1	-15
CA	40	34	23	propres	412	16	-13
Asset	111	56	19	a	1734	162	-12
Administrateur	31	27	19	1997	171	1	-11
Management	120	58	18	fonds	462	26	-10

Lecture des tableaux 19 et 20 : les deux tableaux présentent la même structure, dans laquelle la colonne « Forme » est suivie des colonnes Frq. Tot. (pour fréquence totale pour les parties 1995 à 2001), Fréq. (pour la seule partie 2001) et Coeff. +/-, qui indique le coefficient de spécificité positif ou négatif constaté pour les formes correspondantes. Les *** indiquent un coefficient supérieur à 50, ce qui est une valeur très élevée. Dans le tableau 19, sont rassemblées les dix premières formes dotées du coefficient de spécificité positif le plus élevé pour la partie 2001 ; dans le tableau 20, ce sont les dix dernières formes ayant le coefficient de spécificité négatif le plus élevé pour la partie 2001.

Afin de préparer le commentaire de ces tableaux, on présentera tout d'abord les dénominations et les syntagmes lexicalisés dans lesquels entrent les formes isolées figurant dans les tableaux. En effet, ces formes simples, résultat du travail de segmentation présenté plus haut, appartiennent généralement à des unités complexes qui les englobent, ou polyformes. Pour une plus grande clarté du commentaire, on rétablit donc ces unités complètes artificiellement démembrées pour le calcul¹.

Les formes ayant un coefficient de spécificité positif élevé entrent majoritairement dans les syntagmes suivants :

- « Crédit Agricole S.A. », incorpore les formes « S », « Agricole » et « A »
- « Caisses Régionales », reprend la forme « Régionales »

¹ On verra plus loin que l'utilisation des segments répétés permet de prendre en compte ces unités complexes pour le calcul des spécificités.

- « CA Asset Management », contient les formes « CA », « Asset » et « Management »

Les formes « 2001 » et « 2002 » correspondent, pour la première, à l'année dont il est principalement question dans le rapport d'activité 2001, et pour la seconde, à l'année suivante, entamée au moment de la diffusion du rapport d'activité en mai 2002. Quant à la forme « Administrateur », elle correspond au mot « administrateur », titre conféré aux membres des conseils d'administration de sociétés.

Les formes dotées d'un coefficient de spécificité négatif élevé pour l'année 2001 entrent, pour leur part, dans les syntagmes suivants :

- « Crédit agricole », pour la forme « agricole »
- « Total bilan », pour les formes « Total » et « bilan », expression utilisée pour annoncer le montant du bilan d'une entité.
- « Caisses régionales », pour la forme « régionales »
- « Caisses locales », pour la forme « locales »
- « fonds propres », pour les formes « fonds » et « propres », autre indicateur donné à l'occasion de la présentation d'une banque (comme « total bilan » mentionné plus haut)
- la particule verbale « a » correspond à la forme « a »
- 1996 et 1997, correspondent aux années d'activité dont il n'est pratiquement plus question en 2001.

Afin de commenter la présence de ces formes caractéristiques de l'année 2001 il est commode de les regrouper en fonction de certaines oppositions. On notera au préalable que la plupart des formes concernent soit des dénominations d'entités essentielles du groupe Crédit Agricole, soit des syntagmes relatifs à l'activité de ces entités. On peut donc déjà en conclure que les bouleversements ayant touché l'organisation du groupe bancaire en 2001 ont des répercussions sensibles dans le vocabulaire de la partie du corpus correspondante. Il reste que les formes prises par ces répercussions peuvent être assez différentes. Ainsi, trois regroupements se dégagent, qui concernent chacun un aspect particulier du fonctionnement des formes dans le corpus.

Certaines formes appartenant aux deux tableaux s'opposent en fonction de la graphie qui a été utilisée dans les différentes parties du corpus. C'est ainsi que l'unité « Caisses Régionales », forme originale de la partie 2001, s'oppose à l'unité « Caisses régionales » utilisée jusqu'alors. Il en va de même pour les deux graphies « Crédit agricole » et « Crédit Agricole ». Dans la partie consacrée à l'entité « Crédit Agricole » on reviendra sur la portée qu'il convient d'attribuer à l'apparition de la majuscule dans les graphies de « régionales » et de « agricole ». On se borne ici à enregistrer ces faits mis en relief par la ventilation des variantes graphiques :

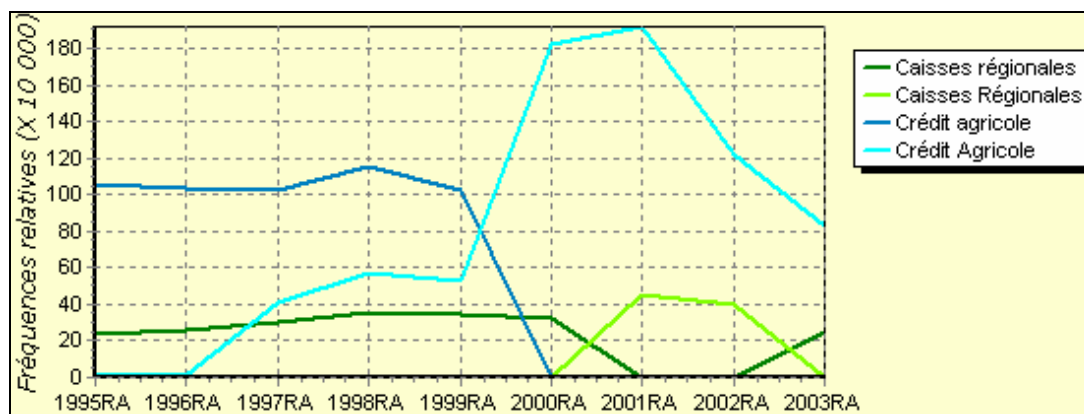


Figure 29.

Ventilation en fréquences relatives des variantes graphiques du syntagme lexicalisé « caisses régionales » et de la dénomination « Crédit agricole »

Cette figure précise la répartition des variantes sur l'ensemble des parties du corpus. Pour le syntagme lexicalisé « Caisses régionales », la période 1995-2000 s'oppose nettement aux deux années 2001 et 2002, 2003 marquant un retour à l'ancienne graphie. Quant à la majuscule attribuée à « agricole » dans « Crédit Agricole », elle donne lieu à une répartition proche mais non identique. En effet, il semble bien qu'il y ait un lien entre l'apparition de la nouvelle graphie et la disparition de l'ancienne mais, d'une part, cette substitution est intervenue dans le rapport de l'année 2000, et d'autre part, la forme avec majuscule existait depuis plusieurs années et perdure jusqu'en 2003. La baisse de fréquence observée pour 2003 enregistre, entre autres, une dilution de la forme dans une partie deux fois plus importante que les autres. La fréquence absolue de « Crédit Agricole » en 2003 est, en effet, voisine de celle qui est observée en 2001.

D'autres formes appartenant à ces tableaux peuvent être rassemblées lorsque leurs forts coefficients de spécificité ont pour origine une modification des dispositifs textuels. En effet, la présence de la forme « Administrateur » attire l'attention sur l'apparition d'une nouvelle rubrique consacrée à la présentation des titres et fonctions des membres du conseil d'administration de Crédit Agricole S.A. D'autre part, la raréfaction de certaines formes signale l'allègement du dispositif non syntactique destiné à présenter les caisses régionales de Crédit agricole. Comme cela est visible sur la figure donnée ci-dessous, les syntagmes « Total fonds propres », « Total bilan » et « Caisses locales », présents dans le rapport 2000, ont disparu du dispositif utilisé en 2001 :



Figures 30 et 31.

Fiches signalétiques de la Caisse régionale de Crédit agricole de Paris et Ile-de-France pour l'année 2000 (figure 30) et l'année 2001 (figure 31)

Enfin, l'apparition de nouvelles dénominations constitue un troisième ensemble regroupant « Crédit Agricole S.A. », tête du groupe Crédit agricole cotée en bourse en 2001, et « CA Asset Management », société de gestion d'actifs rebaptisée en 2000 et dont la dénomination complète « Crédit Agricole Asset Management » tend à être supplantée en 2001 par la forme partiellement abrégée « CA Asset Management ». Si les formes qui correspondent à des millésimes n'appellent pas de plus amples commentaires, on peut, en revanche, rapprocher la sous-représentation de la particule verbale « a » de deux phénomènes. Le premier,

déjà été évoqué précédemment, concerne l'effritement des contextes syntactiques mis en évidence dans la section précédente, et dont la particule « a » constitue l'une des principales formes typiques. Le second phénomène concerne la restitution de l'activité, et plus exactement l'entité au nom de laquelle cette restitution est réalisée dans les rapports d'activité. Ce point sera traité à la fin de cette section.

Au début de cette section, on a émis l'hypothèse selon laquelle une variation importante des fréquences de formes existantes pouvait expliquer la répartition des différentes parties du corpus sur l'axe principal dégagé par l'AFC. Cette hypothèse est vérifiée au moins en ce qui concerne les vingt formes dotées des coefficients de spécificité les plus élevés pour l'année 2001. Modifications graphiques, aménagements des dispositifs textuels et apparition de nouvelles entités dont les dénominations réutilisent des formes existantes, tous ces facteurs concourent à la recomposition d'un fonds de vocabulaire préexistant à partir de la période inaugurée par l'année 2001. Ce réaménagement est à l'image de celui que connaît alors le Crédit agricole : faire du neuf – la cotation de Crédit Agricole S.A. –, avec du vieux, si l'on peut dire – le réseau des Caisses régionales de Crédit agricoles déjà centenaires. Si la nécessité de mettre en avant la nouvelle entité et ses filiales – entre autres, CA Asset Management –, semble impliquer une diminution de l'importance accordée à la description des Caisses régionales, le terme qui sert à désigner celles-ci est gratifié d'une majuscule¹. Ce faisant, l'étude de ces quelques unités caractéristiques de l'année 2001 permet d'introduire une thématique nouvelle : celle des formes linguistiques que peut revêtir la valorisation ou la dépréciation du rôle des principales entités dont le groupe bancaire se compose. C'est sous ce jour particulier que vont être examinées maintenant les formes caractéristiques de l'année 2003.

b. Intégration du groupe Crédit lyonnais

Le tableau ci-dessous reprend, comme pour l'année 2001 les vingt formes les plus caractéristiques du rapport d'activité de 2003.

¹ On reviendra sur ce point dans la troisième partie.

*Tableaux 21 et 22.
Principaux accroissements spécifiques positifs et négatifs de la partie RA2003*

Forme	Frq. Tot.	Fréq.	Coeff. ⁺	Forme	Frq. Tot.	Fréq.	Coeff. ⁻
Lyonnais	162	120	***	agricole	1108	7	***
2003	373	338	***	Indosuez	827	22	-49
Calyon	54	54	40	Total	688	15	-44
2004	51	51	37	Général	485	7	-36
euros	592	213	22	propres	453	16	-24
CLAM	29	29	22	locales	484	23	-21
général	217	97	18	bureaux	223	1	-20
AM	53	38	17	bilan	394	15	-20
d	3438	848	16	M	212	3	-16
S	437	150	14	Caisses	1047	107	-16

Lecture des tableaux 21 et 22 : Dans le tableau 21 sont rassemblées les dix premières formes dotées du coefficient de spécificité positif le plus élevé pour la partie 2003 ; dans le tableau 22 sont regroupées les dix dernières formes ayant le coefficient de spécificité négatif le plus élevé pour la partie 2003.

Comme dans la sous-section précédente, on commencera par rattacher les formes des tableaux aux syntagmes lexicalisés ou non dans lesquels elles figurent majoritairement. Pour les formes dotées d'un fort coefficient de spécificité positif, on peut établir que :

- « Crédit Lyonnais » incorpore la forme « Lyonnais »
- « 2003 » et « 2004 » correspondent aux années plus particulièrement prises en compte dans le rapport 2003, sachant que celui-ci relate des faits survenus en 2003 et au cours du premier trimestre 2004.
- « Directeur général » incorpore la forme « général »
- « CA-AM » incorpore la forme « AM »
- « Crédit Agricole S.A. » incorpore la lettre isolée « S »
- les expressions « milliards d'euros » ou « millions d'euros » incorporent les formes « d » et « euros ».
- « CLAM » est le sigle de Crédit Lyonnais Asset Management.

Pour les formes dotées d'un coefficient de spécificité négatif élevé, on relève les points suivants :

- L'abréviation « M € » (pour millions d'euros) intègre la lettre isolée « M »
- « Caisses locales » rassemble les formes « Caisses » et « locales »

- « bureaux permanents » incorpore la forme « bureaux »
- L'expression « Total bilan » rassemble les deux formes « bilan » et « Total »
- « fonds propres » incorpore la forme « propres »
- « Directeur Général » incorpore la forme « Général »
- « Crédit Agricole Indosuez » incorpore « Indosuez »
- « Crédit agricole » incorpore la forme « agricole ».

Il est facile de voir que les spécificités de l'année 2003 s'organisent de la même manière que celles de 2001. En effet, on retrouve des oppositions graphiques : « Directeur général » s'oppose à la graphie « Directeur Général » des précédentes parties, de même que « M € » s'oppose à « millions d'euros » ou « Crédit agricole » à « Crédit Agricole ». Le changement de dispositif concernant la présentation des Caisses régionales de Crédit agricole est confirmé par les sous-représentations affectant « Caisses locales », « Total bilan », « bureaux permanents » et « fonds propres ». En effet, comme on l'a vu précédemment, ces expressions font partie des fiches signalétiques des Caisses régionales présentes dans les premières parties du corpus. De ce point de vue, l'année 2003 se situe donc dans le prolongement de l'année 2001. Enfin, les nouvelles dénominations figurent en bonne position dans les spécificités positives avec « CLAM » et « Calyon ». Cette dernière entité résulte du rapprochement des activités de banque de financement et d'investissement du Crédit Lyonnais et de la filiale Crédit Agricole Indosuez, ce qui a motivé la création d'une nouvelle dénomination en remplacement de Crédit Agricole Indosuez. On note également la présence des formes « AM » et « Lyonnais » qui signalent l'apparition significative de « CA-AM » pour « Crédit Agricole Asset Management » et de « Crédit Lyonnais ».

Encore une fois, les bouleversements affectant l'organisation du groupe bancaire se traduisent dans les textes des rapports d'activité par la mise en avant de certains syntagmes, correspondant à des noms propres d'entités ou à des termes relatifs à l'activité de ces dernières. On dispose d'un indice sérieux, si ce n'est déjà d'une confirmation, du rôle pertinent que semblent jouer les noms propres comme révélateurs des changements affectant un référentiel spécifique. Il reste qu'à ce stade encore général de l'étude du vocabulaire des rapports d'activité, le phénomène qui

se dégage de l'étude du fonctionnement de la série textuelle chronologique tient dans les traces laissées par le changement de statut de l'émetteur des rapports d'activité.

3. Le changement de statut de l'émetteur

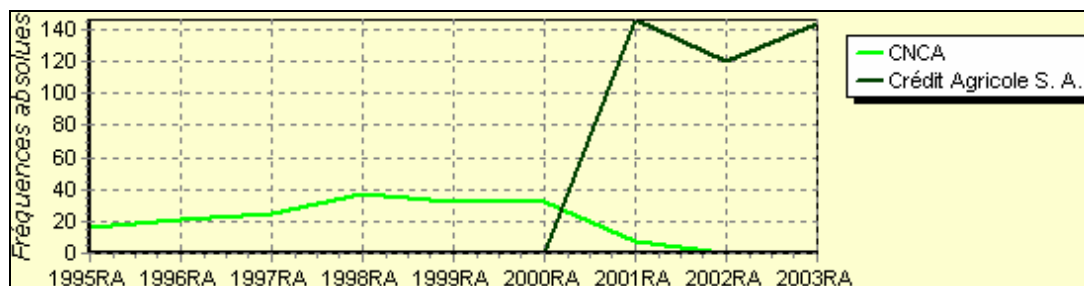
L'examen des formes plus particulièrement impliquées dans les mouvements en apparence discordants constatés à l'échelle du vocabulaire de l'ensemble du corpus (clivage entre les années 1995-2000 et les années 2001-2003, d'une part, et d'autre part, accroissement continu du vocabulaire), repose sur au moins trois sortes de causes. En effet, tout se passe comme si l'apparition d'entités nouvelles se traduisait d'abord par une recomposition du vocabulaire existant, les modifications de graphies, comme l'allègement de certains dispositifs textuels, faisant varier de manière très sensible la fréquence de formes présentes dans toutes les parties du corpus. Les mouvements conjugués de la modification importante de certaines fréquences et de l'apparition de nouvelles formes, caractérise un changement majeur qui oppose également l'année 2001 aux années 1996-1997 et 2003, qui sont caractérisées par les rachats de la banque Indosuez puis du Crédit Lyonnais. Ces achats ont surtout entraîné l'apparition de formes nouvelles sans provoquer une reconfiguration du vocabulaire équivalente à celle de l'année 2001.

Les deux achats s'inscrivent, en effet, dans un mouvement continu qui les a précédé, et que le vocabulaire exprimait déjà : le développement à l'international et la consolidation de la banque de détail en France. Ce qui change en 2001, et que l'examen de l'évolution du vocabulaire rend sensible, c'est le statut de l'émetteur : la Caisse Nationale de Crédit Agricole, organe intégré à l'organisation du groupe Crédit Agricole, et à ce titre porte-parole institutionnel, cède la place à Crédit Agricole S.A., société cotée en bourse, qui s'adresse au marché, d'abord en son nom propre et, en second lieu, au nom du groupe Crédit Agricole. On peut résumer ainsi le changement de statut de l'émetteur des rapports d'activité :

*Tableau 23.
Schéma du changement de statut de l'émetteur à partir de 2001*

Statut de l'énonciateur	Portée du discours
Caisse Nationale de Crédit Agricole - organe central - banque centrale - tête de réseau	Parle au nom du Groupe Crédit Agricole
2001	
Crédit Agricole S.A. - organe central - banque centrale - tête de réseau - société mère d'un groupe dont une partie du capital est coté en bourse, et qui comprend Crédit Agricole S.A., les principales filiales spécialisées du Crédit Agricole et 25% du capital des Caisses régionales de Crédit Agricole.	Parle au nom du groupe Crédit Agricole S.A. et du Groupe Crédit Agricole

Le constat selon lequel le changement de dénomination se double ici d'un changement de rôle que traduisent une augmentation ou une diminution des fréquences trouve une confirmation dans les ventilations de fréquence observées pour « CNCA » (Caisse Nationale de Crédit Agricole) et « Crédit Agricole S.A. » :



*Figure 32.
Ventilation des fréquences absolues de « CNCA¹ » et « Crédit Agricole S.A. » sur
l'ensemble des parties du corpus*

¹ Pour restituer la ventilation des fréquences de « CNCA » sur l'ensemble du corpus, on a au préalable constitué un groupe de formes rassemblant « CNCA » (107 occ.) et les deux variantes de son développé « Caisse nationale de Crédit agricole » (52 occ.) et « Caisse Nationale de Crédit Agricole » (12 occ.).

À la lecture de ce graphique, on peut en effet constater que la place accordée à la dénomination de l'organe central du Crédit Agricole de 1995 à 2000 est sans commune mesure avec celle qui est tenue par le même organe central après l'introduction en bourse de la fin 2001. L'évolution marquée de la fréquence de « CNCA » à partir de 1998¹ (elle passe de 11 à 20 occurrences), qui s'explique pour partie par un changement dans l'importance nouvelle prise par l'organe central du fait de l'acquisition de la banque Indosuez, reste nettement moins élevée que celle correspondant à l'année 2001. Tout se passe donc comme si une telle variation ne pouvait plus s'expliquer par un changement de rôle, partant, de la place tenue par la restitution de l'activité de telle entité participant à l'activité plus générale du groupe bancaire, mais bien par un changement de statut en rapport avec la situation d'énonciation caractérisant les rapports d'activité. La précédente section a fourni sur ce point un indice précieux qu'il convient d'étudier maintenant.

En effet, lors de la présentation des accroissements spécifiques de la partie « 2001 » par rapport à la période 1995-2000, on a relevé que la particule verbale « a » y est sous-représentée dans des proportions importantes (le coefficient de spécificité négatif est de 16). La diminution relative, à partir de 2001, des contextes à dominante syntactique, dont cette forme est aussi l'un des représentants typiques, constitue une première explication. Or, il semble qu'il soit maintenant nécessaire d'éclairer plus précisément le fonctionnement de cette particule verbale lorsqu'elle est en relation avec l'une des entités principales du groupe bancaire. Dans la mesure où les rapports d'activité restituent l'activité de la banque durant une année d'exercice, il est probable qu'une formule du type « Le Crédit agricole a fait telle chose telle année » se rencontre assez fréquemment. Si une formule générique de cet ordre connaît des variations de fréquences avant et après 2001, on aura un indice sérieux concernant le changement de statut de l'entité au nom de laquelle se fait la restitution de l'activité bancaire dans les rapports d'activité.

¹ Cette variation s'explique, pour partie, par quelques paragraphes consacrés à la célébration de la privatisation de la CNCA, établissement public avant 1988, et pour partie, par une montée en puissance du rôle de cette entité du fait de l'acquisition récente de la banque Indosuez. Les filiales créées à la suite de cet achat sont détenues à 100% par la CNCA, qui a ainsi renforcé la position du groupe bancaire dans les activités de gestion d'actifs et de banque de financement et d'investissement.

Afin de vérifier ce point, on a sélectionné les segments répétés les plus fréquents contenant à la fois un nom d'entité (Crédit agricole, CNCA, Crédit Agricole S.A.) et la particule verbale « a ». Un segment répété est « *une suite de formes dont la fréquence est supérieure ou égale à 2 dans le corpus*¹. » Le logiciel Lexico permet de repérer automatiquement les segments répétés présents dans un corpus de textes et d'en constituer un index, comme pour les formes simples. Dans la mesure où le calcul des segments est paramétrable, on a retenu, pour l'exploitation du corpus, non seulement les segments classiquement délimités par des séparateurs de séquences (ponctuations faibles et fortes, point, point d'exclamation, point d'interrogation, virgule, point-virgule, deux points, guillemets, tirets et parenthèses), mais aussi ceux pouvant se trouver à cheval sur de tels délimiteurs. L'utilité de cette option est présentée dans le chapitre de la troisième partie consacré à l'identification des dénominations propres. On retiendra ici que l'on capte de cette manière toutes les suites de formes présentes au moins deux fois dans les textes du corpus, ce qui représente 42 771 segments répétés dont les fréquences varient de 1699 occurrences à 2. Cette liste peut être ordonnée selon l'ordre textométrique, et il est possible de l'explorer en y recherchant des patrons spécifiques tels que ceux qui viennent d'être évoqués et dans lesquels la particule verbale « a » est associée à une dénomination propre.

Lorsque l'on recherche les segments répétés les plus fréquents (dont la fréquence est égale ou supérieure à 50 occurrences), comportant la particule « a », on obtient la liste suivante :

¹ L. Lebart et A. Salem (1994 : 317 et p. 58 sq).

Tableau 24.
Segments répétés de fréquence supérieur à 100 et comportant la particule « a »

Segment répété	Longueur	Fréquence
agricole a	2	272
Crédit agricole a	3	269
le Crédit agricole a	4	189
qui a	2	102
Crédit Agricole a	3	101
Il a	2	81
il a	2	80
Groupe a	2	70
Indosuez a	2	58
Crédit Agricole S A a	5	51

Il n'est pas anodin que lorsque l'on recherche les manifestations d'une formule exprimant au nom de quelle entité se fait la restitution de l'activité bancaire, ce soient les dénominations « Crédit agricole » et « Crédit Agricole S.A. » qui ressortent parmi une liste d'une dizaine de segments répétés très fréquents¹. De ce point de vue, l'absence de « CNCA » vient conforter l'hypothèse selon laquelle une telle formule, qui place le nom propre en position de sujet, ne concerne que les entités au nom desquelles se fait la restitution de l'activité : on a vu plus haut que tel n'est pas le statut de la CNCA dans les textes du corpus. À partir de ce repérage réalisé à l'aide des segments répétés, il est possible de comparer la ventilation des segments incluant les dénominations propres que l'on souhaite étudier. On utilisera également la ventilation du segment « qui a » comme test destiné à faire contraster les variations de fréquence des autres segments.

¹ Les segments incluant « Groupe » et « Indosuez » seront examinés dans la troisième partie.

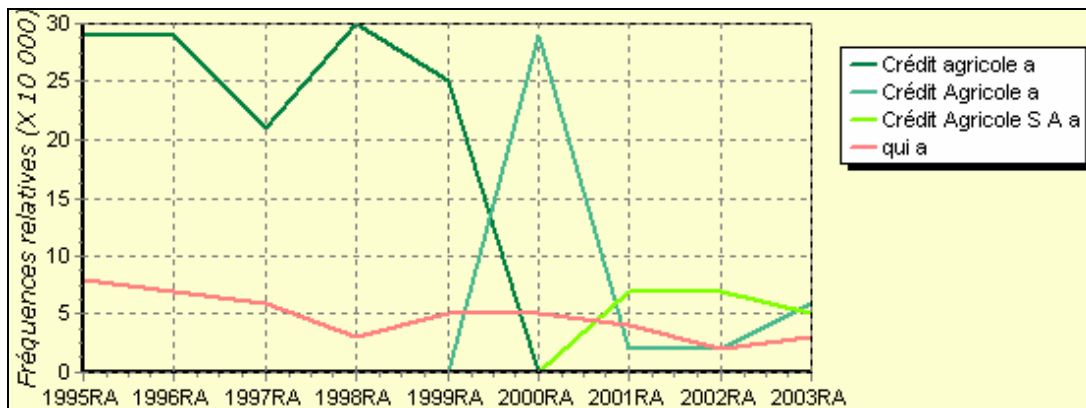


Figure 33.

Ventilation de quatre segments répétés comportant la particule verbale « a »

Le graphique ci-dessus permet de faire contraster la ventilation relativement régulière du segment « qui a » (en dehors de la partie 1995, il appartient au vocabulaire de base du corpus qui regroupe les formes et polyformes banales), avec les à-coups qui caractérisent les courbes des autres segments. Outre le remplacement d'une graphie par une autre concernant la dénomination « Crédit agricole » pour l'année 2000, le fait marquant qui ressort de ce graphique concerne les variations de fréquence pour les années 2001 à 2003. En effet, on constate que 2001 et 2002 sont caractérisées par le fait que « Crédit Agricole S.A. » semble se substituer à « Crédit Agricole » dès lors qu'il s'agit de dire au nom de qui se fait la restitution de l'activité bancaire. Il s'agit donc bien d'un changement de statut de l'émetteur des rapports d'activité qui, de simple instrument agissant au nom du Crédit agricole (CNCA), est hissé au rôle d'acteur à part entière. On note cependant que la relative brutalité du changement opéré en 2001 semble être corrigée de deux manières. La première concerne la visibilité du changement. Il ne s'agit pas d'une substitution à fréquence comparable : « Crédit Agricole S.A. » acteur paraît moins présent que « Crédit Agricole ». De fait, l'acteur qui a « pris la parole » n'est qu'une composante du groupe éponyme auquel il s'est substitué dans les textes des rapports d'activité. Les fréquences de la dernière partie semblent d'ailleurs suggérer l'amorce d'un rééquilibrage de la présence des deux acteurs à l'occasion du rachat du Crédit lyonnais.

III ACCROISSEMENTS SPÉCIFIQUES ET NOMS PROPRES

A. Spécificités évolutives

Compte tenu, d'une part, des observations de la précédente section relatives aux évolutions spécifiques et, d'autre part, de l'intérêt porté aux noms propres dans le premier chapitre, on a cherché à savoir dans quelle mesure ces unités donnent accès à la restitution des évolutions d'un référentiel spécifique. Afin de vérifier ce point, on a lancé un calcul des spécificités évolutives pour chacune des parties du corpus, hormis la première partie, qui ne compte pas de partie plus ancienne à laquelle elle puisse être comparée dans la série textuelle chronologique. On donne ci-dessous les quinze premières formes sélectionnées pour les huit parties du rapport d'activité auxquelles le calcul des spécificités évolutives a été appliqué.

*Tableau 25.
Quinze premières formes des accroissements spécifiques positifs de huit parties du corpus des rapports d'activité¹*

Formes 1996	Formes 1997	Formes 1998	Formes 1999	Formes 2000	Formes 2001	Formes 2002	Formes 2003
1996	1997	1998	euros	2000	A	2002	2003
Indosuez	Agricole	Agricole	M	Agricole	S	S	Lyonnais
Bureaux	général	Mds	€	Management	2001	A	Calyon
Banque	Indosuez	marquants	1999	Asset	Agricole	2003	2004
1997	Indocam	Faits	Fait	€	Régionales	Résultat	CLAM
périmètre	Ltd	Indosuez	Lyonnais	M	2002	Régionales	euros
4	Carr	1999	marquant	euros	CA	euros	général
clientèles	Securities	F	Indocam	2001	Administra. ^f	nomination	AM
pôle	W	Crédit	Agricole	FAIT	Asset	CA	d
5	1998	général	Sofinco	MARQUANT	Management	avant	S
juillet	Crédit	bureaux	2000	CPR	assurances	CAI	CDI
constant	Intesa	Une	Indosuez	Nationale	euros	Première	entreprise
Tel	Banca	Indocam	Comit	Lukas	Nommé	et	rapproch ^t
tandis	Asie	Cheuvreux	Groupe	Crédit	CRCAM	France	interne
acquisition	équipes	CAI	Par	Grèce	Membre	Source	conquête

Les données présentées dans ce tableau permettent de réaliser deux sortes de commentaires de nature à éclairer la suite de la démarche relative à la constitution

¹ La version complète et documentée des accroissements spécifiques positifs et négatifs des huit parties du corpus (1996-2003) est donnée en annexe.

d'un référentiel terminologique adaptable au contexte. Une première lecture consiste à proposer un regroupement des principales sortes d'informations apparaissant dans ce tableau. On a vu dans la précédente section que les spécificités évolutives, positives en l'occurrence, permettent d'isoler des formes qui appartiennent plus spécifiquement à la partie étudiée par rapport à la, ou aux partie(s) qui la précède(ent). Il découle de ce point que chaque partie ne comporte pas nécessairement les mêmes sortes d'informations, ou les mêmes quantités d'informations se rattachant à telle ou telle catégorie. On commentera plus loin cet aspect de la question pour n'aborder ici que le classement des formes en grandes catégories. On disposera ainsi d'une vision synthétique des sortes d'informations véhiculées par les rapports d'activité, mais également d'une forme d'introduction au vocabulaire du genre qui sera étudié plus précisément dans les chapitres suivants.

B. Part des noms propres dans les accroissements spécifiques

On propose de regrouper les formes graphiques du tableau ci-dessus, qui comporte quatre-vingt-dix occurrences, en six grandes catégories. La première de ces catégories est composée des années qui apparaissent de manière récurrente dans toutes les parties (1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004). Ces dates correspondent aux millésimes dont il est question dans chaque rapport d'activité. On remarque à ce propos que l'élaboration des rapports étant décalée de quelques mois par rapport à l'arrêté comptable du 31 décembre de l'année précédente, les faits se rapportent en gros à une période couvrant l'année ainsi qu'un trimestre de l'année suivante.

La seconde catégorie est composée de bribes appartenant à des coordonnées ou à un chiffrage (4, 5, Tel, Bureaux, bureaux). En effet, on a pu voir dans la première section que les rapports d'activité comportent des contextes à dominante non syntactique dans lesquels figurent un certain nombre de tableaux et histogrammes, ainsi que les coordonnées de certaines sociétés du groupe bancaire. Les coordonnées des Caisses régionales de Crédit Agricole sont fréquemment complétées par la mention du nombre de bureaux, c'est-à-dire d'agences, qui relèvent de chaque banque régionale.

Une troisième catégorie regroupe des formes appartenant à des dénominations complexes correspondant à des noms propres. Le rattachement de ces formes, déjà présenté dans la précédente section, sera traité de façon détaillée dans la troisième partie. Pour le moment, on se bornera à relever que cette catégorie d'unités est de loin la plus nombreuse, puisqu'elle rassemble 35 formes sur 87, soit 40%, et 43% des occurrences (52 sur 120). Il s'agit des formes suivantes : « Indosuez, Banque, Agricole, Indocam, Ltd, Carr, Securities, W, Crédit, Intesa, Banca, Asie, Cheuvreux, CAI, Lyonnais, Sofinco, Comit, Groupe, Management, Asset, CPR, Nationale, Lukas, Grèce, A, S, Régionales, CA, Asset, CRCAM, CAI, France, Calyon, CLAM, AM ».

Une quatrième catégorie, plus hétérogène, ressemble différentes sortes d'outils. Outre les classiques mots grammaticaux, on a regroupé ici des intitulés de rubriques récurrents, mais connaissant des variantes (c'est la série des « faits marquants »), ainsi que des éléments de date précise, des abréviations, des unités monétaires ou encore des mentions récurrentes comme « Source », qui introduisent le nom de la source journalistique ou institutionnelle émettrice d'une information citée. Il s'agit des formes « juillet, tandis, Mds, marquants, Faits, F, Une, euros, M, €, Fait, marquant, Par, FAIT, MARQUANT, avant, et, Source, d ».

La catégorie « titres et assimilés » rassemble les mentions de titres des membres des différents organes de direction du groupe bancaire. On a également associé à cette liste « nomination » et « nommé », qui apparaissent systématiquement dans les mêmes contextes. Ce sont les formes « général, Administrateur, Nommé, Membre, nomination ».

Enfin, la catégorie « vocabulaire » rassemble les formes qui présentent la caractéristique négative de ne pas appartenir aux autres catégories, et qui, de façon plus positive, constituent des mots utilisés pour décrire l'activité bancaire telle qu'elle est restituée dans les rapports d'activité. Les substantifs y sont majoritairement présents et une partie des mots de la liste suggère que les activités commerciales sont évoquées (« clientèles, acquisition, Résultat, entreprise, conquête ») ainsi que le personnel d'une entreprise (équipes, CDI). Cette catégorie comporte les formes « périmètre, clientèles, pôle, acquisition, équipes, assurances, Résultat, CDI, entreprise, rapprochement, interne, conquête, constant, Première ».

Ces formes confirment tout d'abord le caractère factuel et opérationnel des informations contenues dans les rapports d'activité, qui se rattachent ainsi sans grande difficulté au genre de discours économique. Elles mettent également en évidence la palette assez riche des moyens utilisés pour rendre compte de l'activité du groupe bancaire. Les coordonnées spatiales et temporelles cohabitent non seulement avec les chiffrages liés à l'activité bancaire, mais aussi avec des informations relatives à l'organisation de l'entreprise (entités appartenant probablement au groupe bancaire, organigramme). C'est dire que, contrairement à ce à quoi on aurait pu s'attendre, la restitution de l'activité elle-même ne semble pas occuper une place centrale, à moins que cet aspect étant au contraire le plus récurrent, les mots utilisés n'apparaissent que relativement peu dans les accroissements spécifiques, qui mettent l'accent sur les nouveautés plutôt que sur les éléments appartenant à « la toile de fond » dont se compose le vocabulaire du corpus. Il convient donc d'essayer de cerner maintenant la part des évolutions d'un référentiel que le calcul des spécificités permet de saisir.

De ce point de vue, toutes les parties du corpus n'apportent pas les mêmes sortes d'informations. On a vu dans la précédente section que deux événements majeurs sont intervenus, l'un en 2001 – la cotation du groupe bancaire –, et l'autre en 2003 – le rachat du Crédit lyonnais. Les cinq nouvelles formes spécifiques apportées par le tableau (on passe de dix formes spécifiques à quinze), viennent confirmer le retentissement de ces changements sur le vocabulaire des rapports d'activité. On a ainsi, pour 2001, « Nommé » et « membre » qui renvoient à la constitution d'un nouveau conseil d'administration. De même pour 2003, « CDI », « rapprochement » et « interne » sont autant de mots qui renvoient au contexte de l'acquisition du Crédit lyonnais par le Crédit agricole. On peut alors se demander s'il en est de même pour les autres parties du corpus.

Il arrive, comme pour l'année 2002, que l'actualité devenant moins prégnante, on assiste à une sorte de consolidation du vocabulaire, qui reprend les formes de la partie précédente ou qui vient compléter des thématiques semblables. Cependant, d'une manière générale, l'activité du groupe bancaire semble avoir été soutenue durant la dizaine d'années couverte par le corpus. L'acquisition de la

banque Indosuez, mentionnée explicitement en 1996 (« acquisition, Indosuez, Banque, périmètre constant [modification du périmètre de consolidation du bilan], pôle, clientèles [la banque constitue le nouveau pôle « grandes clientèles » du Crédit agricole]), continue à produire des effets en 1997 et 1998, en raison de l'apparition de nouvelles entités dans le groupe bancaire (Crédit Agricole Indosuez, Indocam, Ltd, Carr, Securities, W, Asie, Cheuvreux, CAI). Les années 1999 et 2000 sont marquées également par le passage à l'euro et à l'an 2000 et, pour le groupe Crédit agricole en particulier, par de nouvelles acquisitions ou prises de participation (Lyonnais, Sofinco, Comit, CPR, Lukas, Banque Commerciale de Grèce). De ce point de vue, la présence systématique de noms propres de personnes morales dans les formes spécifiques signale que ce type d'unités fonctionne comme un bon indicateur des événements institutionnels qui agitent la vie des entreprises. Ces observations confirment celles qui ont été réalisées précédemment pour les années 2001 et 2003 : les variations de fréquence des noms propres sont liées aux principaux changements qui sont susceptibles d'affecter un référentiel d'objets.

Deux sortes de tests textométriques ont été appliqués au corpus et ont permis de contrôler certaines évolutions notables du vocabulaire. En effet, on a pu voir que, tant dans sa répartition en contextes syntactiques et non syntactiques, que dans son évolution sur la durée, les caractéristiques du vocabulaire des rapports d'activité se transforment au cours du temps. Les tests ont non seulement permis de repérer ces évolutions, mais aussi de les caractériser. Pour ce qui est des différentes sortes de contextes, on peut faire deux constats. Le premier concerne le rôle particulier dévolu aux CDN d'une manière générale. Comme ceux-ci reçoivent des unités qui sont susceptibles de se passer de toute prédication (au moins directe, c'est-à-dire sur l'axe syntagmatique, car on peut considérer que les intitulés de colonnes pour les tableaux, les artifices graphiques des organigrammes, etc., proposent une forme de prédication abrégée et partagée par plusieurs unités simultanément), il n'est pas surprenant qu'ils jouent un rôle de conservatoire pour les entités qui appartiennent depuis déjà quelque temps au référentiel commun. Cependant, ce rôle est lui-même susceptible d'évoluer.

En effet, la rupture introduite par l'année 2001 se traduit par une diminution de l'homogénéité des contextes syntactiques au profit d'enclaves non syntactiques. Le vocabulaire de ces enclaves est différent de celui des contextes non syntactiques de la première période (1995-2000), sans être identique à celui des contextes syntactiques. Ce phénomène permet d'identifier l'émergence d'un nouveau type de contexte, lié au renforcement d'un mode de traitement particulier de l'information prenant la forme de « coups de projecteurs », qui peuvent s'apparenter à la réclame et révèlent un renforcement de la visée communicationnelle destinée à séduire une clientèle grand public. Ce changement de présentation, issu pour une partie de l'information diffusée dans les zones textuelles à dominante syntactique, a pour conséquence prévisible un appauvrissement de l'information associée aux entités nommées. Dans la mesure où cette tendance amorce un léger recul dans la dernière partie du corpus, elle n'a pas eu de retombées significatives sur le balisage, mais elle reste un point à surveiller dans le cadre d'un corpus de suivi.

Par ailleurs, le suivi de l'évolution chronologique du vocabulaire a permis d'identifier, d'une part, un clivage du corpus en deux ensembles (les années 1995-

2000 et les trois dernières parties couvrant la période 2001-2003) et, d'autre part, une évolution chronologique atypique du vocabulaire, dans la mesure où le nombre de formes nouvelles apparaissant chaque année ne diminue pas. L'étude des accroissements spécifiques des années 2001 et 2003 vérifie l'hypothèse d'une apparition de formes nouvelles doublée d'une recomposition du vocabulaire des premières parties. Ces phénomènes sont en relation avec les bouleversements qui affectent alors le référentiel de l'entreprise et se répercutent jusque dans la graphie des dénominations d'entités. Ainsi, l'émergence d'une nouvelle entité liée à la cotation – Crédit Agricole S.A. – se fait-elle au détriment du traitement jusqu'alors réservé aux entités plus anciennes, mais de façon peu visible pour un regard extérieur. C'est l'intérêt des tests textométriques que de faire émerger les interactions entre les bouleversements du référentiel et l'évolution du vocabulaire. Ainsi, l'achat du Crédit Lyonnais se traduit-il de façon typique, pour le vocabulaire caractéristique de la partie 2003, par l'apparition de nouveaux noms d'entités (Crédit Lyonnais, Calyon, CLAM).

Les observations réalisées à partir du calcul des spécificités évolutives des années 2001 et 2003 ont ensuite été étendues aux autres parties du corpus. L'observation des formes ainsi collectées a permis de commencer à caractériser le vocabulaire des rapports d'activité. Celui-ci se compose de diverses catégories d'unités révélant qu'à côté de la restitution de l'activité bancaire à proprement parler, sont également présentes, d'une part, l'insertion de l'activité dans un référentiel spatiotemporel partagé et, d'autre part, la présentation de certains aspects plus spécifiques du référentiel propre au Crédit Agricole, à travers la mention d'un certain nombre d'entités nommées participant à l'évolution de l'activité du groupe bancaire. Ce dernier point confirme donc la capacité des noms propres à révéler certains aspects de l'évolution du référentiel du groupe bancaire. Mais avant de pousser plus avant les investigations sur ce point, il convient de pouvoir situer le vocabulaire des rapports d'activité par rapport aux différentes strates du parler d'entreprise qu'ils mobilisent.

CONCLUSION DE LA PREMIÈRE PARTIE

La proposition de mettre en place des référentiels terminologiques adaptables aux contextes suppose d'interroger certaines conceptions, qui font obstacle à la prise en charge effective des difficultés constatées dans la réalisation des échanges communicationnels sur les intranets d'une grande entreprise. Il s'agit tout d'abord de sortir l'intranet de son terreau technologique, afin de le replacer dans le cadre de l'organisation du travail et, plus spécifiquement, des activités d'information auxquelles sont astreints les agents du secteur tertiaire pour l'accomplissement de leurs tâches. Dès lors, l'intranet peut être conçu comme un espace au sein duquel se déploient des pratiques ayant pour but l'échange médiatisé d'informations à usage professionnel, entre des agents appartenant à une même entreprise et reliés entre eux par un réseau informatique privé ayant recours aux technologies du Web en usage sur Internet. Une telle conception permet de situer l'intranet dans la lignée des systèmes d'organisation de l'information, mais les pratiques mises en œuvre dans ces derniers ne paraissent pas pouvoir lui être appliquées. On a donc cherché à identifier les causes de cette situation paradoxale.

Il est apparu que l'usage de l'intranet, dès lors qu'il est étroitement associé aux activités de production des entreprises, remet en cause non seulement l'ontologie des systèmes d'organisation de l'information, mais aussi les conditions matérielles nécessaires à leur réalisation. Ce faisant, il a été nécessaire d'identifier un modèle général capable de rendre compte de cette forme d'échanges, mais aussi de fournir une base pour la solution que l'on cherche à promouvoir. Cela a été l'occasion de remettre en cause le modèle de la communication dominant dans les sciences de l'information, pour lui substituer un modèle élaboré dans un cadre linguistique et, plus précisément, au sein des théories de l'énonciation. Dès lors, les activités d'information, relatives aux échanges recourant aux sites intranets dans les organisations, ont pu être rattachées à un modèle général, rendant compte des différentes composantes de toute situation de communication verbale. Les référentiels terminologiques sont alors adaptables au contexte à partir du moment où ils permettent de prendre en compte et de restituer une image linguistique correspondant à chacun des publics venant visiter un site intranet. La construction de

tels référentiels terminologiques doit alors s'appuyer sur une étude des discours déjà produits à destination de ces publics, dans le cadre de l'organisation.

La diffusion d'un référentiel spécifique en direction d'un public néophyte est loin de faire l'objet d'une prise en compte particulière dans les écrits diffusés dans les organisations. Ce phénomène peut être expliqué à l'aide des mêmes raisons qui ont servi à rendre compte des carences constatées en matière de diffusion de l'information sur les intranets. Il reste qu'un certain nombre de productions ont pour fonction de communiquer en direction d'un tel public. Il s'agit des écrits de communication, que celle-ci soit publicitaire, commerciale ou institutionnelle. Cette dernière forme de communication s'exprime, entre autres, à l'aide de discours tels que les rapports d'activité. Destinés à un large public aussi bien interne qu'externe à l'entreprise, ces documents ont été retenus afin de documenter les activités de diffusion d'information sur les sites intranets, dès lors que celles-ci visent aussi des publics nouveaux. Lorsque ce ne sont plus le fonds documentaire ou la littérature du domaine qui peuvent servir de base à l'élaboration d'un référentiel terminologique, il convient de leur substituer une entité plus malléable, la composante référentielle présente dans toute situation de communication.

Cependant, une telle option se heurte rapidement à une réelle difficulté, lorsque l'on cherche à faire reposer sur elle la collecte d'unités précises, tirées de textes organisés en série textuelle chronologique. Quelles unités sélectionner, sur quelles bases observer les variations affectant un parler d'entreprise, selon qu'il est mobilisé dans telle ou telle situation de communication ? Afin de répondre à ces questions, on a fait une double hypothèse : un référentiel peut être saisi au moyen de ses noms propres, et le vocabulaire caractéristique d'un corpus peut être capté à l'aide de ces unités. Ces hypothèses reposent sur l'observation, déjà ancienne, selon laquelle l'usage des noms propres indique que des particuliers ont été dénommés de cette manière spéciale au sein d'un groupe social, pour des raisons relevant de la pragmatique. Les conditions que tout être ou objet doit remplir pour bénéficier, en plus de sa catégorisation ordinaire, d'une dénomination spécifique, le désignent comme devant occuper une place privilégiée au sein d'un référentiel. On propose alors de faire reposer la constitution des référentiels terminologiques adaptables au contexte sur une exploration des corpus utilisés comme source, à l'aide de cette

unité particulière que constituent les noms propres. À ce stade, il semble nécessaire d'évaluer la collecte qui pourra être ainsi réalisée. Son évaluation devra répondre aux deux questions soulevées par cette approche : dans quelle mesure une collecte ainsi conduite permet-elle d'accéder au vocabulaire caractéristique du corpus utilisé comme source, et quels éléments du référentiel que l'on cherche à capter permet-elle de saisir ?

Répondre à ces questions suppose que l'on se dote d'une démarche explicite, qui puisse être reproduite autant de fois que nécessaire dans le cadre de la mise en place de référentiels terminologiques adaptables au contexte à l'échelle d'une organisation telle que le groupe bancaire qui a été retenu au titre d'illustration. Une première réponse a été apportée dans cette partie, par la constitution de corpus dont les propriétés communicationnelles, proches de la situation de communication à documenter, ont été décrites et ont pu faire l'objet de contrôles textométriques. Cette évaluation permet de caractériser à un premier niveau les limites de la collecte qui est projetée : celle-ci sera d'abord dépendante de la manière dont l'activité bancaire est verbalisée dans le cadre du genre de discours représenté par un corpus de rapports d'activité. Une telle procédure permettra également d'éviter, à terme, les redites entre corpus et référentiels terminologiques et facilitera la circulation des exemples types d'usage d'un parler d'entreprise. Par ailleurs, l'étude de l'évolution du vocabulaire dont se composent les différentes parties du corpus principal a montré que les noms propres participent aux évolutions affectant le référentiel du groupe bancaire utilisé comme contexte.

L'approche textométrique retenue pour l'exploitation du corpus a permis à ce stade de disposer d'une vision globale sur les propriétés des textes qui ont été rassemblés. La segmentation de ces derniers en formes graphiques rend possible des comptages et des comparaisons qui n'auraient pas pu être possibles autrement. On peut ainsi avoir une vue d'ensemble sur l'accroissement du vocabulaire du corpus et sur la manière dont les stocks des différentes parties se renouvellent d'une année sur l'autre. L'observation de certaines anomalies, repérables par rapport à un modèle théorique attendu lorsque l'on applique une analyse factorielle des correspondances à une série textuelle chronologique, ont conduit à interroger plus avant les répercussions de changements intervenus dans la situation de communication au sein

de laquelle prend place la production des rapports d'activité. L'identification de deux sortes de contextes, à dominante syntactique et non syntactique, a non seulement pu être formalisée, mais le contrôle de cette répartition a été rendu possible à l'échelle du corpus entier, comme à celle d'unités typiques. En effet, l'approche textométrique ne se contente pas de donner accès à une vision clarifiée sur des stocks de vocabulaire relativement volumineux. Elle facilite une mise en relation constante entre les observations macroscopiques et l'étude du fonctionnement discursif d'unités précises à l'échelle d'un paragraphe, d'une phrase ou d'un syntagme.

Deux autres aspects de la réponse doivent maintenant être traités. Il s'agira, tout d'abord, de caractériser le vocabulaire du corpus en tant qu'il offre un exemple d'utilisation d'un parler d'entreprise particulier. Sans cette étape, il ne serait pas possible d'apporter une réponse concernant le caractère représentatif ou non de la collecte réalisée à partir des dénominations propres. Cet aspect de la problématique fait l'objet de la deuxième partie de cette étude. Lorsque cette étape de la démarche aura été franchie, il conviendra d'aborder les questions soulevées par la réalisation de la collecte elle-même, ainsi que l'évaluation des résultats obtenus. Ces différents points seront traités dans la troisième partie.

DEUXIÈME PARTIE

« PARLER D'ENTREPRISE » ET COMMUNICATION INSTITUTIONNELLE

INTRODUCTION DE LA DEUXIÈME PARTIE

Dès lors que la source textuelle devant être utilisée pour la construction d'un référentiel terminologique a été identifiée, décrite et rendue exploitable par des programmes informatiques, il est nécessaire de caractériser son vocabulaire en tant que manifestation d'un parler d'entreprise spécifique. On rappelle que, dans sa thèse, D. de Vecchi a défini cette notion de la manière suivante : « (...) nous parlerons désormais de parler d'entreprise ou PE pour désigner l'ensemble des processus linguistiques qui actualisent les répertoires linguistiques des membres d'une communauté, définie en fonction de l'appartenance à une entreprise. Autrement dit, la cristallisation linguistique de tout moyen de communication mis à la disposition d'une entreprise, pour des conceptualisations ayant des origines diverses¹. » La caractérisation du vocabulaire des rapports d'activité regroupés dans le corpus RA est destinée à positionner cette ressource dans un programme plus vaste, qui vise une forme de cartographie des manifestations d'un parler en fonction de situations de communication caractéristiques. L'idée consiste ici à préparer le partage de telles ressources, mais aussi à conférer une certaine visibilité à leur construction, de manière à éviter, par exemple, les doublons. Cette caractérisation vise également à permettre une évaluation de la collecte projetée à partir des noms propres présents dans le corpus. En effet, on a vu plus haut que cette collecte doit être évaluée sous le rapport de sa propension à capter le vocabulaire caractéristique du corpus source.

On a vu dans la première partie que les rapports d'activité occupent une position particulière dans l'ensemble des discours produits par l'entreprise. En effet, ils sont destinés à un public principalement externe, qui est rarement au fait des caractéristiques de l'activité de l'émetteur. La communication institutionnelle a pour tâche, entre autres, d'exposer une partie du référentiel de l'entreprise auprès d'un public qui l'ignore généralement. Du point de vue de la constitution d'un référentiel terminologique, il apparaît également nécessaire de chercher à identifier les procédés linguistiques utilisés par le destinataire afin de se faire comprendre de ses lecteurs. En effet, l'usage d'un parler d'entreprise dans les rapports d'activité impose une

¹ D. de Vecchi (1999 : 316).

double contrainte au locuteur collectif « direction générale », en ce qui concerne la sélection des unités de son lexique. D'une part, il paraît difficile d'évoquer ce que l'activité d'une entreprise a de particulier par rapport à ses concurrentes, sans utiliser des unités propres à son parler, d'autre part, cette réalité ne peut pas non plus être évoquée en des termes qui risqueraient d'être étrangers au public aussi bien interne qu'externe auquel les rapports sont destinés. À ce titre, l'étude d'un parler à partir des discours des rapports d'activité revêt un caractère exemplaire, dans la mesure où l'on peut y observer les traces de prise en charge énonciative laissées par une situation « d'insécurité linguistique ». Or, c'est précisément cette problématique que l'on cherche à aborder pour la constitution de référentiels terminologiques. En effet, ces derniers sont destinés à documenter des activités d'information à partir d'un retour sur les usages langagiers enregistrés dans des situations de communication similaires.

L'examen des procédés linguistiques utilisés par les « discours seconds »¹ ne vise pas à déterminer si l'énonciateur des rapports d'activité en a fait bon usage. Une telle étude ne saurait être menée ici. On se bornera à constater que la communication institutionnelle, confrontée à une situation analogue à celle que l'on rencontre sur certains sites intranets de l'organe central du Crédit agricole, constitue l'un des rares cadres où la prise en charge des lacunes cognitives d'un public néophyte soit explicitement assumée. On rappelle, par ailleurs, que la constitution de référentiels terminologiques adaptables au contexte repose sur une contrainte importante. Celle-ci consiste à sélectionner, parmi les sources textuelles existantes, celles qui sont les plus adéquates à la question traitée, et non pas à définir ce que serait la ressource idéale pour améliorer la diffusion de l'information dans un contexte précis.

Par ailleurs, la caractérisation des manifestations d'un parler se pose en des termes différents selon les sources utilisées. Les propriétés des rapports d'activité positionnent ces derniers sur plusieurs frontières : interne et externe, référentiel spécifique et référentiel partagé et, en définitive, parler d'entreprise et langue commune. De ce point de vue, le parler d'entreprise consiste plutôt en un feuilletage

¹ Cette problématique est reprise dans le chapitre VII, où l'on adopte la terminologie proposée par J. Authier (1982).

de variétés de langue assemblées de façon particulière en fonction des situations de communication. On doit donc maintenant identifier ces différentes strates lorsqu'elles se manifestent dans les productions d'un genre particulier, les rapports d'activité. Dans cette perspective, la constitution d'une ressource terminologique privilégiera la dimension lexicale du parler d'entreprise. Celle-ci représente d'ailleurs, d'après D. de Vecchi, la dimension la plus caractéristique d'un parler (par rapport aux dimensions phonétique, morphologique ou grammaticale). Par conséquent, parmi les différentes unités appartenant au parler, on retiendra plus particulièrement celles qui sont les mieux à même de restituer le référentiel en jeu dans la situation de communication¹. Il s'agit principalement des substantifs entendus au sens large, qu'il s'agisse de noms communs ou de noms propres, de termes simples ou complexes, ou encore d'unités non linguistiques que l'on dénommera brachygraphiques² (sigles, lettres, symboles, etc.), de syntagmes, voire de phrases lorsque celles-ci jouent le rôle d'une dénomination.

Or, même quand elle est réduite à un sous ensemble des parties de discours, la caractérisation de la composante lexicale d'un parler doit être située dans ce que l'on pourrait appeler l'emboîtement des lexiques : « *On peut ainsi concevoir une série d'ensembles qui méritent tous le nom de « lexiques » et dont chacun contient le suivant : lexique d'une langue (au sens extensif) dans certaines limites chronologiques ; lexique de la même langue en synchronie ; lexique d'un groupe humain limité appartenant à la communauté linguistique ; lexique d'un individu appartenant à ce groupe ; lexique de cet individu dans une situation stylistique donnée, ou dans une situation thématique donnée ; lexique du même, dans une situation stylistiquement et thématiquement définie*³. » Cette réflexion de Ch. Muller permet de préciser qu'un lexique, à quelque niveau qu'on le prenne, hérite partiellement des unités lexicales issues des niveaux supérieurs. Le « lexique de

¹ Ces unités pourront être utilisées comme mots clés destinés à guider la diffusion d'un contenu, mais aussi comme intitulés pour les branches d'arborescence organisant l'information sur un site, ou encore d'unités destinées à alimenter un dictionnaire dédié à un programme informatique reposant sur un TALN (traitement automatisé du langage naturel), et permettant de réaliser des expansions de requête ou des catégorisations de l'information.

² R. Kocourek (1991 : 93-94).

³ Ch. Muller (1967 : 18).

situation », qui correspond au niveau le plus fin de la série (c'est, dans cette étude, le vocabulaire formé par l'ensemble des mots du corpus), incorpore non seulement des mots propres au parler et aux terminologies qui le traversent, mais aussi des vocables relevant de la langue commune. Ce faisant, le fait de rendre compte de l'usage d'un parler, tel qu'il se manifeste dans les réalisations propres à un genre de discours, revient à décrire la manière dont les différentes strates linguistiques dont il se compose s'articulent pour former le vocabulaire d'un corpus. La caractérisation de cet usage se fera à l'aide d'unités typiques plutôt qu'à partir d'un relevé exhaustif, dont on a vu plus haut qu'il ne constitue pas l'objectif assigné à la constitution des référentiels terminologiques adaptables au contexte.

Afin d'identifier les variétés de langue présentes dans les textes du corpus, il sera nécessaire de confronter le vocabulaire de ce dernier à des normes représentant la langue commune (chapitre 4), aussi bien que la terminologie propre aux domaines bancaire et financier (chapitre 5). Cela permettra ensuite de dégager les particularités de la phraséologie des rapports d'activité, dans ses aspects à la fois synchronique et diachronique (chapitre 6). Enfin, il conviendra de s'interroger sur les formes de prise en charge énonciative présentes dans les rapports d'activité. Celles-ci seront abordées à la lumière des opérations combinées d'éliminations stylistiques et de sélections thématiques que l'on aura pu identifier dans le vocabulaire du corpus (chapitre 7).

CHAPITRE 4

LANGUE COMMUNE ET BANQUE GRAND PUBLIC

Dans la mesure où les différentes variétés de langue dont se compose un parler d'entreprise sont susceptibles d'être présentes à des degrés variables dans les productions discursives d'une organisation, on a cherché à déterminer la palette des usages linguistiques mis à contribution dans les rapports d'activité. Pour cela, il a été nécessaire de recourir à des artifices qui, à la manière d'un prisme, ont permis d'isoler telle ou telle variété de langue. On s'attachera ici à déterminer l'usage qui est fait dans les rapports d'activité de la langue commune. Plus précisément, il s'agit de faire émerger du vocabulaire du corpus les unités lexicales relevant de la langue commune et qui ont été mises à contribution pour restituer l'activité du groupe bancaire. Le champ d'investigation a été resserré sur certains types d'unités parmi lesquelles les substantifs, entendus au sens large, car ils sont destinés à former le principal de la collecte envisagée. Afin d'isoler ces mots parmi les autres composantes du vocabulaire du corpus, on a cherché à utiliser une sorte de réactif permettant de révéler des sélections propres à la situation de communication étudiée.

Aborder la question de la langue commune suppose au préalable que cette notion ait été précisée, de même que les caractéristiques de « l'échantillon représentatif » que l'on utilisera comme norme de référence. Ce dernier sera employé de manière à faire ressortir, par comparaison, les mots du corpus dont l'usage est anormalement élevé par rapport à ce que l'on pourrait s'attendre à trouver dans des textes à dominante littéraire. Les résultats obtenus seront documentés par des procédures textométriques, afin de vérifier le fonctionnement réel de ces vocables dans les textes du corpus. Il s'agira alors d'établir une liste de « mots-clés » caractéristiques de l'usage de la langue commune dès lors qu'elle est utilisée pour rendre compte de l'activité bancaire. On proposera alors une caractérisation de l'utilisation de cette variété de langue dans le corpus de rapports d'activité.

I. LA LANGUE COMMUNE, UNE QUESTION DE NORME

A. Valeurs approchées

En procédant de la sorte, on cherche à confronter les éléments composant un vocabulaire particulier – celui du corpus étudié – à une norme externe, c'est-à-dire définie à partir d'autres textes que ceux du corpus, ou tout au moins les englobant largement. Cette norme repose sur la fréquence constatée pour les unités lexicales dans un échantillon jugé représentatif de la langue¹. Elle fonctionne comme norme lorsque la fréquence d'une unité dans ce corpus de référence est comparée à la fréquence observée pour la même unité dans le corpus étudié. Cette comparaison, pour ne pas s'arrêter au constat d'une simple différence de fréquence, doit avoir recours à un modèle probabiliste. En effet, celui-ci permet de porter un jugement sur la fréquence observée dans le corpus étudié : correspond-elle à ce qui était attendu, est-elle en deçà ou delà de ce que la norme pouvait laisser prévoir ? En matière de statistiques lexicales, deux modèles ont principalement été utilisés². Il s'agit, chronologiquement de la loi normale, avec calcul de l'écart réduit³, et du calcul

¹ On ne cherche pas ici à rouvrir le débat sur les fréquences en langue, dont on pense, avec entre autres M. Tournier (1980), qu'elles n'existent pas. Il a semblé néanmoins fructueux d'utiliser certaines méthodes mises au point dans un cadre où la langue était le point focal des recherches, en réduisant les ambitions à l'utilisation de réactifs susceptibles d'interroger un vocabulaire particulier.

² Les deux modèles utilisent les mêmes paramètres, qui sont au nombre de quatre : la taille du corpus de référence, la taille du corpus étudié, la fréquence du mot dans le corpus de référence et sa fréquence dans le corpus étudié.

³ Par exemple P. Guiraud (1954) et C. Muller (1968). La formule de l'écart réduit est la suivante :

$$z = \frac{\text{écart absolu}}{\text{écart type}} = \frac{k - fp}{\sqrt{fpq}}, \text{ où } k \text{ est la fréquence observée dans le corpus étudié pour une unité}$$

donnée, f est la fréquence observée dans le corpus de référence pour la même unité, p est l'étendue du corpus étudié divisée par l'étendue du corpus de référence (fp correspond à la fréquence théorique), q est égale à $1-p$. L'écart réduit obtenu est ensuite rapporté à une table dans laquelle pour des valeurs types de l'écart réduit correspond une probabilité plus ou moins importante. On a, par exemple, pour un écart réduit de $z = 1,0$ une probabilité associée de 0,317 (soit 31,7% de chances de voir l'événement survenir, ce qui est une probabilité importante qui ne permet pas de rejeter l'hypothèse nulle si l'événement survient), alors qu'à un écart réduit de $z = 2,5$ correspond une probabilité de 0,012 (soit 1,2% de chances de voir l'événement survenir, ce qui est une probabilité faible qui permet de rejeter l'hypothèse nulle si l'événement survient). Table fournie, par exemple, dans Muller ([1977] 1992 : 198).

hypergéométrique indiquant pour chaque forme sa spécificité positive ou négative¹, ce dernier calcul s'étant révélé mieux adapté « à la population discrète des occurrences de vocabulaire, et valide pour toute la gamme des fréquences rencontrées². » Dans les deux cas, on obtient une évaluation sur la fréquence constatée pour une forme donnée dans le corpus étudié, en ce sens que l'on peut juger que cette forme a une fréquence banale, supérieure ou inférieure à la norme.

La difficulté de ce type de comparaison ne tient pas tant aux calculs qu'à la constitution préalable d'un corpus de textes censés donner un échantillon représentatif de la langue. Autant dire tout de suite qu'il n'y a pas en la matière de situation satisfaisante. Ce point n'a pas été jugé rédhibitoire, dans la mesure où l'on n'attend pas de cette comparaison la possibilité d'extrapoler sur l'usage de la langue commune en général. Il s'agit d'utiliser un réactif capable, compte tenu de ses caractéristiques propres, de faire ressortir une strate particulière du vocabulaire du corpus. Il est nécessaire de commencer par expliciter les limites du corpus de référence, de manière à en tenir compte au moment du commentaire des résultats obtenus à la suite de la comparaison envisagée. La langue commune, définie ici comme la partie banale du lexique de la langue à une époque contemporaine de la production des discours rassemblés dans le corpus, comporte implicitement une référence à un état de langue en synchronie.

En dehors du sentiment d'évidence qu'elle fait naître chez tout locuteur, il s'agit d'une réalité difficile à cerner, d'abord pour des raisons historiques : elle résulte de la transformation d'un dialecte dont l'aire de diffusion a été étendue au-delà de ses limites d'origine, et ce processus dynamique est en constante évolution

¹ Lafon P. (1980). On reprend ici les définitions données dans L. Lebart et A. Salem (1994 : 318) : « **spécificité positive** : pour un seuil de spécificité fixé, une forme *i* et une partie *j* données, la forme *i* est dite spécifique positive de la partie *j* (ou forme caractéristique de cette partie) si sa sous-fréquence est « anormalement élevée » dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou supérieures à la sous-fréquence constatée est inférieure au seuil fixé au départ.

Spécificité négative : pour un seuil de spécificité fixé, une forme *i* et une partie *j* données, la forme *i* est dite spécifique négative de la partie *j* si sa sous-fréquence est anormalement faible dans cette partie. De façon plus précise, si la somme des probabilités calculées à partir du modèle hypergéométrique pour les valeurs égales ou inférieures à la sous-fréquence constatée est inférieure au seuil fixé au départ. »

Pour une présentation simplifiée du modèle hypergéométrique, on renvoie à L. Lebart et A. Salem (1994 : 172-175).

² P. Lafon (1980 : 128).

jusqu'au stade ultime de la langue morte¹. En synchronie, la question est surtout envisagée sous l'angle des rapports entre langue commune et langue spécialisée. D'une manière générale, on peut dire qu'il n'existe pas de frontière nette entre les deux sortes de langue. D'une part, les langues spécialisées se développent à l'intérieur d'une langue commune : « (...) le français de l'automobile n'est en aucune façon une sorte de langue ou de dialecte, et l'usage de termes analogiques comme « technolecte » induit un parallélisme fallacieux. Le français de l'automobile est l'usage du français pour rendre compte de connaissances en matière d'automobile, ce qui est bien différent (...).»² D'autre part, la langue commune reçoit en permanence un enrichissement de son lexique en provenance des langues spécialisées³. Par ailleurs, les langues spécialisées connaissent elles-mêmes des gradations internes : B. Quemada invite à prendre en compte les domaines techniques et le niveau de technicité des termes⁴ ; M.T. Cabré partage ce point de vue en l'affinant⁵. Enfin, lorsqu'il s'agit d'établir un lexique, Ch. Muller rappelle qu'il y a deux manières d'aborder la question : l'une consiste à rechercher ce qui se trouve à l'intersection des idiolectes dans une communauté à une époque donnée, et cela peut donner le *Dictionnaire fondamental de la langue française* de G. Gougenheim, à l'opposé, on peut vouloir constituer un ensemble à partir de l'union de ces mêmes idiolectes, et l'on a l'expérience du *Trésor de langue française*⁶. Dans l'approche que l'on propose, ces deux options sont mises à contribution avec les réserves qui ont paru nécessaires.

Lorsque l'on retient l'approche par comparaison externe, il convient dans un premier temps de s'assurer que l'on dispose d'une valeur approchée de la norme, qui ne soit pas trop éloignée de la date de production des discours rassemblés dans le corpus à étudier. Par ailleurs, il faut également pouvoir compter sur un contraste suffisant entre les deux corpus, mais pas au point qu'il n'y ait de commun entre eux qu'un vocabulaire très pauvre (on pense aux mots grammaticaux) ; car en procédant

¹ J. Vendryès : [1923] 1968 : 287-307.

² P. Lerat (1995 : 18).

³ W. Von Wartburg et S. Ullmann : [1946] 1969 : 160-170.

⁴ B. Quemada (1978 : 1148-1150).

⁵ M.T. Cabré (1998 : 124-133).

⁶ Ch. Muller (1967 : 16).

de la sorte, on fait l'hypothèse que, les rapports d'activité traitant exclusivement des questions relatives à l'activité d'un groupe bancaire particulier, la confrontation avec le vocabulaire de textes dans lesquels on peut penser a priori qu'une telle thématique occupe une place très marginale, devrait contribuer à mettre en relief des mots communs aux deux ensembles, mais surreprésentés dans le corpus RA.

B. Une comparaison avec l'usage du *TLF* : attentes et réserves

Il est utile dans ce type d'entreprise de pouvoir s'appuyer sur des programmes informatiques permettant de s'affranchir de relevés manuels longs et fastidieux. C'est ce que rend possible *Hyperbase*, logiciel de textométrie conçu par Étienne Brunet¹. Ce logiciel offre, en effet, la possibilité de comparer le vocabulaire d'un corpus avec les fréquences de 100 000 formes distinctes rendant compte de l'usage observé dans le *Trésor de la langue française* et ayant pour source le corpus *Frantext*². Il est par ailleurs possible de choisir une période allant du XVI^e au XX^e siècle afin de cibler un état de la langue en rapport avec celui du corpus. On a retenu la période disponible la plus récente (1950-1990) pour la comparaison avec le corpus des rapports d'activité. Ce choix ouvre déjà la possibilité d'introduire un biais, puisque l'on se prive ainsi des néologismes susceptibles d'avoir enrichi la langue commune dans la dernière décennie du XX^e siècle et dans la première du XXI^e siècle³. Mais le concepteur du logiciel attire l'attention des linguistes sur d'autres points susceptibles de diminuer la valeur de la comparaison⁴.

En effet, le corpus du *TLF* reflète un usage littéraire de la langue et qui plus est, dans un registre soutenu. On sort donc sensiblement du cadre donné pour le corpus par un sous-genre de discours économique. D'autre part, les faibles fréquences ne sont pas prises en compte pour la comparaison, le calcul de l'écart

¹ On utilise ici la version 5.2 du logiciel datée de 2001.

² Pour une présentation à jour au site <http://www.atilf.fr/tlfi.htm> ; le site de la base de données *Frantext* ayant pour adresse : <http://www.frantext.fr>.

³ Une mise à jour pourrait être réalisée à l'aide d'un fichier de fréquences contenant les dernières données disponibles. L'exercice n'a pas été poussé jusque là car les résultats attendus pouvaient accepter un léger décalage chronologique, compte tenu du fait que la langue commune est théoriquement moins sujette aux variations brutales que les vocabulaires spécialisés.

⁴ E. Brunet (1999 : 54) *Manuel de référence de la version 4.0 d'Hyperbase*. Dans le cadre de cette étude, on utilise la version 5.2 du logiciel.

réduit perdant de sa pertinence lorsque la fréquence théorique d'une forme est trop faible. Par ailleurs, si une forme est absente de la liste de référence, quand bien même elle serait particulièrement caractéristique du corpus étudié, elle n'est pas relevée ; on a vu précédemment que les néologismes sont concernés comme les noms propres. Enfin, il faut noter que la comparaison porte sur des fréquences de formes et non de lemmes. Cela rend la comparaison possible puisque le corpus RA n'a pas été lemmatisé. En revanche, dans le cas des homographes, il n'est pas possible de trancher à partir de la liste de référence qui donne les formes sorties de leur contexte.

Une fois ces réserves faites, ce type de comparaison semble devoir être maintenu pour plusieurs raisons. Tout d'abord, on ne recherche pas ici des variations fines du lexique observables entre un auteur et ses pairs, mais, plus grossièrement, l'usage qui est fait d'une variété de langue particulière dans les rapports d'activité. De ce point de vue, le corpus du *TLF* garantit un usage littéraire contemporain qui, s'il ne résume pas à lui seul la langue commune, hérite probablement d'une part importante de ses caractéristiques dans ce qui peut l'opposer aux langues spécialisées, autre variété qui sera étudiée plus loin. En second lieu, on devine que cette comparaison risque d'être surdéterminée par les aspects thématiques qui opposent les deux corpus : là encore, ce point ne semble pas rédhibitoire. Au contraire, il s'agit bien d'explorer la manière dont la thématique bancaire des rapports d'activité est restituée dans ses aspects lexicaux et en particulier pour ce qui concerne les substantifs entendus au sens large. Enfin, comme cela a été indiqué plus haut, on rappelle qu'à ce stade de l'enquête, on recherche une caractérisation à valeur exemplaire et partant non exhaustive. Il reste que les résultats de la comparaison ne devront pas être pris tels quels mais soumis au préalable à une évaluation destinée à réduire les risques d'une mauvaise interprétation. En effet, le « réactif » utilisé pour faire apparaître des vocables relevant de la langue commune pourrait fort bien faire émerger également d'autres vocables que l'on devrait plutôt qualifier d'unités terminologiques.

C. Résultats et méthode d'évaluation des « mots-clés » en contexte

On donne ci-dessous la liste des cinquante premiers mots-clés issus de la comparaison. Comme il sera à nouveau question de mots-clés plus loin dans cette étude, on précise que le sens retenu ici est celui autrefois proposé par P. Guiraud de « *mots dont la fréquence s'écarte de la normale*¹. » On s'éloigne sur ce point du vocabulaire propre à Hyperbase afin de réserver l'expression « forme spécifique » pour le résultat d'un calcul de spécificité réalisé avec le logiciel Lexico et qui met en œuvre une autre méthode de calcul (hypergéométrique).

¹ P. Guiraud (1954 : 64-66). À cette époque P. Guiraud avait utilisé la liste Vander Beke comme référence censée refléter la norme dans l'usage de la langue (George Vander Beke, *French Word Book*, New York, The Mac Millan Company, 1931).

Tableau 26.

Cinquante premiers mots clés résultant de la comparaison entre les fréquences du TLF pour la période 1950-1990 et celles du corpus Rapports d'activité

Écart réduit	TLF	Corpus rapports	Mots clés	Écart réduit	TLF	Corpus rapports	Mots clés
2899.00	36	2882	agricole	196.38	44	217	international
1538.07	175	3375	Crédit	193.56	25	161	taux
660.86	57	828	milliards	191.26	104	326	produits
558.22	132	1066	Caisses	187.72	24	153	crédits
557.13	12	320	actifs	182.80	28	161	sociétés
444.13	48	511	locales	175.11	10	92	acquisition
410.55	204	977	banque	169.62	19	123	financière
326.94	243	851	total	168.15	56	210	épargne
324.03	10	170	filiale	168.06	95	274	réseau
313.10	59	400	bilan	165.01	29	148	banques
311.41	2	73	actionnaires	164.45	97	271	capital
302.98	142	602	fonds	160.83	716	733	Président
296.83	36	296	croissance	159.16	48	184	consommation
288.91	43	315	marchés	158.19	161	337	bureaux
276.78	83	420	développement	149.74	150	308	clientèle
272.06	14	169	Banco	146.47	23	117	financiers
260.76	14	162	Lyonnais	144.14	47	165	professionnels
256.75	296	740	Directeur	143.35	332	442	millions
224.96	87	350	entreprises	142.99	58	182	métiers
218.18	16	145	contrats	141.65	203	340	activité
214.16	22	167	commerciale	141.29	96	232	activités
212.95	645	914	Groupe	139.96	643	606	marché
210.58	48	243	progression	138.34	198	328	assurance
209.89	18	148	enregistré	137.59	705	625	francs
202.91	31	188	gamme	133.50	24	109	augmentation

Lecture du tableau 26 : le tableau a été présenté en deux blocs afin de regrouper les mots clés sur une même page. Les formes sont classées par ordre décroissant d'écart réduit (présent dans la première colonne de chaque tableau). La colonne TLF correspond à la fréquence relative de la forme enregistrée dans le TLF (base *Frantext*) et la colonne suivante présente la fréquence absolue observée dans le corpus de rapports d'activité. La quatrième colonne fournit pour chaque ligne la forme qui a été documentée par les chiffres des trois colonnes précédentes.

On notera tout d'abord qu'il était possible de prendre une liste plus longue. Cependant, en recherchant une plus grande complétude à ce niveau, on risquait d'obtenir un ensemble difficile à manipuler et d'autant plus inutile, que les principaux commentaires peuvent être réalisés à partir de ces cinquante premières formes. Comme on l'a mentionné plus haut, la liste ainsi obtenue doit faire l'objet

d'une évaluation, qu'il faut maintenant préciser. Compte tenu de la différence relevée entre les textes composant les deux corpus, celui du *TLF* et celui des rapports d'activité, il convient d'abord de se demander si les formes ainsi recensées correspondent au même mot dans les deux corpus (cas des homographes), ou bien à l'utilisation des mêmes zones de la même palette sémantique dans le cas des mots polysémiques.

Il fallait donc établir une règle d'évaluation permettant de limiter le risque d'erreur dans la caractérisation. Pour le *TLF* on tient que l'usage correspond *grosso modo* à celui qui est enregistré dans deux dictionnaires de langue, le *Nouveau Petit Robert (NPRE)* et le *Trésor de la Langue Française (TLFi)* dans leurs versions électroniques¹. On compare ensuite avec l'usage du même mot dans le corpus des rapports d'activité (corpus RA) : si l'on constate que la diversité des acceptions est attestée, on considère que le mot appartient à la langue commune ; si, au contraire, les occurrences du corpus RA indiquent qu'il s'agit d'un emploi contraint limité à une seule acception, ou bien si le mot est manifestement employé dans un sens qui n'est pas répertorié dans les deux dictionnaires, alors on tiendra qu'il relève dans ce corpus d'un usage spécialisé et que par conséquent sa sélection dans les mots-clés relevant de la langue commune est abusive. Restent enfin les cas litigieux pour lesquels on a, par exemple, un usage manifestement partagé entre langue commune et langue spécialisée. Le cas est assez fréquent et on a résolu cette difficulté en observant la répartition de l'usage entre les deux variétés de langue, un usage majoritaire s'élevant à une proportion d'au moins deux tiers des occurrences emportant finalement le rattachement à l'une ou l'autre des variétés.

Pour aborder ce point, il est également nécessaire d'établir si la forme recensée correspond à un mot simple ou à un syntagme complexe. La question ne se pose pas dans les mêmes termes pour les deux corpus. En effet, pour ce qui est du *TLF*, la liste de référence ne comporte que des unités simples. En revanche, la question se pose avec plus d'acuité dans le corpus RA compte tenu du fait que son vocabulaire est susceptible de contenir des unités terminologiques dont on sait

¹ *Version électronique du Nouveau Petit Robert (NPRE)*, Dictionnaires Le Robert / VUEF, version 2.1, Paris, 2001. *Trésor de la Langue Française informatisé (TLFi)*, CNRS éditions, Paris, 2004.

qu'elles se présentent majoritairement sous la forme d'unités complexes¹. De ce point de vue, cette question est liée à la précédente, car, comme on le verra, le problème ne s'est véritablement posé que pour les unités terminologiques et les noms propres.

Ce point permet d'introduire une dernière remarque relative à un traitement du programme *Hyperbase* lors d'une étape antérieure d'analyse du corpus RA (segmentation, puis constitution de l'index). Ce traitement, dans le souci louable de rassembler toutes les occurrences d'un même mot, a assimilé dans certains cas les occurrences des formes avec et sans majuscule initiale (ce dernier cas étant typiquement lié à un emplacement en début de phrase). Par ailleurs, si un mot apparaît toujours avec une majuscule initiale, quelle que soit sa place dans la phrase, il est rangé parmi les noms propres. Reste le cas difficile à arbitrer a priori des mots apparaissant tantôt avec, tantôt sans majuscule initiale, à tous les emplacements possibles de la phrase². Il semblerait, au vu des résultats fournis par l'index du corpus, que la loi du plus fréquent l'ait emporté. Cela explique, par exemple, que la forme « Crédit », qui occupe la deuxième position dans la liste, totalise 3375 occurrences ; celles-ci proviennent de l'addition abusive des occurrences des formes de « crédit » (383 occ.) et de « Crédit » (2992 occ.). En effet, la forme avec minuscule entre dans la composition de nombreuses unités terminologiques complexes, alors que la forme commençant par une majuscule entre dans un nombre non moins important de noms propres³.

En plus des concordances⁴, deux techniques ont principalement été utilisées avec le logiciel Lexico 3 afin de vérifier, dans les contextes d'apparition des formes, si celles-ci pouvaient ou non être rattachées à la langue commune. La première s'appuie sur les segments répétés⁵ du corpus et consiste à rechercher la chaîne de

¹ R. Kocourek (1991 : 135-151).

² On peut signaler également le cas des mots en capitales.

³ On peut voir à cette occasion que les traitements ne sont jamais neutres ; ce qui est paradoxal ici tient au fait que la règle principale, délibérément écrite en faveur des unités appartenant au lexique, favorise abusivement les noms propres dans une routine de deuxième plan.

⁴ « L'ensemble des lignes de contexte se rapportant à une même forme pôle ». L. Lebart et A. Salem (1994).

⁵ A. Salem (1987). Segment répété (SR) : « suite de formes dont la fréquence est égale ou supérieure à 2 dans le corpus. » L. Lebart et A. Salem (1994).

caractères correspondant au mot-clé dans l'ensemble des segments répétés du corpus RA classés par ordre décroissant de fréquence. Cette méthode permet d'explorer les environnements les plus contraints d'une forme. En commençant par cette étude, on dispose d'un moyen rapide et fiable pour estimer le comportement d'une forme donnée dans un corpus, y compris lorsque celle-ci possède un grand nombre d'occurrences. La seconde technique d'exploration repose sur le calcul des co-occurrences d'une forme ou d'un groupe de formes dans un contexte défini au préalable et qui peut, par exemple, correspondre à la phrase ou au paragraphe.

II. LANGUE COMMUNE ET VOCABULAIRE DE LA BANQUE

A. Documentation des « mots-clés » par les segments répétés et les co-occurrences

1. Le cas de « agricole »

Pour illustrer ce point, on appliquera cette méthode à la forme « agricole » dotée du plus fort écart réduit dans le tableau calculé par Hyperbase. On pourrait penser a priori que cet adjectif épïcène a bien sa place dans la thématique d'une banque dénommée « Crédit agricole », si ce n'était sa fréquence très élevée qui la place respectivement au quatorzième (« Agricole », 1714 occ.), au vingt et unième rang (« agricole », 1108 occ.) et plus accessoirement au quatre cent trente-cinquième rang (« AGRICOLE », 60 occ.), dans les fréquences les plus élevées de l'index du corpus RA. Dans le tableau ci-dessous, on donne les principaux segments correspondant à des dénominations propres ou à des syntagmes lexicalisés incluant l'une des graphies de « agricole ».

*Tableau 27.
Principaux segments répétés correspondant à une dénomination et incluant l'une
des formes du mot-clé « agricole »*

Mot-clé	Segments répétés - dénominations	Fréq.
agricole		
(formes <i>Agricole</i> , <i>agricole</i> et <i>AGRICOLE</i> réunies)	Crédit Agricole	1699
	Crédit agricole	1062
	Crédit Agricole Indosuez	593
	Crédit Agricole S A	410
	groupe Crédit Agricole	123
	Crédit Agricole Asset Management	96
	Crédit Agricole Indosuez Cheuvreux	82
	CREDIT AGRICOLE	60
	la Caisse régionale de Crédit agricole	59
	Caisse nationale de Crédit agricole	52
	groupe Crédit Agricole S A	43
	Président de la Caisse régionale de Crédit agricole	40
	Fédération nationale du Crédit agricole	25
	Caisses régionales de Crédit Agricole	24
	Crédit Agricole Lazard Financial Products	20
	Caisses régionales de Crédit agricole	18
	Groupe Crédit agricole	17
Fédération Nationale du Crédit Agricole	16	

Lecture du tableau 27 : la colonne de gauche reprend le mot-clé étudié et les formes qui, assimilées par le traitement du logiciel Hyperbase, ont été associées dans un groupe de formes dans le logiciel Lexico (le groupe correspond à l'expression régulière 'Agricole|agricole|AGRICOLE'). La colonne centrale présente les segments répétés calculés par Lexico correspondant à des dénominations incluant l'une des trois formes du mot-clé dans le corpus (lorsqu'un segment d'une longueur supérieure à un segment plus petit ayant la même fréquence est repéré, Lexico ne retient que le segment le plus long des deux. Cela explique, par exemple, que le segment « la Caisse régionale de Crédit agricole » ait été repris au lieu de la seule unité terminologique sans déterminant. La dernière colonne présente la fréquence observée dans le corpus pour chaque segment répété.

Il apparaît à la lecture de ce tableau que le mot-clé « agricole » dans ses différentes formes est très majoritairement utilisé à l'intérieur de dénominations propres (Crédit agricole) ou d'unités terminologiques complexes (Caisses régionales de Crédit agricole). On voit, en effet, que la forme « AGRICOLE » est absorbée à 100% par la dénomination propre « CREDIT AGRICOLE », le taux est de 99%

pour la forme « Agricole » et de 96% pour la forme « agricole » relativement moins contrainte. Celle-ci est, en effet, utilisée également dans d'autres contextes très minoritaires. Cependant, comme il s'agit ici de repérer un usage dominant, ces emplois marginaux de l'adjectif dans son acception ordinaire ne suffisent pas à maintenir le mot-clé dans la liste de ceux qui appartiennent à la langue commune.

Il reste qu'historiquement c'est bien en tant qu'adjectif que ce mot a été associé au terme « crédit » pour former l'unité « crédit agricole », en précisant par la destination qu'il s'agissait d'une sorte de crédit caractérisé par le financement des agriculteurs par d'autres agriculteurs organisés en sociétés coopératives, les caisses de crédit agricole mutuel. En outre, dans l'usage actuel de la presse, cette banque est très fréquemment appelée « banque verte », expression où le remplacement de l'adjectif « agricole » par « vert » montre que le sens, fut-il métaphorique, est encore présent à l'esprit des locuteurs contemporains. Cependant, l'inclusion dans une dénomination ne peut pas être considérée comme un fait anecdotique, d'autant qu'après plus d'un siècle d'existence, cette banque s'est banalisée et n'a plus grand-chose d'agricole, en dehors de certains pans de son organisation et d'une faible partie de sa clientèle. De ce point de vue, la dénomination fait figure de survivance du passé en complet décalage avec la réalité du groupe bancaire, et sa persistance, au-delà des tactiques du marketing, semble bien avoir été rendue possible par une « désémantisation » de l'adjectif « agricole¹ » dans la dénomination propre.

Cette première méthode est particulièrement adaptée, lorsque l'on pressent que le mot-clé a des chances d'être inclus dans des dénominations propres. En revanche, le repérage des usages s'avère plus difficile lorsque de telles unités utilisent peu ou partiellement le mot-clé, mais que l'on pressent par ailleurs que l'usage attesté dans le corpus ne reflète pas celui que l'on attendrait d'un mot relevant de la langue commune. Une deuxième méthode est alors utilisée pour procéder à la vérification souhaitée. Elle s'appuie sur l'exploration d'un contexte moins contraint que le segment répété, puisqu'elle consiste à repérer les co-

¹ On notera de ce point de vue que l'apparition de la majuscule dans les dénominations est venue renforcer cette désémantisation. En effet, on a à partir de 2000 « Crédit Agricole » et à partir de 2001 « Crédit Agricole s.a. ».

occurents de la forme situés à l'intérieur d'une fenêtre pouvant correspondre à la phrase ou au paragraphe. On montrera aussi qu'elle se combine heureusement avec la première approche.

2. Le cas de « *international* »

Afin d'illustrer ce point, on utilisera un autre mot-clé appartenant à la même catégorie grammaticale que le premier, du moins en apparence. Le mot-clé « international » regroupe, comme dans le cas précédent, les occurrences de trois formes : « international » (176 occ.), « International » (35 occ.) et « INTERNATIONAL » (6 occ.). Contrairement à ce qui a été fait pour l'exemple précédent, on étudiera séparément les variantes graphiques, la forme en minuscules étant séparée des formes avec majuscule initiale et capitales. La méthode utilisée précédemment permet de raccrocher rapidement et de façon exclusive les variantes avec majuscules à des dénominations diverses. Comme pour « agricole », les variantes utilisant la majuscule sont très majoritairement incluses dans des dénominations propres, ce qui ne correspond pas à un usage ordinaire de l'adjectif :

- « Project Finance **International** » (nom d'un magazine),
- « Cheuvreux de Virieu **International** Ltd » (nom d'une filiale basée à Londres),
- « **International** et grandes clientèles » ou « GRANDES CLIENTELES ET **INTERNATIONAL** » (nom d'un pôle d'activité au sein du groupe bancaire), etc.

La méthode des co-occurents est utilisée pour explorer les contextes dans lesquels apparaît la graphie en minuscules d'« international ». Les co-occurents sont calculés grâce à la méthode des spécificités qui isole, du fait de leur fréquence anormale¹, des formes apparaissant dans les mêmes fenêtres de texte que « international ». La fenêtre retenue est le paragraphe, car elle est censée fournir le meilleur rapport entre pertinence et volume d'informations captées, car outre les collocations sur l'axe syntagmatique, elle a l'avantage de proposer des formes ou

¹ C'est-à-dire plus ou moins élevée que la fréquence attendue par rapport à une norme qui cette fois fournit une estimation reposant sur la fréquence de la forme dans le corpus, la taille de ce dernier, la fréquence de la forme dans les fenêtres d'observation et la taille des fenêtres en question.

des segments répétés rattachés à la forme par des liens thématiques. On donne ci-dessous une sélection des principales formes et segments co-occurents apparaissant dans au moins l'un des cent soixante paragraphes où la forme « international » est présente :

Tableau 28.

Extrait des quarante premiers co-occurents formes ou segments répétés de la forme « international » dans le corpus RA

Formes et segments	Frq. Tot.	Fréquence	Coeff.
l international	98	98	***
à l international	73	73	***
international	176	176	***
développement international	17	17	24
en France et à l international	12	12	17
réseau international	9	9	13
Ratio international de solvabilité	8	8	12

Lecture du tableau 28 : la première colonne du tableau reprend les formes et les segments co-occurents de la forme « international » ; la colonne « Frq. Tot. » (fréquence totale), donne la fréquence de la forme ou du segment pour l'ensemble du corpus ; la colonne « Fréquence » donne la fréquence de la forme ou du segment dans l'ensemble des fenêtres sélectionnées par la présence concomitante de la forme « international » ; la colonne « Coeff. » donne le coefficient de spécificité de la forme ou du segment pour les mêmes fenêtres (l'ensemble des fenêtres constituant un sous-ensemble du corpus RA). Le coefficient de spécificité peut être positif ou négatif. Dans le premier cas, il indique que la forme et le segment sont particulièrement présents. Seuls sont ici retenus les principaux coefficients positifs (lorsque le coefficient est remplacé par « *** », c'est que sa valeur dépasse 50).

L'analyse de cet extrait des quarante premiers co-occurents de la forme « international » montre que contrairement, aux attentes¹, « international » fonctionne d'abord comme substantif dans le corpus RA (le segment « l'international » a une fréquence de 98 pour une fréquence de 176 affectée à la

¹ Le *NPRE* ne connaît le substantif « international » que comme ellipse de « championnats internationaux », comme manière de désigner un sportif (« un international » pour un sportif participant à des rencontres internationales) et comme nom propre, « l'Internationale » désignant l'organisation ouvrière fondée par Karl Marx et l'hymne révolutionnaire. Le *TLFi* ajoute les désignations anciennes des membres des brigades internationales durant la Guerre d'Espagne et des partisans de l'Internationale (attestation relevée par J. Dubois – 1962 : 325 – dans une lettre de Flaubert datée de 1871), ainsi qu'un synonyme rare de cosmopolite.

forme, soit une proportion d'utilisation de 56%); l'opposition reprise dans le segment « en France et à l'international » indique que cet adjectif substantivé sert à désigner une aire géographique destinée à opposer le « national » au « reste du Monde », pour reprendre le vocabulaire des économistes¹.

Les segments « développement international », « réseau international » et « Ratio international de solvabilité » montrent l'existence d'un *continuum* des emplois allant de l'usage courant à l'unité terminologique. Si l'on additionne les formes incluses dans des dénominations aux emplois de la forme en minuscules comme substantif (néologique ou spécialisé ?) ou comme membre d'une unité terminologique complexe, on obtient une proportion des usages « atypiques » représentant plus des deux tiers des occurrences. Cela conduit donc à considérer que les emplois de « international » dans le corpus RA s'écartent de l'usage en langue commune, compte tenu de la règle présentée en introduction.

La combinaison de ces deux techniques a permis d'opérer une distinction entre mots-clés correspondant à l'usage de la langue commune, et mots-clés relevant d'usages atypiques de ce point de vue car, selon les cas, il s'agit soit d'une spécialisation de la langue, soit de dénominations propres. On reprend dans la section suivante les éléments dont les deux groupes sont composés, en adoptant une organisation propre à mettre en valeur les particularités des uns et des autres.

B. Dénominations cachées, sens uniques et « mots-clés » relevant de la langue commune

Dans la catégorie des « dénominations cachées » regroupant unités terminologiques et dénominations propres, entrent les mots-clés suivants : « actifs, agricole, assurance, Banco, banque, bilan, bureaux, Caisses, capital, consommation, contrats, Crédit, épargne, financiers, fonds, international, locales, Lyonnais, total ».

¹ « Ensemble des comptes qui retracent les flux entre les unités institutionnelles résidentes et les unités non résidentes. (...) ». Une unité institutionnelle étant un « centre élémentaire et autonome de décision économique dans l'accomplissement de sa fonction principale et disposant en principe d'une comptabilité complète. » A. Silem et J.-M. Albertini (2006).

Le cas de « bureaux » est exemplaire pour cette catégorie : le sens usuel de ce mot évoque plutôt de nos jours une sorte de meuble ou le local type dédié aux tâches administratives. On serait donc tenté de penser qu'il occupe ainsi une place bien méritée dans un corpus de rapports d'activité de banque. Cependant, dans la mesure où, comme l'on a déjà vu, cette activité n'est retracée que de manière schématique, on peut s'étonner de cette intrusion du lexique sur le terrain concret des moyens matériels. En réalité, les « bureaux » des rapports d'activité sont presque exclusivement des « bureaux permanents » (218 occurrences sur 223). Ces derniers correspondent, dans la terminologie du Crédit agricole, à des « *bureaux auxiliaires, ouverts toute l'année, pendant au moins cinq jours ouvrables par semaine, quelle que soit la durée de l'ouverture journalière*¹. » En cela, ils se distinguent des bureaux périodiques (ouverts pendant moins de cinq jours ouvrables par semaine) et des bureaux saisonniers (ouverts seulement pendant certaines périodes de l'année). Dans la terminologie bancaire et, du reste, également dans la langue commune, on parle plus volontiers d' « agence » ou de « guichet ». Il s'agit donc d'une unité terminologique propre au cadre étudié et qui n'appartient pas à la langue commune. La mention des « bureaux » dans les rapports d'activité est destinée à fournir pour chaque caisse régionale de Crédit agricole le nombre d'agences dont elle dispose, ce qui est un indicateur de l'importance de son réseau commercial.

Une deuxième catégorie regroupe ce que l'on pourrait appeler les mots à sens unique. La forme correspond bien à un mot recensé par les dictionnaires de langue, mais l'usage dans le corpus RA est restreint à l'un des sens répertoriés, ce qui nous paraît relever d'un usage intermédiaire entre langue commune et langue spécialisée. On peut considérer qu'il s'agit plutôt ici d'une spécialisation liée au genre de discours. En effet, on retrouve des mots ou des acceptions plutôt en usage dans la littérature économique. Dans cette catégorie ont été rangés : « entreprises, filiale, gamme, Groupe, marché, marchés, professionnels, réseau ».

La forme « filiale » permettra d'illustrer le cas de figure ainsi décrit. La faible fréquence enregistrée dans le *TLF* et la thématique générale d'un corpus

¹ G. Decourselle (1987 : 84).

littéraire plaident en faveur d'un usage de l'adjectif au féminin (« *Qui émane d'un enfant à l'égard de ses parents* » selon le *NPRE*). L'adjectif est inexistant dans le corpus RA qui privilégie le substantif homographe avec le sens de « *société jouissant d'une personnalité juridique distincte (à la différence de la succursale) mais dirigée ou étroitement contrôlée par la société mère* », d'après la même source. La sélection de ce mot-clé par le calcul est donc biaisée par l'homographie existant entre les deux mots ; chaque corpus utilisant très probablement l'homographe dans un sens différent, l'usage commun est amputé de l'un de ses membres, ce qui a entraîné ce classement de la forme « filiale » hors langue commune¹.

Le troisième groupe rassemble les vingt-trois mots-clés (une petite moitié de l'ensemble étudié), qui relèvent à proprement parler de la langue commune². Il s'agit de : « acquisition, actionnaires, activité, activités, augmentation, banques, clientèle, commerciale, crédits, croissance, développement, Directeur, enregistré, financière, francs, métiers, milliards, millions, Président, progression, société, taux »³. On va maintenant chercher à organiser cet ensemble pour voir dans quelle mesure il permet de traiter différents aspects de l'activité bancaire ; enfin, on vérifiera l'enracinement de ce groupe de mots dans la langue commune en le confrontant à la nomenclature du *Dictionnaire fondamental* de G. Gougenheim.

¹ Une vérification réalisée début 2007 dans la base *Frantext* donne des chiffres différents pour la période 1950-1990, ceci étant lié à un état du corpus qui n'est plus le même que celui de 1997, date de constitution du fichier utilisé pour la comparaison. On note cependant que l'usage de la graphie « filiale » comme adjectif ou substantif est directement liée aux textes du corpus. Ainsi, pour la période allant de 1950 à 1990 avec un corpus de travail de 362 textes tous genres confondus excepté les traités et essais, il y a 40 occurrences de « filiale », dont 10 correspondent au substantif. Pour la même période, avec un corpus incluant les traités et essais (613 textes), on a 84 occurrences de « filiale », 52 pour le substantif et 32 pour l'adjectif.

² Ces vingt-trois mots totalisent près de 7000 occurrences dans le corpus RA. Il s'agit, évidemment, d'une partie seulement de ce que représentent en termes d'occurrences les unités relevant de la langue commune.

³ Dans *Le Poids des mots*, D. Mayaffre compare au *TLF* un corpus de discours de deux hommes politiques représentant le parti de l'Ordre dans l'entre-deux guerre (il s'agit de P.-É Flandin et de A. Tardieu). Parmi les spécificités positives, ressortent, entre autres, « milliards » (+15), « agriculture » (+13), « finances » (+13) et « millions » (+13), qui ne sont pas sans rappeler certaines unités spécifiques identifiées comme relevant de la langue commune dans le corpus RA. Ce lointain écho tendrait à étayer la remarque de l'auteur selon laquelle le discours de Droite puiserait son inspiration dans l'économie : « *Le discours de droite se caractérise par ses passés composés et ses datations qui enferment la réalité dans un passé maîtrisé ; il se caractérise par son souci comptable, et la précision de ses analyses chiffrées, qui apprivoisent une réalité présente.(...) Et ces caractéristiques transcendent à notre avis les époques.* » D. Mayaffre (2000 : 384).

C. Rôle de la langue commune dans les rapports d'activité et retour au « français fondamental »

1. Mots-clés et composants de l'activité bancaire

De manière à permettre par la suite une comparaison avec la langue spécialisée, on propose de regrouper maintenant cette vingtaine de mots relevant de la langue commune dans des ensembles pouvant correspondre aux principaux ingrédients de l'activité bancaire. En procédant de la sorte, on s'inscrit dans le prolongement de la démarche proposée par B. Quemada dans la présentation qu'il fait de la langue spécialisée : « *L'expérience née de la pratique quotidienne permet à chacun de distinguer grossièrement ce que peuvent être des langages techniques en regard d'une langue commune ou générale. À un meilleur niveau d'analyse, on peut tenir pour acquis, dans une culture donnée, que les moyens de nommer les finalités (produits et productions), les chaînes d'actions (opérations), les façons ou le savoir-faire (procédés), les moyens (matériaux et outils), les spécialistes (agents et producteurs), etc., constituent, pour chaque domaine, un ensemble caractéristique¹.* »

On propose à la suite de B. Quemada une répartition générale légèrement différente : agents, opérations, objets (regroupant produits, productions, matériaux et outils), ensembles complexes (regroupant ces trois ingrédients) et propriétés ou attributs (éléments secondaires qui pour exister ont besoin d'un support appartenant à l'un des quatre éléments premiers). On peut illustrer cette classification, dont la vertu principale est d'ordre heuristique, à l'aide de l'exemple suivant : pour un contrat (objet) passé entre deux parties (agents économiques) en vue d'un prêt (opération), les clauses du contrat, les droits et obligations des parties, le loyer de l'argent (taux d'intérêt) sont considérés comme des propriétés. En appliquant cette grille aux mots-clés relevant de la langue commune, on obtient la répartition suivante :

¹ B. Quemada (1978 : 1148).

*Tableau 29.
Répartition des mots-clés relevant de la langue commune en fonction des grandes
catégories de composants de l'activité bancaire*

Agents	Opérations	Objets	Propriétés
- actionnaires - banques - clientèle - métiers - société	- acquisition - activité - activités - crédits	- francs - produits	- augmentation - commerciale - croissance - développement - Directeur (titre) - enregistré - financière - milliards - millions - Président (titre) - progression - taux

Tout d'abord, à la lecture de cette grille, on est frappé par le caractère général de cette évocation de l'activité bancaire. Cependant, les ingrédients nécessaires à une présentation sommaire de l'activité sont bien présents : la clientèle, les métiers de la banque et les actionnaires pour ce qui est des agents ; les opérations, à part les crédits, ne sont pas détaillées, mais la présence d' « acquisition » permet d'évoquer une activité spécifique liée au développement externe de la banque, qui mène une politique d'achat d'entreprises du secteur bancaire et financier ; les objets sont limités au générique « produits » et à l'unité monétaire servant à quantifier les résultats obtenus ; c'est ce dernier aspect qui est surtout mis en valeur. En effet, les mots servant à qualifier l'évolution positive de l'activité et son dynamisme (augmentation, croissance, développement, progression, commerciale, financière, taux), et à quantifier les résultats obtenus (enregistré, milliards, millions) sont très bien représentés. Enfin, les aspects institutionnels sont restitués à l'aide des titres associés aux plus hautes fonctions exercées dans le groupe bancaire. Pour qui ne rechercherait dans la lecture des rapports d'activité qu'une indication de tendance générale (globalement, les choses vont-elles plutôt bien ou mal ?), ce vocabulaire de base semble suffisant et n'a pas été négligé par le locuteur collectif qui s'est exprimé à travers les rapports d'activité.

2. Mots-clés et français fondamental

Enfin, on peut tenter de vérifier l'enracinement de ce groupe de mots dans la langue commune en le confrontant à la nomenclature, déjà ancienne, du *Dictionnaire fondamental de la langue française*¹. On peut voir que la sélection des mots-clés appartient majoritairement à un fonds lexical de base établi il y a plus de quarante ans (seuls « actionnaires », « croissance » et « taux » en sont absents ; on notera que les deux derniers sont également des termes qui se sont banalisés).

Tableau 30.

Comparaison des mots-clés relevant de la langue commune avec la nomenclature du *Dictionnaire fondamental de la langue française*

Mots-clés « langue commune »	Français fondamental	Note
acquisition	acquérir	dérivé de...
actionnaires	-	absent
activité	activité	
activités	activités	
augmentation	augmentation	
banques	banque	
clientèle	clientèle	
commerciale	commercial	
crédits	crédit	
croissance	-	absent
développement	développer	dérivé de...
Directeur	directeur	
enregistré	enregistrer	
financière	financière	
francs	franc	
métiers	métiers	
milliards	milliard	
millions	millions	
Président	président	
produits	produit	
progression	progrès	même origine latine que progrès : <i>progredi</i> , au supin <i>progressum</i> , « aller en avant » ; en passant par le nom d'action correspondant à <i>progredi</i> : <i>progressio</i> , <i>progressionis</i> . ²
sociétés	société	
taux	-	absent

¹ G. Gougenheim (1961).

² D'après le *Dictionnaire historique de la langue française*, Le Robert, [1992] 1998.

Les trois mots absents du *Dictionnaire fondamental* mais isolés par le test comme relevant de la langue commune (« actionnaires », « croissance » et « taux »), semblent suggérer une évolution de la langue dans le sens d'une plus grande perméabilité aux termes relevant de la sphère économique. En tout état de cause, cette évolution de la norme à laquelle le corpus de rapports d'activité a été comparé vient conforter la validité d'une telle comparaison.

Dans ce chapitre, on a pu établir que la langue commune en usage dans les rapports d'activité ne se réduit pas aux outils grammaticaux nécessaires à l'expression. Il semble au contraire qu'un rôle important lui soit assigné, celui de fournir aux lecteurs une image de l'activité bancaire que l'on pourrait qualifier de premier niveau. Elle propose un tableau général de cette activité, mais suffisant pour indiquer des tendances et quantifier les résultats obtenus au cours d'une année.

Afin de mettre au jour cette composante du vocabulaire des rapports d'activité, on a été conduit à utiliser un double artifice : d'une part, on a eu recours à une valeur approchée de la langue commune, d'autre part, les traitements nécessaires à cette mise en exergue ont été réalisés par un programme informatique induisant des contraintes d'un autre ordre. On a donc été amené à ne pas considérer les résultats fournis en sortie par le logiciel comme étant directement exploitables. Au préalable, une série de vérifications en contexte ont été systématiquement menées. Les fonctionnalités d'un logiciel de textométrie ont alors été à nouveau sollicitées pour documenter l'usage des mots-clés dans le corpus RA. Cela a été réalisé à l'aide des segments répétés incluant les formes correspondant aux mots-clés, ainsi que des co-occurents apparaissant dans les mêmes paragraphes. Ces vérifications ont permis d'identifier des utilisations contraintes pour certains mots-clés : inclusion dans des dénominations propres ou dans des unités terminologiques complexes, ou bien la sélection récurrente d'un trait sémantique lorsque plusieurs sont disponibles en langue ; enfin, certaines sélections ont été jugées erronées, car reposant manifestement sur une homographie. À l'issue de ces vérifications en contexte, vingt-trois unités sur cinquante ont finalement été étiquetées comme relevant de la langue commune. La confrontation avec un vocabulaire de base a globalement confirmé cette sélection.

Cette première approche du vocabulaire des rapports d'activité a certes permis d'identifier et de caractériser le rôle qui semble être assigné à la langue commune, mais elle a également été l'occasion de repérer au moins deux autres variétés de langue (dans le deuxième cas, le terme de variété pouvant être jugé trop fort, on parlera plutôt d'usage sélectif) : d'une part, l'existence d'une langue plus spécialisée utilisant des unités terminologiques et des dénominations propres

spécifiques, d'autre part, la présence d'unités de la langue commune dont l'usage dans le corpus ne sélectionne systématiquement qu'un seul versant de leur aire sémantique lorsque plusieurs sont disponibles. Il est donc nécessaire de poursuivre les investigations, afin de mieux cerner le fonctionnement et le rôle de ces deux autres facettes du vocabulaire des rapports d'activité.

CHAPITRE 5

UN USAGE « GÉNÉRIQUE » DE LA TERMINOLOGIE BANCAIRE ET FINANCIÈRE

Après avoir tenté de caractériser l'usage qui est fait de la langue commune dans les rapports d'activité, il convient d'examiner les questions soulevées par la contribution des unités terminologiques à la restitution de certains aspects de l'activité bancaire. Dans cette perspective, trois points peuvent être abordés : comment les termes sont-ils utilisés pour évoquer l'activité économique d'un groupe bancaire dans ses aspects les plus généraux ? Les termes utilisés dans les rapports d'activité possèdent-ils des caractéristiques spécifiques susceptibles de les rendre plus accessibles pour le grand public ? Quelles conclusions tirer des observations fournies par les deux points précédents en ce qui concerne l'usage de la terminologie dans ce type de discours ?

Avant de répondre à ces interrogations, il faut au préalable aborder la question du statut des unités terminologiques rencontrées en dehors des textes technoscientifiques¹. D'un certain point de vue, on pourrait en effet considérer que, sorties de tels contextes, ces unités perdent leur statut de termes, pour devenir des quasi-termes ou des unités de la langue commune². Cette question est liée à celle du statut des termes dans la durée : leur présence dans des textes non spécialisés peut également révéler une banalisation en cours ou achevée, les langues spécialisées alimentant en permanence le fonds commun de la langue à l'aide d'unités nouvelles. Dans ce travail, on admettra avec M. T. Cabré « (...) *qu'une langue de spécialité n'est pas un ensemble structurellement monolithique, mais qu'elle présente des variantes en fonction des usages et des circonstances de communication. Selon cette perspective, nous prenons en compte deux paramètres :*

Le degré d'abstraction, qui dépend du sujet en tant que tel, de ses destinataires et des fonctions de communication que vise l'émetteur.

¹ R. Kocourek (1991 : 48).

² « *Les termes sont, ensuite [en plus d'être des syntagmes lexicaux], des unités lexicales dont le sens est défini par les spécialistes dans les textes de spécialité. Si c'est le cas, ce sens est en même temps une notion (un concept) de la spécialité. Au cas où la définition ne serait pas jugée satisfaisante, on peut employer l'étiquette quasi-terme ou unité lexicale spéciale ; mais ce jugement, n'est pas, en principe, du ressort linguistique.*

Soulignons surtout l'observation suivante : les termes étant des unités lexicales définies, ils ne représentent, potentiellement, que certaines acceptions de l'aire sémantique de l'unité lexicale, à savoir celles qui sont définies par des spécialistes dans les textes spécialisés, et non par les lexicographes dans les dictionnaires généraux. » R. Kocourek (1991 : 180).

Les intentions de communication, qui déterminent les types de textes différents selon les objectifs que l'on souhaite atteindre. Les ressources utilisées pour les différents types de textes coïncident avec celles qui s'emploient dans des communications non spécialisées pour réaliser les mêmes opérations¹. »

En adoptant ce point de vue, il devient alors possible d'envisager les rapports d'activité sous un double aspect : en tant que documents émanant d'un domaine d'activité particulier, ils sont susceptibles de réutiliser une partie de la terminologie bancaire et financière, et en tant que supports de communication destinés à un large public comportant nécessairement des néophytes, ils peuvent jouer le rôle de discours seconds, à la manière des discours de vulgarisation scientifique. Par ailleurs, on considèrera que l'entrée d'une unité terminologique dans un dictionnaire de langue peut être vue comme un indice de banalisation plutôt qu'un empiètement de la lexicographie sur le terrain de la terminographie.

Afin d'étudier l'usage qui est fait de la terminologie bancaire dans les rapports d'activité, on cherchera à identifier une norme de référence extérieure au corpus. On vérifiera alors les modalités de l'implantation de cette terminologie. De cette étude on tirera une caractérisation de l'usage de la terminologie bancaire dans les rapports d'activité et l'on proposera de répartir les termes identifiés selon la grille des composants de l'activité bancaire, de manière à disposer d'une base de comparaison avec les observations réalisées dans le précédent chapitre.

¹ M. T. Cabré (1998 : 125-126).

I. IMPLANTATION DE LA TERMINOLOGIE DU DOMAINE

A. Adoption d'une référence externe en matière de terminologie bancaire et financière

Pour caractériser l'usage de la terminologie bancaire et financière dans les rapports d'activité, on ne dispose pas de la facilité offerte pour la langue commune, c'est-à-dire l'existence d'une base textuelle considérable d'où peuvent être extraites les fréquences observées pour des dizaines de milliers de mots, et d'un programme informatique adapté pour effectuer une comparaison. Il a donc été nécessaire d'aborder la question de la caractérisation du vocabulaire d'une autre manière. Comme pour la langue commune, le but recherché n'est pas une collecte exhaustive de toutes les unités terminologiques que comportent les textes du corpus. En revanche, il a été nécessaire de s'appuyer sur une norme afin de délimiter ce que l'on souhaitait considérer comme relevant de la terminologie bancaire et financière. Comme pour la langue commune, on a admis implicitement l'existence d'une terminologie bancaire et financière susceptible de traverser les textes des rapports d'activité, selon le modèle adopté pour les parlers d'entreprise. En effet, il ne s'agit pas tant de connaître le jugement d'un expert quelconque du Crédit agricole sur le vocabulaire des rapports d'activité, que de voir dans quelle mesure la terminologie propre au domaine est réellement utilisée et selon quelles modalités.

Afin de procéder à cet examen, une première possibilité consisterait à faire contraster le vocabulaire de deux corpus à l'aide de la méthode des spécificités, dans le but d'obtenir les unités caractéristiques de l'un des deux ensembles de textes. Cette technique a notamment été employée dans le cadre de l'extraction automatisée des termes¹. Cependant, son utilisation n'a pas été retenue ici pour plusieurs raisons. Implicitement, cette méthode repose sur une opposition tranchée entre les vocabulaires de deux corpus que l'on caractérise a priori comme étant représentatifs, pour l'un, de la langue commune, et pour l'autre, d'une langue spécialisée. Or, on a

¹ P. Drouin (2004).

vu dans le chapitre précédent, avec l'utilisation du corpus *Frantext* et le calcul l'écart réduit, que le résultat de ce type d'opposition n'est ni tranché ni homogène et qu'il doit être soumis dans tous les cas à vérification. La caractérisation de la variété de langue doit donc plutôt être questionnée que postulée¹. Par ailleurs, pour aborder les questions relatives à la terminologie bancaire et financière, on a préféré partir d'une liste fermée de termes pour laquelle on dispose systématiquement de définitions homogènes. Ce choix est apparu d'autant plus nécessaire, que le constat de l'usage terminologique d'une unité devait précisément être fait dans un contexte linguistique difficile à rattacher à une variété de langue dominante.

Il est alors apparu que l'approche proposée par les enquêtes d'implantation terminologique pouvait être une alternative fructueuse. Certes, il n'existe pas d'unanimité sur les méthodes utilisées, comme en témoigne le recueil collectif intitulé *La mesure des mots – cinq études d'implantation terminologique*², mais on a retenu l'esprit qui a animé ces études, c'est-à-dire la confrontation d'une liste de termes définie a priori (il s'agissait de termes regroupés sous forme de listes et ayant fait l'objet d'une publication au *Journal officiel*), avec les usages réels constatés dans la littérature³. Cependant, il faut préciser également que l'optique retenue dans ce travail n'est pas celle de l'aménagement linguistique (qui englobe l'implantation terminologique), tel que le définit P. Auger : « *Aux fins de cette étude et de façon générale, nous considérons que l'aménagement terminologique est une facette particulière de l'aménagement linguistique et vise ultimement l'implantation sociale de terminologies officialisées par l'autorité légitime, ici l'État, qui s'appuie sur sa loi linguistique et, par voie de conséquence, sur une politique linguistique qui découle directement de la loi*⁴. » En effet, il ne s'agit pas ici de vérifier l'implantation des quelques deux cents termes officiels du domaine des finances recommandés par la Commission générale de terminologie et de néologie⁵. L'usage

¹ En revanche, on obtient plus sûrement de cette manière des contrastes thématiques.

² L. Depecker et G. Mamavi (1997) ; voir aussi le numéro 16 de *Terminologies nouvelles* (1997) consacré aux enquêtes terminologiques.

³ Certaines études n'ont porté que sur des interviews.

⁴ P. Auger (1999 : 18).

⁵ Fascicule « *Vocabulaire de l'économie et des finances – Enrichissement de la langue française – Termes et définitions publiés au Journal officiel* », 2006, 272 p., Premier ministre, Commission générale de terminologie et néologie, diffusé par la Délégation générale à la langue française et aux

que l'on a cherché à repérer est celui de la terminologie bancaire et financière dans son ensemble, sans se limiter aux termes qui ont fait l'objet d'une recommandation officielle.

Comme cela a été fait pour la langue commune dans le chapitre précédent, on a recherché une norme dont la date de publication ne soit pas trop éloignée de celle des textes du corpus. On se place donc dans une perspective synchronique, en considérant provisoirement que la variable temporelle (il s'agit d'une période de dix ans seulement), est relativement neutre pour déterminer le caractère plus ou moins terminologique d'un vocabulaire. D'autre part, on a recherché une source dont le public est relativement proche de celui des rapports d'activité. En effet, l'utilisation d'une terminologie uniquement destinée à certains spécialistes aurait probablement conduit à accentuer inutilement le contraste avec le vocabulaire du corpus.

B. Le *Lexique de la banque et des marchés financiers*

Le *Lexique de la banque et des marchés financiers* (LBMF dorénavant) a semblé correspondre à ces critères. Il s'agit de ce que l'on peut appeler un usuel¹ regroupant environ 2500 entrées appartenant au vocabulaire bancaire et financier. Il est destiné à un public assez large et proche de celui des rapports d'activité, la quatrième de couverture l'estimant utile pour « *les professionnels de la banque et des marchés financiers, les avocats, experts-comptables, chefs d'entreprise, étudiants, épargnants, investisseurs (...)* ». Cet ouvrage, dont on utilise la cinquième édition² datant de 2001, est l'œuvre d'un juriste spécialisé dans les domaines bancaires et financiers, ce qui permet également de préciser le point de vue adopté. Sa portée peut néanmoins être étendue au-delà du droit financier *stricto sensu* dans la mesure où les définitions qu'il propose ont, d'après l'auteur, « *souvent bénéficié du conseil de professionnels qui, à Paris, Bruxelles, Francfort ou New York pratiquent quotidiennement ces activités bancaires ou financières*³. »

langues de France. Ces données sont également accessibles à partir de la base de données CRITER (Corpus du réseau interministériel de terminologie) à l'adresse <http://www.criter.dglf.culture.gouv.fr>.

¹ « *Ouvrage de référence mis à la libre disposition du public dans une bibliothèque* » (TLFi).

² La première édition, intitulée « Banque et bourse » date de 1983.

³ Avertissement pour la 5^e édition.

Avant d'utiliser cet ouvrage comme « réactif » pour caractériser le vocabulaire du corpus, on a pu procéder à son évaluation à l'aide des mots-clés étudiés dans la première section. En effet, il est apparu utile à ce stade de dégager plus finement les limites prévisibles de l'ouvrage, compte tenu du point de vue qui a pu être adopté, au moins implicitement, par ses auteurs¹.

On constate, par exemple, que si un certain nombre d'organismes et d'institutions du système bancaire sont mentionnés dans le *LBMF*, il s'agit majoritairement d'organismes publics nationaux ou internationaux. Le « Crédit agricole mutuel » est mentionné, mais d'autres banques françaises ne le sont pas, comme le Crédit Lyonnais, la Société Générale ou BNP Paribas. L'organisation propre à chaque établissement n'est pas reprise dans le détail : les caisses locales et régionales, les bureaux permanents sont absents, mais la Caisse nationale de Crédit agricole est mentionnée. Tout se passe comme si un double découpage avait été opéré. Horizontalement d'abord, le *Code monétaire et financier* et les textes réglementaires européens et internationaux relatifs aux marchés financiers, donnent le périmètre général de l'ouvrage. Verticalement ensuite, il semble que les éléments génériques aient été privilégiés. Ce dernier point doit néanmoins être relativisé car on constate aussi l'existence de familles terminologiques constituées avec le même terme de tête pour des expansions différentes (par exemple, « obligation à lots », « obligation à prime », « obligation assimilable du Trésor », etc.). Cela peut s'expliquer par le fait que retenir l'un des membres d'une telle série contraint souvent le terminologue à faire état des autres membres pour rendre compte de certaines particularités.

Le cadre implicite du *Code monétaire et financier* a imposé une autre limitation portant sur la diversité des aspects économiques de l'activité bancaire. C'est ainsi que les domaines voisins de l'assurance (*contrats, assurance*), de la

¹ Il faut néanmoins reconnaître à B. Sousi-Roubi, d'avoir pris la mesure de cette difficulté : « *Le choix porte d'abord sur l'étendue et les limites des deux grands ensembles qui composent l'ouvrage : le domaine de la banque et celui de la bourse. Ce choix est souvent conventionnel, mais toujours difficile. Le choix porte sur le terme à intégrer dans les ensembles : faut-il inclure tel mot ? Faut-il exclure tel autre ? Le souci de logique et de rigueur est alors aussi permanent que le sentiment de doute et d'inquiétude. Les risques s'appellent arbitraire, illogisme et omissions.* » *LBMF*, Avertissement pour la première édition (1983).

comptabilité (*bilan*), la mention des agences ou des partenariats (*réseau*) et, plus largement, tout ce qui a trait à la clientèle et à la commercialisation (marché des *professionnels, produits, gamme, filiale, entreprises*), ne sont pas traités. Enfin, on mentionnera également la tendance à privilégier les termes pour lesquels une définition précise a pu être apportée, au détriment de notions plus vagues mais très utilisées, comme les « services *financiers* » ou le « crédit à la *consommation* ».

Les limites qui viennent d'être identifiées ne sont pas de nature à remettre en cause le choix du *LBMF* pour tester le vocabulaire des rapports d'activité. Au contraire, la proximité observée avec le *Code monétaire et financier* confirme plutôt la portée terminologique de l'ouvrage. De ce point de vue, celui-ci semble occuper une position intermédiaire, entre un ouvrage terminologique probablement plus complet mais moins maniable, comme le *Grand Dictionnaire Terminologique*¹, et des discours moyennement ou peu spécialisés, comme les rapports d'activité. Il constitue à ce titre un « réactif » adapté au projet d'évaluation que l'on conduit.

C. Usage de la terminologie bancaire et financière dans le corpus de rapports d'activité

1. Termes attestés et termes récurrents

Comme cela a été mentionné plus haut, cette évaluation repose sur la projection de la nomenclature du *LBMF* sur les textes du corpus, un peu à la manière des études d'implantation qui examinent la reprise des néologismes officiels dans des textes spécialisés². Cette projection ne repose pas sur une comparaison de listes (nomenclature du *LBMF* et termes du corpus), qui aurait nécessité au préalable un étiquetage du corpus, l'extraction automatique et la lemmatisation de candidats termes, puis une sélection manuelle des termes soumise au jugement extérieur d'un « spécialiste » du domaine. On a préféré, dans cette étude, concentrer les tâches de contrôle sur la vérification du fonctionnement des unités dans leurs contextes

¹ Grand dictionnaire terminologique de l'Office québécois de la langue française : <http://www.granddictionnaire.com>.

² L. Depecker *et alii* (1997).

d'origine. De toute manière, il s'agit encore d'un passage obligé pour les options qui privilégient l'extraction automatique¹.

On a donc vérifié dans le corpus la présence en tant que terme de chaque entrée du *LBMF*. Pour cela, on a eu recours aux expressions régulières de manière à rassembler les différentes graphies d'une même unité ; les variantes ont ensuite été contrôlées à l'aide des concordances, ou dans les contextes de la phrase ou du paragraphe lorsque le nombre d'occurrences le permettait (en général, moins d'une centaine d'occurrences). Lorsque pour une forme simple ce nombre dépassait la centaine, le contrôle s'est d'abord fait à partir des segments répétés incluant la forme recherchée (ou ses variantes graphiques), comme cela a été exposé dans le précédent chapitre. Le contrôle a consisté à vérifier que la graphie correspondait bien à un terme, et dans ce cas, si l'usage était conforme à la définition qui en est donnée dans le *LBMF*. Dans les cas où l'usage majoritaire constaté (soit les des deux tiers des occurrences au moins) était conforme à la définition, le terme a été retenu dans une liste spécifique, sinon il a été écarté. Pour les autres cas, comme un usage terminologique minoritaire ou la présence d'un homonyme, une deuxième liste a été constituée².

À l'issue de ce repérage, on relève un total de 296 unités terminologiques attestées sur les 2406 que compte la nomenclature du *LBMF*. Cet ensemble de près de trois cents unités répertoriées représente plus de 8800 occurrences dans le corpus. On donne en annexe la liste des 296 unités attestées. Il est alors possible de proposer une caractérisation générale des unités composant la liste ainsi obtenue.

Du point de vue de la fréquence, on note une amplitude très importante puisque les termes sont répartis entre les hapax et la fréquence de 1549 observée pour l'unité « Crédit agricole [mutuel] ». Globalement, les termes de faible fréquence sont plus nombreux que les termes de fréquence moyenne ou élevée. On compte ainsi 191 termes ayant une fréquence inférieure à 9 contre 105 ayant une fréquence supérieure ou égale à 9. De même, tous les cas de figure possibles sont rencontrés pour ce qui est de la répartition des termes dans les neuf parties du

¹ M.-C. L'Homme (2004) et P. Drouin (2002).

² Ces listes figurent dans les annexes.

corpus. Ils s'échelonnent du hapax présent dans une seule partie au vocable présent dans toutes les divisions du corpus.

Ces deux éléments combinés (fréquence et répartition) permettent d'isoler deux grandes populations sur un *continuum* : les termes récurrents sont ceux qui sont présents dans toutes les parties du corpus, ce qui suppose une fréquence minimum de 9, le corpus étant divisé en neuf parties. Cependant, cette répartition théorique d'une occurrence dans chaque partie du corpus n'est que très rarement attestée, y compris pour des vocables non identifiés comme étant des termes¹. La fréquence la plus basse observée pour un terme présent dans chaque partie du corpus étant de 20, on a retenu cette limite. De manière à obtenir une population significative de termes, on a ensuite assoupli le critère d'une présence systématique dans les neuf parties à une présence attestée dans plusieurs parties consécutives. Le groupe des termes récurrents se compose donc d'unités terminologiques dont la fréquence dans le corpus est supérieure ou égale à 20 et dont la présence est vérifiée dans plusieurs parties consécutives du corpus. Les termes occasionnels sont ceux qui ont une fréquence inférieure à 9 (moins d'une occurrence en moyenne par partie du corpus) et qui sont localisés dans une seule partie du corpus (forme originale) ou dans des parties non consécutives. Le premier ensemble se compose de 57 unités (14,5% des 296), représentant près de 8000 occurrences soit, à elles seules, pratiquement 90% des occurrences des unités terminologiques attestées. On donne ci-dessous la liste des 57 termes récurrents.

¹ Il y a neuf formes communes de fréquence 9 dans le corpus RA, dont huit éléments de dénominations propres et un nom commun, « expression ».

*Tableau 31.
Liste des cinquante-sept termes du LBMF récurrents dans le corpus RA*

<ul style="list-style-type: none"> - actif (s) - action - actionnaire - affacturage - arrangeur - banque - banquier - bourse - capital (social) - carte bancaire - change - CNCA (Caisse Nationale de Crédit Agricole) - compte - courtage - courtier [de banque] 	<ul style="list-style-type: none"> - crédit - Crédit aAgricole [mutuel] - crédit-bail - dépôt - dette - émission - encours [bancaires] - épargne - euro - FCP (fonds commun de placement) - fonds propres - futures - gestion alternative - holding - indice 	<ul style="list-style-type: none"> - intérêts - intermédiation [des ordres / financière] - introduction en bourse - investisseurs institutionnels - LBO (Leverage Buy Out) - leasing - mandat - marché financier - marché primaire - moyen de paiement - OPCVM (Organisme de placement collectif en valeurs mobilières) - organe central 	<ul style="list-style-type: none"> - PEA - place - placement - portefeuille - prêt - risque - Sicav - société de bourse - souscription - taux d'intérêt - terme - titre - titrisation - trading - trésorerie
--	---	--	---

La répartition entre termes simples et termes complexes se fait nettement en faveur des premiers parmi les termes récurrents. En effet, que ce soit pour la nomenclature du *LBMF* ou pour la liste des termes attestés, on constate une répartition similaire, de l'ordre de 30% d'unités simples contre 70% d'unités complexes¹. Ces proportions sont inversées dans la population des termes récurrents où les termes simples représentent 76% des unités attestées contre 24% pour les termes complexes. On peut donner une autre idée de cette sélection en l'illustrant à l'aide de la manière dont les champs terminologiques² constitués par un même terme de tête, ou régissant, sont sollicités dans les rapports d'activité. Pour cela, on partira d'un terme simple et l'on comparera le champ présent dans *LBMF* avec ceux des termes attestés et des termes récurrents dans le corpus RA. On illustrera ce point à l'aide des deux termes les plus fréquents et les plus récurrents dans le corpus.

2. Les termes « banque » et « crédit »

Le terme « **banque** » est défini dans le *LBMF*, comme une « *catégorie d'établissement de crédit qui, d'une façon générale, est habilité à recevoir du public* »

¹ Ces proportions ne sont pas très éloignées de celles que donne R. Kocourek à propos des terminologies en général, dans lesquelles les termes simples représentent de 10 à 20% des terminologies (1991 : 139).

² « *Les champs à base formelle (...) regroupent les unités lexicales liées par la répétition des unités significatives : des morphèmes et des mots. Ces champs sont, en plus, caractérisés par des rapports sémantiques dus au sens de ces morphèmes et de ces mots répétés.* » R. Kocourek (1991 : 188).

des fonds à vue ou à moins de deux ans de terme, et qui peut effectuer toutes les opérations de banque. (...) ».

*Tableau 32.
Le champ du terme « banque » dans le LBMF et le corpus RA*

Champ du LBMF	banque ; banque (mutualiste ou) coopérative ; banque à distance (BAD) ; banque centrale ; banque centrale de compensation ; Banque centrale européenne (BCE) ; banque centrale nationale (BCN) ; banque d'affaires ; banque de dépôts ; Banque de France ; banque de l'Etat ; banque d'émission ; banque des banques ; Banque des règlements internationaux (BRI) ; Banque européenne d'investissement (BEI) ; Banque européenne pour la reconstruction et le développement (BERD) ; Banque fédérale des banques populaires ; Banque française du commerce extérieur (BFCE) ; banque inscrite ; Banque internationale pour la reconstruction et le développement (BIRD) ; Banque mondiale ; banque mutualiste ou coopérative ; Banques populaires.
Champ attesté dans le corpus	banque ; banque (mutualiste ou) coopérative ; banque à distance (BAD) ; banque d'affaires ; Banque de France ; Banque européenne d'investissement (BEI) ; Banque mondiale ; BEI ; BERD (Banque européenne pour la reconstruction et le développement) ;
Champ des termes récurrents	banque

Le terme « **crédit** » est défini de la sorte dans le LBMF : *« Acte par lequel une personne agissant à titre onéreux, met ou promet de mettre des fonds à la disposition d'une autre personne ou prend, dans l'intérêt de celle-ci, un engagement par signature tel qu'un aval, un cautionnement, ou une garantie. »*

*Tableau 33.
Le champ du terme « crédit » dans le LBMF et dans le corpus RA*

Champ du LBMF	crédit ; crédit [à l'exportation] export ; crédit à court terme ; crédit à l'exportation ; crédit à l'importation ; crédit à long terme ; crédit à moyen terme ; Crédit adossé ; Crédit agricole mutuel ; crédit cartellaire ; crédit commercial ; crédit confirmé ; crédit consolidé ; crédit consortial ; Crédit coopératif ; Crédit d'acceptation ; crédit d'anticipation ; crédit de calendrier ; crédit de campagne ; crédit de courrier ; crédit de décaissement ; crédit de mobilisation de créances commerciales (CMCC) ; crédit de mobilisation de créances nées sur l'étranger ; crédit de relais ; crédit d'embouche ; Crédit d'équipement des petites et moyennes entreprises (CEPME) ; crédit d'escompte ; crédit différé ; crédit d'impôt ; crédit documentaire ; Crédit documentaire confirmé ; crédit documentaire irrévocable ; crédit documentaire notifié ; crédit documentaire provisionné ; crédit documentaire révocable ; crédit documentaire subsidiaire ; crédit documentaire transférable ; crédit en blanc ; crédit en participation ; crédit en pool ; crédit extra-bancaire ; crédit face à face ; crédit financier ; Crédit foncier de France ; crédit fournisseur ; Crédit global d'exploitation ; crédit hypothécaire ; crédit intercalaire ; crédit interentreprises ; crédit irrévocable ; Crédit local de France (CLF) ; Crédit maritime mutuel ; crédit mobilisable ; crédit mobilisé ; Crédit municipal (caisses de) ; Crédit mutuel ; Crédit mutuel populaire ; crédit national ; crédit non mobilisable ; Crédit par acceptation ; Crédit par aval ; crédit par caisse ; crédit par cautionnement ; crédit par signature ; crédit populaire ; crédit revolving ; crédit roll over ; crédit stand by ; crédit subsidiaire ; crédit-acheteur ; crédit-bail ; crédit-bail adossé ; crédit-bail immobilier ; crédit-bail mobilier
Champ attesté dans le corpus	crédit ; crédit [à l'exportation] export ; crédit à court terme ; crédit à long terme ; crédit à moyen terme ; Crédit aAgricole [mutuel] ; crédit documentaire ; crédit hypothécaire ; crédit revolving ; crédit-bail ; crédit-bail immobilier ; crédit-bail mobilier.
Champ des termes récurrents	crédit ; Crédit aAgricole [mutuel] ; crédit-bail.

L'examen de la manière dont les champs terminologiques de « banque » et de « crédit » sont utilisés dans les rapports d'activité confirme une nette préférence pour les unités terminologiques génériques, les termes spécifiques, plus précis et susceptibles d'aborder de manière détaillée l'activité bancaire, étant relativement peu sollicités. Or, la généricité des termes récurrents les rend plus aptes que les autres à pénétrer la langue commune. De ce point de vue, leur utilisation pourrait révéler le souci du locuteur de choisir de façon délibérée ces unités pour la plus grande facilité avec laquelle elles sont susceptibles d'être comprises par les lecteurs des rapports d'activité. On s'attachera donc à vérifier ce point.

II. BANALISATION DU VOCABULAIRE UTILISÉ POUR LA RESTITUTION DE L'ACTIVITÉ BANCAIRE

A. Le recours à de termes banalisés

Afin de vérifier si la sélection des termes génériques revient aussi à choisir des termes banalisés ou en cours de banalisation, on aura recours à deux dictionnaires de langue bénéficiant d'une large diffusion, y compris sur des supports électroniques. On fait donc l'hypothèse que l'entrée de ces termes dans le *Nouveau Petit Robert électronique* (2001), *NPRE*, et dans le *Trésor de la Langue Française informatisé*, *TLFi*, publié entre 1971 et 1994 (2004), traduit un changement de statut assimilable à une certaine banalisation du fait de leur rattachement au fonds commun de la langue. Pour cela, on vérifie dans la nomenclature et les définitions de ces deux ouvrages la présence des 55 termes attestés et récurrents dans le corpus RA¹. Cette entrée dans les dictionnaires de langue est parfois accompagnée d'une marque spéciale intégrée à la définition du mot et destinée à préciser l'origine de son sens ou son domaine d'usage. L'absence d'une telle marque a été reprise comme un indice de banalisation, et sa présence a été considérée comme une banalisation non encore achevée². Cependant, cet indice doit être manipulé avec précaution, dans la mesure où son emploi varie en fonction des conceptions des lexicographes. En effet, alors que le *TLFi* revendique un découpage extralinguistique permettant de rattacher un sens au domaine dont il est issu, le *NPRE* cherche avant tout à opérer une distinction entre l'usage courant et l'usage spécialisé, ce dernier étant marqué par la mention d'un domaine d'usage³. Cette différence d'approche se traduit

¹ On a retiré les deux termes qui correspondent également à des noms propres d'organismes et que les dictionnaires de langue excluent *a priori* (il s'agit de « Crédit agricole [mutuel] » et de « CNCA »).

² « La marque lexicographie partiellement un terme, qui représente un concept dans le monde terminographique, c'est-à-dire qu'elle le métamorphose en mot. On passe alors de l'axe du savoir sur le monde à l'axe du savoir sur le linguistique. Une forme est normale, autrement dit dépourvue de marque, ou opposée à différenciée, c'est-à-dire marquée. » J-C Boulanger et M-C L'Homme (1991 : 38).

³ « L'adjuvant le moins immédiatement contextuel est le **domaine**. Quand ce domaine est institutionnel, parce qu'il représente un corps de métier ou une branche cataloguée du savoir, nous l'avons détaché en l'imprimant en **capitales italiques** ; quand il n'a qu'un caractère occasionnel, et que donc il est plus **thématique** qu'institutionnel, nous disons ordinairement **dans le domaine de...**, imprimé en minuscules italiques. Le changement de domaine entraîne généralement un changement de sens par inclusion dans son sens d'éléments empruntés aux implications du domaine. Quand ce

probablement par une apparition moins fréquente des marques de domaines dans le *NPRE*, du moins pour ce qui est de la cinquantaine de termes retenus pour la vérification¹.

Cette investigation a d'abord permis de distinguer plusieurs groupes parmi les différents termes retenus. Une minorité d'entre eux est présente dans les deux dictionnaires de langues sans marque d'origine ou de contexte d'usage. Il s'agit de « crédit », « actionnaire », « placement », « change » et « banquier ». Deux autres sens sont présentés sans marque de domaine dans le *TLFi* mais avec une marque dans le *NPRE* : « compte » et « épargne ». Pour vingt-neuf autres cas recensés², le terme est traité par le *TLFi* à l'aide d'une marque de domaine alors que celle-ci est absente dans la proportion des deux tiers pour le *NPRE*. Huit autres termes sont absents du *TLFi* mais présents dans le *NPRE*, soit sans marque de domaine (« euro », « affacturage », « mandat », « FCP »), soit avec une telle marque (« intermédiation », « courtier », « société de bourse », « marché primaire »). Enfin, un quatrième groupe rassemble les termes absents des deux dictionnaires : « moyen de paiement », « arrangeur », « encours [bancaires] », « futures », « OPCVM », « organe central », « PEA », « gestion alternative », « trading » et « LBO ».

On a indiqué plus haut que la présence d'une marque de domaine doit être utilisée avec prudence comme indice de rattachement à une spécialité ou comme

changement est minime, il n'est pas marqué par une découpe du sens : on le considère comme une acception particulière en relation avec une variante par exemple technologique, qui ne touche pas à l'invariant du noyau sémique. Une division de sens peut ainsi embrasser plusieurs domaines, la définition fournissant une représentation générique du noyau sémique. » P. Imbs, Profil général du *TLF*, p. XXXII-XXXIII, tome 1.

Dans leur préface du *Nouveau Petit Robert*, J. Rey-Debove et A. Rey adoptent un point de vue différent : « Certains termes ne sont en usage que dans un domaine particulier de la communication ou dans un type de discours, et les abréviations qui correspondent à des domaines du savoir dans les encyclopédies sont employées ici pour caractériser un emploi didactique, technique ou scientifique du français et qui ne serait ni naturel ni normal dans l'usage général. Au contraire des mots et sens familiers, ces termes spéciaux sont le plus souvent partagés par toutes les communautés francophones et appartiennent à ce fonds commun que l'on peut qualifier de " français général " et qui rend possible la communication, malgré les divergences, dans toute la francophonie. » *NPR électronique*, 2001.

¹ Le tableau reprenant les termes et les marques associées aux sens dans les deux dictionnaires est donné en annexe.

² Il s'agit de « banque, action, bourse, émission, intérêts, dépôt, taux d'intérêt, souscription, trésorerie, fonds propres, introduction en bourse, courtage, prêt, actif, Sicav, titre, portefeuille, carte bancaire, titrisation, place, dette, indice, holding, risque, crédit-bail, leasing, terme, marché financier, investisseurs institutionnels ».

indice de banalisation. On se limitera donc à des remarques d'ensemble. Ainsi, quand elle concorde avec le *TLFi* (absence ou présence d'une marque), l'absence de marques d'usage spécialisé dans le *NPRE* permet de regrouper un ensemble de vingt-sept termes qui semblent être passés dans l'usage courant, soit la moitié de la liste. Cette proportion importante tend à confirmer que l'usage qui est fait du vocabulaire spécialisé dans les rapports d'activité privilégie des termes répandus et a priori accessibles pour « l'homme cultivé moderne (...) » du *TLFi*¹ ou la population nombreuse des « amis lecteurs » et de « (...) ceux qui le deviendront » du *NPRE*². Le groupe des dix-sept termes présents dans le *NPRE* avec une marque de domaine d'usage, rassemble une majorité de mots appartenant au champ lexical de la bourse et des marchés financiers (« titrisation », « place », « indice », « holding », « marché financier », « investisseurs institutionnels », « intermédiation [des ordres / financière] », « courtier », « société de bourse », « marché primaire »). La déréglementation des marchés, qui a encouragé la financiarisation de l'activité bancaire et de l'économie en général, est un phénomène contemporain que les médias ont contribué à rendre très visible. L'entrée de ces termes spécialisés dans un dictionnaire de langue enregistre une banalisation sur laquelle semblent également tabler les auteurs des rapports d'activité. L'absence de huit termes de la nomenclature du *TLFi* arrêtée plus tôt que celle du *NPRE* s'explique également par la banalisation récente de termes relatifs aux marchés financiers : « intermédiation », « courtier », « société de bourse », « marché primaire », « euro », « affacturage », « mandat », « FCP ».

Il reste cependant une proportion non négligeable de termes (une dizaine) qui échappent aux deux dictionnaires utilisés pour le test. On note dans cette population une surreprésentation de syntagmes terminologiques ou de sigles (« moyen de paiement », « encours [bancaire] », « OPCVM », « organe central », « PEA », « gestion alternative », « LBO »), ainsi que des termes d'origine anglo-saxonne (« arrangeur » – calque de « *arranger* » –, « *futures* », « *trading* », « *leverage buy out* » - développé du signe *LBO*). Certaines de ces unités feront probablement leur

¹ TLF – Profil général du dictionnaire TLF p. xvii, tome 1 (1971).

² *Préface du Nouveau Petit Robert*, par J. Rey-Debove et A. Rey (2001).

entrée dans les prochaines éditions des dictionnaires. On note, par exemple, que « moyen de paiement » est cité plusieurs fois dans le *NPRE* sans pour autant faire l'objet d'une définition. Le terme « OPCVM » n'est pas non plus recensé alors qu'il est le générique de deux termes qui le sont « Sicav » et « FCP ».

Il semble probable que, s'il advient un jour, le passage dans l'usage courant de ces dix termes aura à franchir un obstacle important : ils représentent majoritairement le point de vue du banquier et ont, de ce fait, moins de chances d'être partagés par tout autre agent économique. Les Sicav et les FCP peuvent être conseillés par un banquier à son client comme des placements possibles. Le terme d'« OPCVM » sert plutôt à regrouper ces mêmes placements par familles, ce classement étant lié à l'organisation de la collecte et de l'intermédiation boursière. Dans le même ordre d'idée, il est possible d'opposer la « carte bancaire », qui est d'un usage répandu, et son générique, les « moyens de paiement », qui permet de regrouper tout un pan de l'activité bancaire. L'organisation juridique de l'entreprise bancaire (organe central), la quantification de son activité (« encours bancaires »), ou certains pans d'activité éloignés du grand public (« gestion alternative », « LBO », « futures », « trading »), tous ces aspects peu visibles sont pourtant mentionnés. Si les rapports d'activité sont destinés à être lus, certes par des spécialistes, mais surtout par un assez large public, on pourrait s'attendre à ce que ces termes fassent l'objet d'un traitement discursif particulier¹.

À partir de ces constats relatifs à la terminologie bancaire et financière présente dans les rapports d'activité, il est possible de proposer une première caractérisation. Les unités terminologiques utilisées de façon récurrente sont en nombre relativement limité : une soixantaine. Elles sont majoritairement simples (les trois quarts), alors que la terminologie de référence utilisée pour la comparaison comporte des unités complexes dans une proportion voisine. Il apparaît également que les champs terminologiques ayant un même terme pour tête sont relativement peu sollicités dans les rapports d'activité, et deviennent quasiment inexistantes lorsque l'on examine les termes récurrents. Enfin, la comparaison avec les

¹ Ce point sera abordé dans le dernier chapitre.

nomenclatures de deux dictionnaires de langue montre que la moitié de ces termes peut être considérée comme banalisée ou en voie de l'être, alors que la présence d'une dizaine d'entre eux, absents des deux nomenclatures, devra recevoir une tentative d'explication. Ce dernier point n'est cependant pas de nature à mettre en cause le constat général selon lequel la terminologie bancaire et financière des rapports d'activité est caractérisée par le recours sélectif à des unités génériques pour le traitement ordinaire de l'activité bancaire. L'utilisation de termes plus spécifiques est également constatée, mais elle subit une double restriction d'emploi : elle n'est pas récurrente et, lorsqu'elle est utilisée, c'est avec parcimonie.

B. Termes récurrents et composants de l'activité bancaire

1. Agents économiques

Cette première caractérisation gagne à être complétée par une description plus fine qui examinera l'usage qui est fait de la terminologie bancaire et financière à l'intérieur des grandes catégories destinées à regrouper ce que l'on pourrait appeler les composants de base de l'activité bancaire¹ (agent, objet, opération, ensemble complexe) et les composants secondaires, ou « propriétés ». Cette analyse plus précise sera conduite à partir de l'ensemble des 296 termes attestés.

La population des agents économiques représente soixante unités terminologiques simples réparties à parts égales entre noms propres et termes simples et complexes. Les fréquences observées pour les trente noms propres attestés varient de 1 (AFB, BEI, CEPME, Commission bancaire, Conseil des bourses de valeurs, IMRO, SEC, Sicovam) à 1549 (Crédit [Aa]gricole) en passant par une fréquence de 9 (Banque de France) et une fréquence de 191 (Caisse Nationale de Crédit Agricole (CNCA)). On remarque que la fréquence des noms

¹ On précise à nouveau que ces catégories ont essentiellement une vertu heuristique afin de permettre de dégager des régularités dans l'usage des différentes variétés de langue mises à contribution pour restituer l'activité bancaire. Elle permet une meilleure manipulation des termes ; il est alors plus facile de voir comment opère la sélection propre au genre de discours. De la sorte on obtient une vue globale sur les caractéristiques de la sélection tout en maintenant la possibilité d'une réflexion au niveau des unités terminologiques prises séparément. Enfin, il y a un avantage pratique en plus de celui qui a été évoqué au-dessus : c'est la possibilité d'avoir une idée des lacunes prévisibles de l'extraction d'information à partir d'un tel corpus.

propres d'agents économiques semble liée à l'importance jouée par l'entité nommée qu'ils désignent dans le référentiel Crédit Agricole des rapports d'activité. Pour « Crédit agricole » et « CNCA » cela va sans dire. Mais on a, par exemple, pour « Coface¹ » une fréquence de 11 occurrences, qui est la plus élevée pour les entités situées en périphérie du référentiel Crédit Agricole. En effet, on voit que la fréquence tombe à une occurrence pour la mention d'une entité liée à un événement ponctuel, comme c'est le cas pour « Admical » : « *En juin 2000, la Fondation du Crédit Agricole " Pays de France " a reçu, pour l'ensemble de son action, l'Oscar du mécénat culturel attribué par l'Admical (Association pour le développement du mécénat industriel et commercial)².* »

Comme on l'a indiqué plus haut, l'autre moitié des agents économiques est représentée par des termes. Les fréquences de ces derniers varient de 1 (agent de change, clearing house, courtier en ligne, introducteur teneur de marché, maison de titres, SCPI) à 675 (banque), en passant par 48 (investisseur institutionnel) et 10 (banque coopérative). Dans cet ensemble, on remarque qu'un clivage existe entre termes au singulier et termes au pluriel.

En effet, les termes au pluriel semblent généralement utilisés pour désigner une classe, comme dans « établissements de crédit » (4 occurrences) : « *Pour l'ensemble des établissements de crédit, les encours de prêts à l'économie sont demeurés globalement stables³.* », ou dans « (...) *un fort développement commercial auprès des investisseurs institutionnels et des distributeurs locaux⁴.* », pour « investisseurs institutionnels ». Les termes utilisés au singulier sont fréquemment sollicités (dans les deux tiers des cas, et quelle que soit la fréquence) pour catégoriser le référent d'un nom propre. Pour reprendre le découpage proposé par J.D. Gollut et J. Zufferey⁵, le rattachement de l'entité nommée à une classe se fait par des opérations cognitives de définition utilisant les mécanismes de la prédication

¹ La Compagnie financière d'assurance pour le commerce extérieur (COFACE) est une participation importante du Groupe Crédit agricole.

² *Rapport d'activité de l'année 2000.*

³ *Rapport d'activité de l'année 1995.*

⁴ *Rapport d'activité de l'année 2002.*

⁵ Gollut J.D. et Zufferey J. (2000).

attributive (« Lukas Bank est une banque de détail » ou de l'extension attributive (« la banque russe Rossiyskiy Kredit »).

Parmi les termes servant à désigner des sortes d'agents économiques, ceux qui pourraient être utilisés également pour désigner des personnes physiques (agent de change, gérant de portefeuille, broker, dealer, courtier, banquier, actionnaire), désignent en général des collectifs, soit au travers des personnes morales qu'ils catégorisent (voir plus haut), soit par un recours systématique au pluriel. Le fonctionnement du terme « actionnaire » est, de ce point de vue, remarquable (actionnaires, actionnaires individuels). Le terme compte 40 occurrences au singulier et 69 occurrences au pluriel qui se répartissent systématiquement comme cela a été rapporté au-dessus. Ainsi, on a pour le singulier des utilisations du type « catégorisation d'une entité » : « *Avec près de 30 % des droits de vote, le Crédit agricole en est le premier actionnaire.* » (1995). Quant au pluriel, il sert généralement à désigner le collectif des actionnaires de Crédit Agricole S.A.: « *Depuis son introduction en bourse, Crédit Agricole S. A. compte plus de 2 millions d'actionnaires individuels en France et un grand nombre d'investisseurs institutionnels, principalement en France et en Europe.* » (2001).

2. Objets économiques

Les objets économiques représentent une population de 99 termes. À l'intérieur de celle-ci, les termes servant à désigner des réalités proches du grand public (chéquier, compte courant, livret A, carte de crédit, compte bancaire, distributeur automatique de billets, chèque, etc.) ont des fréquences faibles, ce qui surprend moins quand les réalités désignées sont moins visibles ou répandues (bons du Trésor, lettre de garantie, titre subordonné à durée indéterminée, obligation indexée). Ce point confirme les observations formulées à l'occasion de la caractérisation du genre de discours et de l'utilisation de la terminologie en général : l'usage constaté dans le vocabulaire des rapports d'activité est celui d'une terminologie privilégiant les aspects les généraux de l'activité, quel que soit le degré de visibilité de la réalité évoquée.

Cette sélection favorise donc des termes généraux comme les « moyens de paiement », les « OPCVM », ou génériques, au sens de régissant de nombreuses

expansions (Sicav, compte, titre, carte, action, etc.). Parmi les termes génériques, quand ils concernent des réalités proches de celles qui sont susceptibles d'être connues du grand public, comme « carte », on observe, comme pour les agents, une répartition différente des usages selon le singulier et le pluriel. Là également, le pluriel est utilisé pour désigner un ensemble (« *Le Crédit agricole compte 9,9 millions de cartes bancaires, soit 25 % du marché.* » (1998)), et le singulier est préféré pour catégoriser le référent d'une dénomination commerciale qui n'est pas nécessairement partagée. Pour reprendre l'exemple des cartes bancaires, la construction la plus fréquente (et probablement aussi la plus économique), consiste à placer le nom du produit en position d'épithète du terme générique, comme dans « *Enfin, le Crédit agricole affirme l'originalité de son offre en associant les avantages de la carte Mozaïc aux Livrets Jeunes récemment mis en place par les pouvoirs publics.* »(1995), ou dans « *Le Cedicam a créé une carte " Business " destinée aux petites entreprises pour gérer les dépenses professionnelles des dirigeants, et développé une offre de cash management domestique et international.* » (2002).

3. Opérations économiques

Les opérations économiques totalisent 72 unités terminologiques. Comme dans les cas précédents, on constate que les termes ayant peu d'occurrences sont, soit concrets (facilité de caisse, prélèvement automatique, virement, prêt personnel, découvert), soit techniques et peu accessibles au grand public (adjudication du Trésor, Offre publique de retrait, swap de taux, rachat d'actions, opération de crédit). Là aussi, le générique (crédit) est préféré au spécifique (crédit à moyen terme, crédit à long terme, etc.). En ce qui concerne ce dernier point on renvoie au champ terminologique donné plus haut.

En revanche, pour les opérations, la restitution des aspects généraux de l'activité bancaire emprunte une voie spécifique. Celle-ci consiste, par métonymie, à parler du domaine d'activité en utilisant le nom plus ou moins concret d'un type d'opération bancaire ou d'activité économique. Il arrive que le niveau de généralité exprimé par le terme soit moindre. Dans ce cas, il est question de l'ensemble des opérations ainsi désignées sans que celles-ci fassent l'objet d'une présentation

détaillée. À aucun moment il n'est question de l'activité elle-même, décrite sous ses différents aspects. Pour un terme relativement peu fréquent, « conservation de titres » (13 occurrences), on a des contextes de ce type : « *Dans le domaine de la conservation de titres, les trois centrales de titres du Crédit agricole, Brunoy, Mer et Sud, représentent une conservation de 1 100 milliards de francs.* » (1997), ou bien encore : « *Rassemblant les activités européennes du Groupe en matière de banque dépositaire et de conservation de titres, d'administration de fonds et de services aux émetteurs, Crédit Agricole Investor (...).* » (2003).

Lorsqu'une opération bénéficie d'une stabilité et d'une visibilité plus importantes dans le référentiel Crédit Agricole, tel qu'il est restitué dans les rapports d'activité, la transformation de l'activité concrète en domaine d'activité entraîne diverses modifications comme, par exemple, l'utilisation de la majuscule pour certaines occurrences d'un même terme. Ce phénomène est lié, entre autres, à la présence simultanée du terme dans les CDS et les CDN, où il est appelé à jouer des rôles différents. Par exemple, on a pour l'affacturage¹ les types de contextes CDS suivants, où le terme apparaît en général en minuscules :

« *Dans l'affacturage, Transfact a conclu un accord (...)* » (2000).

« *(...) dans des métiers de l'affacturage (...)* » (2001).

En revanche, pour les CDN, on a des configurations différentes dans lesquelles le terme apparaît toujours avec une majuscule initiale, comme dans les exemples suivants :

« *TRANSFACT (Affacturage)* » (2000)

« *Affacturage : n°1 en France via Eurofactor et Transfact* » (2003).

¹ La définition du terme est la suivante : « *Technique par laquelle un client appelé adhérent ou fournisseur, transmet ses créances à un factor (en général un établissement de crédit) qui moyennant rémunération, se charge de leur recouvrement et les lui paie soit à l'échéance (affacturage à l'échéance ou maturity factoring), soit avant l'échéance (affacturage traditionnel ou old line factoring) ce qui constitue alors une opération de crédit. (...)* » in LBMF, article « Affacturage ». Le terme compte dans le corpus RA 72 occurrences, réparties en 59 occurrences en minuscules et 12 occurrences avec majuscule initiale et une occurrence en capitales.

4. Ensembles complexes et propriétés

Les ensembles complexes, avec 16 occurrences, forment la catégorie de termes la moins bien représentée. On a accordé à ces ensembles un statut particulier parce qu'ils regroupent de fait les trois composants de base déjà présentés. C'est, par exemple, le cas de « marché » qui sert à regrouper à la fois les opérations des agents économiques d'un secteur, les agents eux-mêmes et les objets qu'ils manipulent. Il s'agit par définition d'un terme général, mais renvoyant à un ensemble hétérogène ou composé de classes différentes. Les ensembles complexes héritent pour l'essentiel des propriétés de leurs composants.

Dernier ensemble, les propriétés comptent 49 termes. Elles sont définies par défaut comme ce qui ne peut pas exister si l'un des quatre objets de base décrits plus haut n'existe pas. On fait l'hypothèse que le recours à des termes désignant des propriétés de composants peut être considéré comme un indicateur du niveau de détail utilisé dans la restitution de l'activité. Une présence plutôt faible des propriétés permettrait d'indiquer également la préférence pour les éléments quantifiables au détriment de ceux qui seraient susceptibles de détailler le fonctionnement de l'activité bancaire (pour reprendre l'exemple donné plus haut, les contrats sont plus aisément quantifiables et présentent à ce titre plus d'intérêt que les clauses qu'ils comportent, dès lors que l'on cherche à rendre compte globalement d'un pan de l'activité bancaire). C'est, par rapport aux trois ensembles principaux, (agent, objet et opération économiques), celui qui est le moins bien représenté. Trois termes ont une fréquence supérieure à 20 dans cet ensemble : « taux » (26), « risque » (145) et « terme » (134), entrant tous dans des syntagmes lexicalisés de fréquence moindre.

Comme pour les ensembles complexes, la caractérisation de l'usage des termes relevant de la catégorie des propriétés n'apporte pas d'éléments nouveaux. Le général et le générique y sont privilégiés par rapport au détaillé et au spécifique. En revanche, dans la mesure où le recours à ces termes peut fonctionner globalement comme un indicateur du niveau de généralité adopté dans la restitution de l'activité bancaire, on peut comparer la répartition des différentes catégories selon qu'elles

sont présentes dans la nomenclature du *LBMF*, ou qu'elles regroupent des termes attestés ou récurrents. C'est ce que présente le graphique donné ci-dessous :

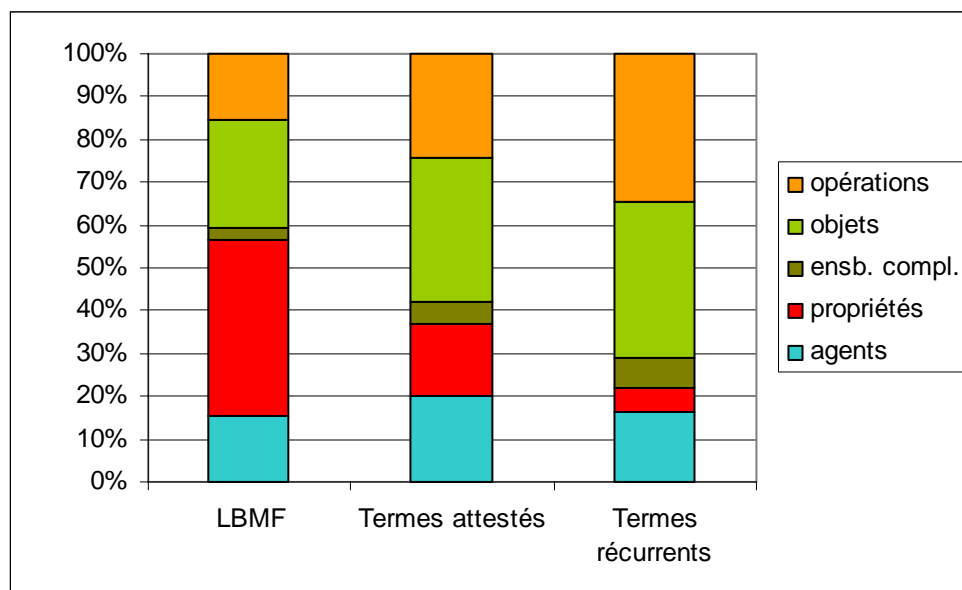


Figure 34.
Répartition des termes en grandes catégories selon leur contexte

Lecture de la figure 34 : cet histogramme reprend les grandes catégories entre lesquelles les termes ont été répartis (opérations, objets, ensembles complexes, propriétés et agents) en fonction de leurs provenances. Celles-ci sont au nombre de trois : la nomenclature du *LBMF*, les termes de cette nomenclature attestés dans le corpus RA, parmi ces termes attestés, ceux qui ont en outre la propriété d'être récurrents. Le nombre de termes (il s'agit des vedettes ou des vocables et non des fréquences) présents dans chaque catégorie a été ramené à une échelle allant de 0 à 100% pour la comparaison des proportions.

Il ressort du graphique présenté ci-dessus que les « propriétés », qui représentent plus de 40% de l'ensemble de la nomenclature du *LBMF*, descendent à une proportion de 5% dans la liste des termes récurrents attestés dans le corpus RA. Cette sélection va dans le sens d'un traitement relativement général de l'activité bancaire que l'on constate dans les rapports d'activité (sur 300 termes attestés, 1/5 seulement sont fréquents et récurrents, les champs terminologiques de type régissant-expansion sont peu sollicités ; enfin, les termes simples sont privilégiés par rapport aux termes complexes). Ces indications vont permettre d'éclairer la collecte d'information qui sera réalisée dans la troisième partie. En effet, elle rendra possible une évaluation du voisinage des noms propres selon qu'il sera conforme à ces

observations ou qu'il s'en écartera ; dans ce dernier cas, elle autorisera une caractérisation en termes d'apports nouveaux ou de lacunes.

Cette étude de la terminologie utilisée dans les rapports d'activité a permis de mettre en lumière les principales caractéristiques de leur sélection. Celle-ci privilégie les unités simples, générales et génériques, ces dernières étant pour certaines en usage également dans la langue commune. Les éléments de base de l'activité bancaire (agents, objets et opérations) y sont mieux représentés que les propriétés. La faible présence de ces dernières fonctionne comme un test, car si cette classe d'unités avait été plus importante, il est fort probable qu'elle aurait indiqué l'existence d'un traitement détaillé de l'activité bancaire. Par ailleurs, l'usage discursif va généralement dans le sens du général et du collectif : ce sont, par exemple, les termes pouvant désigner des agents physiques, qui renvoient en fait à des collectifs (personnes morales), ou encore, les noms d'opérations bancaires qui servent à désigner dans ce contexte des secteurs d'activité. Ce mouvement vers le général s'exprime aussi dans l'utilisation des termes, lorsqu'ils sont associés à certains noms propres afin de rattacher leur référent particulier à une catégorie générale (ce point sera repris dans la troisième partie). On a vu que la langue commune est utilisée afin de produire un discours généraliste sur l'activité bancaire. On vient de voir que cette dernière peut également être décrite dans ses grandes lignes à l'aide de termes banalisés. On peut se demander maintenant si les mots dont le sens est contraint de façon récurrente sont sollicités de la même manière dans les rapports d'activité. Cet examen sera l'occasion de préciser les caractéristiques de la phraséologie des rapports d'activité.

CHAPITRE 6

ACTIVITÉ BANCAIRE ET PHRASÉOLOGIE DES RAPPORTS D'ACTIVITÉ

L'usage et le rôle de la langue commune et de la terminologie bancaire ayant été précisés, un certain nombre de points relatifs à la sélection lexicale commandée par le genre restent à aborder. En effet, on a laissé en chemin, dans les précédents chapitres, un certain nombre d'interrogations à ce sujet. Ainsi, il est apparu que des mots relevant de la langue commune étaient présents dans les rapports d'activité, mais avec une restriction de leur palette sémantique. De même a-t-on pu remarquer, lors de l'évaluation menée à partir de la terminologie du *LBMF*, que la thématique de l'activité bancaire exercée dans une économie de marché était bien présente dans le corpus, mais absente d'une nomenclature à dominante juridique. Cela suggère l'existence d'une autre facette du vocabulaire qu'il s'agit maintenant de caractériser. Enfin, lorsque l'on a évoqué les propriétés du « sous-genre » de discours économique représenté par les rapports d'activité, il a été question de filtrages thématiques et stylistiques de divers ordres ; dans ce contexte, comment les variétés de langue utilisées dans le corpus sont-elles articulées pour restituer l'activité bancaire ? Pour tenter de répondre à ces questions, on partira du socle lexical du corpus, son vocabulaire commun, dont on étudiera les vocables ayant les plus fortes fréquences, à la lumière des variétés de langue et des principaux composants de l'activité bancaire. La phraséologie des rapports d'activité sera ensuite explorée à partir des syntagmes lexicalisés ayant pour tête des vocables communs. Enfin, l'étude de la ventilation de ces vocables permettra de mettre au jour des phénomènes de variation sémantique caractéristiques de la phraséologie des rapports d'activité¹.

¹ La notion de phraséologie telle qu'on l'entend ici est proche du premier sens donné par le *NPRE* : « Ensemble des expressions (terminologie et particularités syntaxiques) propres à un usage, un milieu, une époque, un écrivain. » On limitera l'étude des particularités syntaxiques à l'existence de locutions que l'on a déjà qualifiées de syntagmes lexicalisés. De ce point de vue, la conception adoptée ici se rapproche également, toutes proportions gardées (on recherche des phénomènes plutôt caractéristiques d'un lexique de situation et non de la langue en général), de celle proposée dans le *Dictionnaire de linguistique et des sciences du langage* (Dubois et alii, 1994) : « (...) Un dictionnaire phraséologique s'assigne pour objet le recensement et la présentation des expressions figées spécifiques à une langue. (...) Il devra également exclure le cliché de son champ d'étude : à la différence du cliché, écart stylistique banalisé par la répétition (*l'astre des nuits, des doigts de fée*), la phraséologie se définit non par l'écart qu'elle représente par rapport à la langue, mais par le caractère stabilisé de la combinaison qu'elle constitue. »

I. VOCABULAIRE COMMUN ET PHRASÉOLOGIE DES RAPPORTS D'ACTIVITÉ

A. Le vocabulaire commun thématique

Pour tenter de répondre à ces questions, on peut aborder le vocabulaire des rapports d'activité par un nouvel angle d'attaque, puisqu'il s'agit maintenant de caractériser une phraséologie propre, et non plus la manière dont des variétés de langue, identifiées à l'aide de moyens externes, sont mises à contribution. Une approche possible consiste à partir du vocabulaire commun à l'ensemble des parties du corpus. Le vocabulaire commun correspond à « *l'ensemble des formes attestées dans chacune des parties du corpus*¹ ». Lorsque ces formes sont classées par ordre décroissant de fréquence, ce sont les mots grammaticaux qui arrivent en tête ; on a ainsi « de, et, la, des, en, d, à, le » qui totalisent à eux seuls 38 611 occurrences soit plus du quart des occurrences des 731 formes correspondant au vocabulaire commun². Dans la mesure où il ne s'agit pas de faire un commentaire exhaustif du vocabulaire commun, on opérera sur celui-ci deux sélections. La première consiste à éliminer le vocabulaire qui ne correspond pas à des substantifs ou à des adjectifs qualificatifs. Cette sélection drastique s'explique par le but poursuivi : caractériser la toile de fond lexicale au milieu de laquelle des entités sont nommées, catégorisées ou rattachées à des ensembles plus vastes. Cette approche n'est pas très éloignée de celle des mots-thèmes de P. Guiraud³. Partant des mots qui ont la plus forte valeur absolue dans un corpus donné, cet auteur n'avait conservé que les substantifs, les verbes, les adjectifs qualificatifs et les adverbes, en accord avec les exploitations stylistiques qu'il comptait mener dans un deuxième temps. Ce vocabulaire thématique est ensuite restreint aux cinquante plus fortes fréquences, ce qui rend les données plus facilement manipulables. On parlera donc maintenant d'un extrait du

¹ Lebart et Salem (1994).

² Ce vocabulaire est donné en annexe. Il comporte 133 674 occurrences sur les 201 902 du corpus pris dans son ensemble, soit les deux tiers des occurrences.

³ P. Guiraud (1953).

vocabulaire commun, le vocabulaire commun thématique. On donne ci-dessous ces cinquante premières formes, par ordre décroissant de fréquence :

*Tableau 34.
Cinquante premières formes du vocabulaire commun thématique du corpus RA*

Crédit (2992)	marché (588)	année (406)	entreprises(316)	capital (235)
Agricole (1713)	fonds (567)	Jean (399)	services (315)	place (234)
agricole (1108)	France (542)	bilan (393)	actifs (304)	filiales (233)
Caisses (1047)	banque (528)	développement(388)	groupe (301)	progression (230)
milliards (828)	Général (485)	crédit (382)	marchés (296)	activités (229)
Indosuez (827)	locales (484)	encours (348)	clientèle (295)	bancaire (227)
Président (724)	gestion (471)	activité (339)	croissance (288)	Caisse (222)
Directeur (724)	propres (452)	clients (327)	assurance (280)	épargne (210)
Total (688)	millions (442)	offre (326)	réseau (267)	actions (205)
Groupe (603)	Banque (435)	produits (323)	Europe (236)	ensemble (204)

En prenant pour point de départ le vocabulaire commun thématique du corpus RA (corpus des rapports d'activité), on s'appuie sur ce que le vocabulaire du corpus a de plus banal¹. Dans le cadre d'une série textuelle, il est également possible d'identifier, au sein du vocabulaire commun, un vocabulaire de base : « *ensemble des formes du corpus ne présentant, pour un seuil fixé, aucune spécificité (négative ou positive) dans aucune des parties, (i.e. l'ensemble des formes qui sont « banales » pour chacune des parties du corpus)*². » Ces formes banales sont donc non seulement réparties sur l'ensemble des parties du corpus, mais en outre, leur répartition se fait dans des proportions similaires, quelle que soit la partie. Parmi les cinquante formes communes présentées au-dessus, on compte neuf formes banales (activités, assurance, bancaire, croissance, ensemble, marché, offre, progression,

¹ Cette banalité revêt un aspect décevant, lorsque la recherche vise à commenter le contraste entre le vocabulaire de corpus représentatifs de discours d'opinions opposées ou divergentes. Dans son étude sur le *Vocabulaire des proclamations électorales de 1881, 1885 et 1889*, A. Prost, qui est l'un des rares historiens à s'attarder sur les caractéristiques du vocabulaire commun d'un corpus, consacre un développement à ce sujet. Il en ressort que les mots outils du français fondamental, le vocabulaire technique de l'élection et les thèmes également développés par les principaux courants politiques semblent de peu d'intérêt comparés aux vocables marquant les discours conservateurs ou républicains (A. Prost : 1974, p. 35). Dans la perspective adoptée pour cette étude, le vocabulaire commun revêt au contraire un intérêt particulier, dans la mesure où il est caractéristique d'un même locuteur collectif, et où il permet ainsi d'accéder aux unités formant la base de la thématique des rapports d'activité.

² L. Lebart et A. Salem (1994).

réseau). Dans les développements qui vont suivre, les propos seront illustrés à l'aide de formes communes fréquentes, banales ou non.

Les mots-clés de la première section ont permis d'introduire la question de la caractérisation de l'usage des variétés de langue dans le corpus ; on peut maintenant revenir sur ces constats initiaux en utilisant l'éclairage fourni par une sélection reposant cette fois sur la fréquence et la récurrence de certaines unités. Trente-cinq des cinquante formes dont se compose la sélection du vocabulaire commun thématique sont des mots-clés appartenant à la liste présentée dans la première section¹.

Plusieurs facteurs permettent d'expliquer ce recoupement. Dans les deux cas, les mots grammaticaux ont été éliminés, respectivement par calcul et manuellement, du fait de leur relative neutralité thématique. En second lieu, la comparaison du chapitre 5 portait sur un ensemble de formes communes aux deux corpus (*Frantext* et *Rapports d'activité*), correspondant à une sorte de plus petit commun dénominateur thématique, compte tenu du fait que les formes trop spécifiques du corpus RA avaient peu de chances de se trouver dans l'autre corpus, à quelques exceptions erronées près, reposant sur une homographie. Cela est particulièrement frappant pour la forme « Banco ». Celle-ci correspond majoritairement à l'interjection « Banco ! » ou à une ellipse de l'expression « faire banco » dans la base *Frantext*², alors que dans le corpus RA elle est systématiquement employée dans des dénominations complexes de banques de pays lusophones ou

¹ Il s'agit de : « actifs, actions, activité, activités, Agricole, agricole, assurance, banque, bilan, Caisse, Caisses, capital, clientèle, Crédit, crédit, croissance, développement, Directeur, encours, entreprises, épargne, filiales, fonds, Groupe, groupe, locales, marché, marchés, milliards, millions, Président, produits, progression, réseau, Total ».

² La vérification faite en 2007 doit tenir compte de ce que le fichier utilisé pour la comparaison ne porte que sur la période 1950-1990 avec l'état du corpus en 1997, qui donne une fréquence de 14. La forme compte dix occurrences de plus dix ans plus tard, sans que l'on puisse identifier avec certitude les sens et les sortes de mots qui ont été comptabilisés (le corpus de travail *Frantext* comprend tous les genres disponibles, soit 613 textes). Cette réserve faite, l'observation des contextes présente une réalité plus complexe. Le sens de l'interjection et de l'enjeu tenu représentent 19 occurrences, les cinq autres occurrences se répartissant entre l'adjectif « banco » (1 occ.), anciennement associé à une monnaie pour signifier que sa valeur est indépendante des variations de change, un nom de personnage de *Macbeth* de Shakespeare donné avec une orthographe erronée (la bonne orthographe est « Banquo »), dans la première édition de *Frêle bruit* (1976) de Michel Leiris, deux noms de banques italiennes et un parc récréatif de Côte d'Ivoire...

hispanophones : « Banco Espirito Santo » (55 occ.), « Banco Bisel » (39 occ.), « Banco del Desarrollo » (23 occ.), « Banco Acac » (18 occ.), « Banco Boavista » (10 occ.), « Banco Bradesco » (7 occ.), etc.

Le troisième facteur, et non le moindre, repose sur ce que l'on pourrait appeler le jeu de la langue. On sait qu'un même mot peut être utilisé dans des situations de communication très différentes, qui ne sélectionnent qu'une partie de son aire sémantique. En effet, les mots utilisés dans les langues spécialisées sont fréquemment des mots, soit déviés de leur sens vulgaire du fait de leur utilisation dans un contexte scientifique ou technique, soit associés avec d'autres mots simples pour former des unités complexes spécifiques¹. Dès lors, il n'est pas surprenant que ces mots soient communs aux deux corpus lorsqu'il s'agit de mots simples représentant des notions générales. Ce que mettent en lumière la comparaison avec le *TLF* et la liste des fréquences élevées du vocabulaire commun thématique, c'est le recours privilégié à certains de ces mots dans le corpus RA, soit en les réutilisant avec leur sens vulgaire (environ 50% des cinquante premiers mots-clés), soit en déviant ce sens (termes simples et complexes, mots au statut incertain dont l'usage dans le corpus limite l'aire sémantique – phénomène de réfringence évoqué par A. Phal²). Les quinze formes du vocabulaire commun thématique qui ne figurent pas dans la liste des cinquante premiers mots-clés³ comportent deux nouveautés : ce que l'on pourrait appeler un mot outil, « ensemble » (utilisé très majoritairement comme substantif et non comme adverbe – 204 occurrences sur 205), et une division temporelle, « année » (attendue dans un corpus de rapports décrivant l'activité d'une entreprise durant une année civile). Les autres formes correspondent à des catégories déjà identifiées dans les précédentes sections.

¹ A. Phal (1969).

² *Ibid.*

³ Il s'agit de « année, bancaire, Banque, clients, ensemble, Europe, France, Général, gestion, Indosuez, Jean, offre, place, propres, services ».

B. Vocabulaire commun thématique, variétés de langue et composants de l'activité bancaire

Dans le but d'en proposer une restitution organisée, on peut répartir le vocabulaire commun thématique de deux manières : selon le recours à certaines variétés de langue et selon les principales catégories adoptées dans les chapitres 4 et 5, enrichies des dimensions spatio-temporelles et d'une catégorie générale « divers ». Ainsi, on a pour la langue commune les mots suivants : « activité, activités, année, bancaire, clientèle, clients, croissance, développement, Directeur (pour « Directeur Général », ou « Président Directeur Général »), ensemble, Général, milliards, millions, offre (de services), place (première place), Président, produits, progression, services ». On voit que les mots généraux dominant dans cet ensemble et qu'une thématique commerciale généraliste s'en dégage.

Pour ce qui est de la terminologie bancaire et financière, les mots suivants sont utilisés seuls ou au sein d'unités complexes : « actifs, actions, banque, capital, Crédit, crédit, encours, épargne, fonds (pour « fonds propres »), gestion, propres ». D'une manière générale, c'est le versant économique de l'aire sémantique de ces mots qui est sélectionné, et plus particulièrement les aspects bancaires. Il s'agit, bien sûr, d'une tendance qui n'exclut pas certaines exceptions. Par exemple, si les « actions » sont en général des valeurs mobilières (dans 90% des cas environ), elles peuvent également concerner « les actions » du groupe bancaire en matière de mécénat ou de développement durable.

Un troisième groupe est constitué de noms propres ou de parties de dénominations propres : « Agricole, agricole (Crédit Agricole), Banque (a un statut mixte, tantôt présente dans une dénomination propre complexe de banque ou de département à l'intérieur d'une banque, tantôt utilisée comme nom de secteur d'activité), Caisse (dans des dénominations du type « Caisse régionale Nord-de-France »), Europe, France, Indosuez (Banque Indosuez, puis Crédit Agricole Indosuez), Jean (ce prénom n'est jamais utilisé seul et entre dans des noms de ce type : « Jean Laurent (32 occ.), Jean Pierre Pargade (13 occ.), « Jean Bouysset (12 occ.), etc. ». Cette catégorie rassemble principalement des agents personnes morales ou physiques, ce dernier point constituant une entorse relative au principe de généralité déjà évoqué.

Un quatrième groupe rassemble des usages restreints du fait de la terminologie propre au Crédit Agricole et qui est absente du *Lexique de la banque et des marchés financiers* utilisé comme référence dans le chapitre 6, ainsi que des mots dont l'usage dans le corpus sélectionne une facette particulière de leur aire sémantique. Quoique non répertoriés comme termes, ces mots partagent avec les représentants patentés de la terminologie bancaire la propriété d'être utilisés pour le versant économique de leur aire sémantique. Il s'agit de : « assurance (contrat d'assurance et métonymies à partir de cette facette), bilan (les trois-quarts des 394 occurrences sont absorbées par l'expression « Total bilan », servant à annoncer les chiffres des bilans des entités du groupe bancaire des les contextes non syntactiques), Caisses (« Caisses locales » (372 occ.), « Caisses régionales » (653 occ.)) – « *Établissement où l'on dépose des fonds pour les faire valoir ou les administrer* » d'après le *NPRE*, entreprises (seule l'unité économique est concernée), filiales (sortes d'entreprises), Groupe, groupe¹ (dans les deux cas il s'agit très majoritairement des groupes « Crédit agricole » et « Crédit Agricole S.A. », la majuscule marquant graphiquement une individuation par catachrèse), locales (voir « Caisses » au-dessus), marché, marchés (il s'agit ici des quatre principaux sens de ce mot ayant un rapport avec les domaines bancaire et financier : lieu d'échange de valeurs, ensemble des échanges considérés à un moment donné, secteur économique et segment de clientèle ou débouché), réseau (d'agences bancaires), Total (voir « bilan » au-dessus) ».

¹ On notera cependant que « filiale » et « groupe » ont reçu une définition précise dans le *Code monétaire et financier* : « Art. L. 511-20. - I. - Est une filiale d'un établissement de crédit, d'une entreprise d'investissement, d'une compagnie financière ou d'une compagnie financière holding mixte l'entreprise sur laquelle est exercé un contrôle exclusif au sens de l'article L. 233-16 du code de commerce, ou une influence dominante en raison de l'existence de liens de solidarité importants et durables résultant d'engagements financiers, de dirigeants ou de services communs. (...)

III. - Est un groupe l'ensemble d'entreprises composé d'une entreprise mère, de ses filiales et des entités dans lesquelles l'entreprise mère ou ses filiales détiennent des participations, ainsi que des entités liées de telle sorte que leurs organes d'administration, de direction ou de surveillance sont composés en majorité des mêmes personnes ou qu'elles sont placées sous une direction unique en vertu d'un contrat ou de clauses statutaires. Les établissements affiliés à un réseau et l'organe central au sens de l'article L. 511-31 sont considérés comme faisant partie d'un même groupe pour l'application du présent code. Il en est de même pour les entités appartenant à des groupes coopératifs régis par les dispositions similaires dans la législation qui leur est applicable. »

Comme on vient de le voir, la proportion de mots dont l'usage relève de la langue commune dans les principales fréquences du vocabulaire commun thématique est proche de celle observée dans le cadre du test par les mots-clés, soit une vingtaine dans les deux cas. La répartition en grandes catégories susceptibles de rendre compte de l'activité bancaire offre un éclairage différent, qui permet en outre une comparaison avec la répartition obtenue pour la terminologie bancaire :

*Tableau 35.
Répartition des cinquante premières fréquences du vocabulaire commun thématique selon les grandes catégories de composants de l'activité bancaire*

agents	objets	opérations	ensembles complexes	propriétés	divers
Agricole agricole banque Banque Caisse Caisses clientèle clients Crédit entreprises filiales Groupe groupe Indosuez Jean locales	actifs actions bilan capital encours fonds offre produits propres Total	activité activités assurance crédit épargne gestion services	marché marchés réseau	bancaire croissance développement Directeur Général milliards millions place Président progression	année ensemble Europe France

Afin de l'éclairer, on peut rapprocher cette répartition de celle observée dans le chapitre 5 à propos des termes récurrents présents dans le corpus.

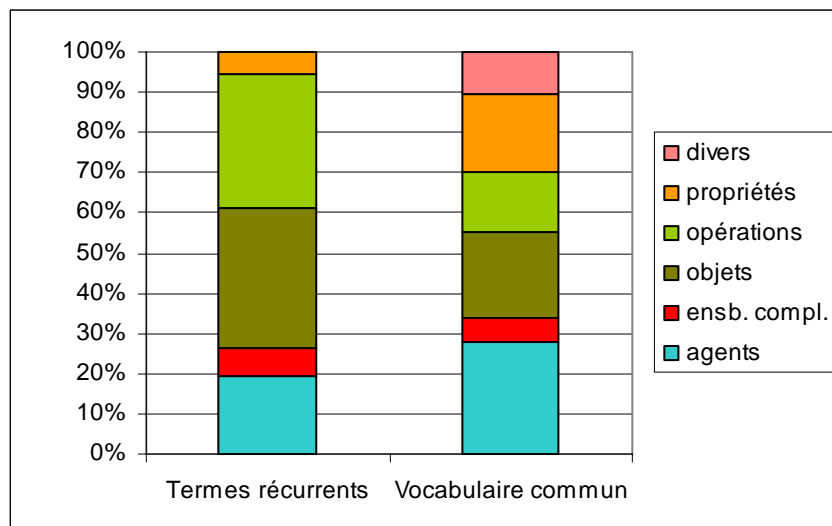


Figure 35.
Répartition des termes récurrents et des principales fréquences du vocabulaire commun thématique¹ selon les composants de l'activité bancaire

La comparaison entre les répartitions des termes récurrents et les mots du vocabulaire commun thématique suggère plusieurs remarques. On note tout d'abord que dans le corpus RA, lorsqu'il s'agit de traiter d'éléments qui se trouvent au cœur de l'activité bancaire (opérations et objets), les termes semblent privilégiés. Parmi les cinquante-sept termes récurrents recensés dans le chapitre 5, plus d'une vingtaine ont été qualifiés de récurrents du fait de leur fréquence et de leur présence dans des parties contiguës, mais non sur l'ensemble du corpus. Par rapport au vocabulaire commun thématique, le vocabulaire spécialisé semble donc également plus proche de l'actualité de l'activité bancaire telle qu'elle est restituée dans les textes du corpus, avec des mises en exergue temporaires ou des phénomènes de variation brutale auxquels le vocabulaire commun est globalement moins sensible (même s'il

¹ Pour les agents économiques les graphies différentes de « Agricole », « banque » et « groupe » ont été comptées pour une forme au lieu de deux.

existe des exceptions, comme l'apparition de la majuscule au « a » d'« agricole » à partir de 2000, qui entraîne d'importantes modifications de fréquence)¹.

À la différence des termes récurrents, le vocabulaire commun thématique possède une répartition plus équilibrée entre les différentes catégories. On note l'apparition de mots classés dans la catégorie « divers »; ces mots ont majoritairement pour fonction d'ancrer le discours sur l'activité bancaire dans un système de référence spatio-temporel. On note également la présence d'un mot outil applicable aux mots des autres catégories. On retrouve ce caractère de généralité chez les mots rangés dans la catégorie des propriétés. À la différence de ce que l'on avait observé dans le chapitre 5, où la rareté des termes relevant des propriétés fonctionnait assez bien comme un indice de la minoration des détails, les mots du vocabulaire commun relevant de cette catégorie ont un sens général bien adapté dès lors qu'il s'agit de caractériser l'activité de secteurs d'activité, voire du groupe bancaire pris dans son ensemble. On a ainsi une nouvelle confirmation de la prédilection du vocabulaire des rapports d'activité pour la généralité.

Enfin, on note que parmi les agents économiques, les noms propres ou les parties de dénominations complexes représentent la moitié des mots et permettent l'introduction d'un niveau de détail élevé avec les noms des principaux dirigeants du groupe bancaire. Si l'on rapproche ce constat de celui de l'apparition d'une nouvelle catégorie, où sont regroupés des mots destinés à fixer son cadre spatio-temporel au discours, on voit que les noms propres jouent un rôle non négligeable dans le vocabulaire commun thématique, dont ils représentent un cinquième des effectifs. Cette importance accordée aux acteurs, personnes morales ou physiques, et au cadre spatio-temporel dans lequel leur activité se déploie réintègre le référentiel propre au groupe bancaire dans le discours, à l'aide d'unités échappant aux variétés de langue. Afin de parfaire cette caractérisation du vocabulaire des rapports d'activité, il est maintenant nécessaire de s'attacher à la phraséologie qui leur est propre.

¹ Cependant, dans ce cas, la variation ne concerne plus directement l'interaction entre usage du vocabulaire et restitution de l'activité, mais se situe plutôt au niveau des représentations de cette activité.

C. La phraséologie des rapports d'activité : une abondance d'économismes

Pour aborder cette question, on a choisi d'explorer la phraséologie des rapports d'activité à l'aide d'unités déjà identifiées, appartenant soit aux mots-clés du chapitre 5, soit au vocabulaire commun. En suivant l'orientation rappelée plus haut, on considèrera que la phraséologie des rapports d'activité peut être abordée à travers les syntagmes lexicalisés qui la composent. Pour repérer ces syntagmes, on part des mots simples relevant de la langue commune, correspondant aussi à des termes banalisés, ou encore dont le sens a été restreint du fait de leur apparition dans les textes du corpus. En effet, si l'on tient compte des conclusions des précédents chapitres, ces mots sont très probablement utilisés comme régissants ou expansions dans un certain nombre de syntagmes. On a retenu quatre ensembles de trois mots chacun. Pour la langue commune on a retenu « activité(s) », « produits » et « clientèle ». La terminologie bancaire et financière est représentée par « banque », « crédit » et « épargne » ; trois mots représentent la catégorie des unités dont une partie seulement de leur aire sémantique a été sélectionnée de façon récurrente dans le corpus (« assurance », « entreprises », « marché(s) »). Enfin, trois mots ont été sélectionnés comme échantillon de contrôle du fait de leur appartenance au vocabulaire commun non retenu par le test des mots-clés (« gestion », « services », « bancaire(s) »). Au préalable, il est nécessaire d'explicitier les critères qui ont permis la sélection de syntagmes probablement lexicalisés.

En effet, il faut commencer par préciser que la notion de syntagme lexicalisé est relative, ou plus précisément qu'elle renvoie à un *continuum* d'unités complexes plus ou moins figées. Dès lors, l'identification de tels syntagmes doit se faire à l'aide de critères susceptibles d'établir leur degré de figement. G. Gross en propose un certain nombre :

« (...) nous définissons comme composés des groupes nominaux:

- dont aucun élément n'est actualisé de façon autonome et, en particulier, dont la détermination interne ne peut faire l'objet d'aucune variation,*
- qui ne constituent pas de prédication interne,*
- entre les éléments desquels on ne peut pas faire d'insertion,*
- dont aucun des éléments ne peut faire l'objet d'une substitution synonymique,*

- *et dont le sens global correspond à un concept existant dans la langue et qui pourrait, à l'occasion, être exprimé par un substantif unique¹. »*

Pour sa part, R. Kocourek, qui adopte le point un vue terminologique mais qui accorde toutefois une large part aux apports de la lexicologie, propose la hiérarchie de critères suivante pour identifier des unités terminologiques complexes lexicalisées² :

- 1) existence d'une définition,
- 2) position occupée au sein d'une série synonymique ou au sein du système terminologique – l'existence d'un synonyme déjà lexicalisé tendrait à disqualifier le nouveau syntagme,
- 3) structure syntaxique hiérarchique (régissant et modificateur),
- 4) maniabilité syntagmatique (avec le raisonnement selon lequel un terme peu maniable du fait du nombre important d'unités qui le composent est voué à n'être que très peu utilisé dans les discours),
- 5) imprévisibilité sémantique,
- 6) récurrence attestée (dans plusieurs textes spécialisés),
- 7) cohésion syntaxique.

Si les deux auteurs s'accordent pour convenir qu'il n'y a pas de critère décisif (nécessaire et suffisant), la position qu'ils accordent à la définition est emblématique de deux approches différentes. R. Kocourek l'adopte comme premier critère de sa liste, car son approche est de type onomasiologique, alors que G. Gross, partant d'une approche sémasiologique, en fait le point de contrôle du processus inverse, consistant à partir du fonctionnement des signes pour établir leur signification. Dans la mesure où cette dernière approche a été retenue dans le cadre de ce travail, la « caution » définitionnelle n'a pas été retenue³ il a donc été nécessaire d'adopter des critères explicites pour l'observation des syntagmes caractéristiques.

¹ G. Gross (1996 : 42).

² R. Kocourek (1991 : 151).

³ Cette démarche a néanmoins été utilisée dans chapitre 5, afin de mesurer l'écart éventuel entre la définition terminologique et les occurrences du mot rencontrées dans les textes.

Par conséquent, les critères suivants ont été utilisés afin de rechercher des syntagmes « bien formés » parmi les segments répétés incluant l'une des formes retenues pour le test¹ : parmi les critères syntaxiques, ont été utilisés les indices graphiques, l'absence de certaines classes lexicales (conjonctions, pronoms, verbes auxiliaires, etc.), le respect de formules productives (comme N. Adj., N. Prép. Adj., etc.), l'étendue du syntagme (R. Kocourek estime qu'une terminologie est composée de moins de 20% de termes simples et de moins de 10% de termes complexes de plus que quatre mots), les tests de figement (déplacement du régissant ou du modificateur, expansion, ellipse) ; parmi les critères sémantiques, ont été utilisés le test de la commutation, l'étude de l'imprévisibilité sémantique (plus le sens du syntagme est imprévisible, plus celui-ci a de chances d'être lexicalisé) ; enfin, la « force de l'usage », pour reprendre l'expression d'A. Darmesteter, a été appréhendée à travers la récurrence de la forme ou du segment (fréquence supérieure à 9 et présence dans au moins deux parties consécutives). On donne ci-dessous, pour les quatre groupes annoncés plus haut, les syntagmes lexicalisés présents dans le corpus. Le premier ensemble est constitué de mots pivots représentatifs de l'usage de la langue commune et présents dans les mots-clés de la première section² :

¹ On reprend le classement proposé par R. Kocourek dans sa synthèse très éclairante de R. Kocourek (1991 : 135-151).

² Les termes, au sens de la terminologie du *LBMF* sont marqués en gras.

Tableau 36.
« activité(s) », « produits » et « clientèle » et leurs principaux syntagmes incluants

Mots	Syntagmes	Fréq.
activité / activités		
	activité commerciale	16
	activités de marchés	15
	activités de banque privée	11
	activités de financement	11
	activités de gestion	10
produits		
	gamme de produits	33
	offre de produits	23
	produits d assurance	13
	produits dérivés	12
	produits bancaires	10
	produits d épargne	10
clientèle		
	banque de grande clientèle	19
	clientèle haut de gamme	12
	grande clientèle	11

On remarque que les mots représentatifs de la langue commune ont une productivité relativement faible, compte tenu de leur caractère assez général. En fait, c'est la barrière de la récurrence qui filtre le plus efficacement les syntagmes ayant recours à ces unités comme régissant ou comme expansion. Si l'on abaisse le seuil de fréquence à 5 occurrences, le nombre de syntagmes double ou triple. Ainsi, on a pour « clientèle » les syntagmes suivants ayant une fréquence de 9 à 5 occurrences : « clientèle de particuliers, clientèle institutionnelle, clientèle d'entreprises, clientèle internationale, Grande clientèle, clientèle des Caisses régionales, segments de clientèle, clientèle des jeunes, clientèle privée, clientèle des grandes entreprises. » On note donc la présence d'une strate de syntagmes faisant des apparitions sporadiques selon l'actualité de l'activité telle qu'elle est restituée dans les textes du corpus. Les syntagmes les plus fréquents ne reprennent que très peu la terminologie bancaire du *Lexique de la banque et des marchés financiers*, sauf dans un cas qui tient de l'exception (il s'agit de « produits dérivés » qui est signalé en gras) ; et pour cause : il s'agit principalement de notions économiques. Certes, les « produits bancaires », absents de la nomenclature utilisée pour le test, pourraient figurer dans une terminologie plus ouverte aux notions un peu floues qui caractérisent le jargon bancaire (les produits bancaires s'opposent sur le terrain de l'épargne aux produits

reposant directement ou indirectement sur des portefeuilles d'actions). Il n'en reste pas moins que ces syntagmes restituent assez bien le versant commercial des rapports d'activité avec l'usage de notions qui relèvent d'un flou entretenu parfois par les visées du marketing. On présente ci-dessous le groupe des formes de test correspondant aux « mots-clés » relevant de la terminologie bancaire (au sens du *LBMF*) et servant de tête pour la construction de certains syntagmes lexicalisés.

Tableau 37.
« banque », « crédit » et « épargne » et principaux syntagmes incluants

Mot	Syntagmes	Fréq.
banque		
	banque privée	118
	banque de proximité	56
	Banque de proximité	46
	Banque de proximité en France	36
	Banque de détail	29
	Banque privée	27
	banque de détail	26
	banque de proximité en France	26
	banque de financement	19
	Banque de grande clientèle	19
	Gestion d actifs assurances et banque privée	19
	banque universelle	19
	Banque de financement	17
	banque à distance (terme du LBMF)	17
	banque française	16
	banque commerciale	13
	banque d affaires (terme du LBMF)	12
	banque au quotidien	11
	Banque d investissement	11
	banque d investissement	11
	banque à accès multiples	11
	banque centrale	11
	banque de financement et d investissement	10
	banques à réseau	10
crédit		
	crédit à la consommation	128
	crédit bail (terme du LBMF)	80
	encours [de crédit] (terme du LBMF)	26
	Crédit bail	19
	Crédit à la consommation	15
	encours [de crédits] (terme du LBMF)	13
	crédit bail mobilier (terme du LBMF)	13
	crédit bail immobilier (terme du LBMF)	11
épargne		
	épargne salariale	36
	épargne retraite	24
	épargne logement (terme du LBMF)	16
	épargne bancaire	16
	produits d épargne	10
	épargne longue des salariés	10

Dans ce groupe, la productivité syntagmatique est plus élevée que dans le précédent. Cela s'explique probablement par les trois propriétés cumulées par deux d'entre eux : mots-clés relevant de la langue commune, termes banalisés présents dans le *LBMF* et, enfin, mots appartenant au vocabulaire thématique commun. La

proportion de termes dans lesquels ces mots jouent le rôle de régissant (dans des syntagmes nominaux hiérarchisés, condition nécessaire à leur statut d'unité terminologique complexe lexicalisée), est plus élevée que dans les autres groupes (les sept termes sont signalés en gras). Les remarques relatives à l'ouverture sur les aspects commerciaux de l'activité bancaire (crédit à la consommation, banque de proximité), ainsi que le bon accueil réservé à certains néologismes, y compris avec les flottements dénominatifs qui les accompagnent (banque à distance VS banque à accès multiples), demeurent valables. On note également un certain nombre de doublons du fait de la présence d'une majuscule initiale (banque d'investissement / Banque d'investissement), qui révèle un usage propre à ces unités.

Le troisième groupe de formes-tests rassemble des mots ne relevant pas pleinement de la langue commune (une facette seulement de leur aire sémantique est sélectionnée de manière récurrente dans le corpus RA), sans pour autant appartenir à la nomenclature du *LBMF*. On note pour ce groupe une productivité syntagmatique de moyenne à élevée. Les mots et les syntagmes rassemblés peuvent hériter des propriétés caractérisant les deux autres groupes. On constate à nouveau que la sélection des facettes des aires sémantiques est presque toujours orientée en fonction des thématiques économiques.

Tableau 38.
 « assurance », « entreprises » et « marché(s) » et leurs syntagmes incluant

Mot	Syntagmes	Fréq.
assurance		
	assurance vie	114
	assurance dommages	36
	Assurance vie	21
	Gestion d actifs assurances et banque privée	19
	Assurance dommages	15
	compagnie d assurance	15
	produits d assurance	13
	compagnie d assurance vie	11
entreprises		
	grandes entreprises	41
	petites entreprises	15
	marché des entreprises	12
	financement des entreprises	11
	entreprises françaises	10
marché / marchés		
	part de marché	107
	parts de marché	38
	marché français	36
	marchés financiers	20
	marché primaire	16
	marchés actions	16
	marché du crédit	14
	marchés de proximité	14
	marchés de capitaux	13
	marchés émergents	13
	marché des entreprises	12
	marchés boursiers	10

Le quatrième et dernier groupe réunit des mots du vocabulaire commun ne faisant pas partie des cinquante premiers « mots-clés ».

Tableau 39.

« gestion », « services » et « bancaire(s) » et leurs principaux syntagmes incluant

Mot	Syntagmes	Fréq.
gestion		
	gestion d actifs	117
	Gestion d actifs	59
	Gestion pour compte propre	24
	gestion alternative	23
	Gestion pour compte propre et divers	19
	Gestion d actifs assurances et banque privée	19
	société de gestion	17
	gestion de patrimoine	15
	sociétés de gestion	15
	mandats de gestion	13
	activités de gestion	10
	gestion des ressources humaines	10
services		
	services bancaires	45
	services financiers	33
	offre de services	23
	services de banque	12
	Services financiers	11
	offre de produits et de services	11
bancaire / bancaires		
	services bancaires	45
	groupe bancaire	27
	Produit net bancaire	23
	collecte bancaire	19
	réseau bancaire	19
	épargne bancaire	16
	cartes bancaires	14
	secteur bancaire	12
	relation bancaire	11
	participations bancaires	11
	premier réseau bancaire	11
	produits bancaires	10
	produit net bancaire	10
	carte bancaire	10
	système bancaire	10

Dans ce dernier groupe, à la productivité relativement élevée, la dominante économique de la phraséologie rencontrée dans les rapports d'activité est à nouveau perceptible. Comme dans les autres groupes, on relève un *continuum* allant du terme propre à l'organisme émetteur jusqu'à des expressions grand public promues par le marketing à des fins de distinction commerciale. On peut donc maintenant s'attacher à caractériser plus précisément cette phraséologie et examiner la position qu'elle occupe par rapport aux autres variétés de langue utilisées dans le corpus.

II. DEUXIÈME APPROCHE DIACHRONIQUE : FLUCTUATION DE L'ACTIVITÉ BANCAIRE ET VARIÉTÉS DE LANGUE

A. Syntagmes lexicalisés ayant « banque » pour régissant

La phraséologie des rapports d'activité recourt à des syntagmes dont le sens, en apparence relativement transparent, est fréquemment construit à l'aide des relations « à destination de » (comme « banque de grande clientèle », « crédit à la consommation », « épargne retraite », « financement des entreprises », etc.), ou, plus vague encore, « relatif à » (« activité commerciale », « produits d'assurance », « banque au quotidien », « marchés actions », « secteur bancaire », etc.). Ce type de construction permet au locuteur de rester à la surface des choses tout en offrant à son interlocuteur la possibilité de deviner ce dont il s'agit. La présence de tels syntagmes peut également s'expliquer par le souci de valoriser une sorte de grande épicerie de l'argent, en employant des locutions empreintes d'un économisme ronflant.

Cependant, la langue commune, qui est fortement sollicitée dans ces constructions, est également source d'ambiguïtés dès lors que l'on cherche à saisir la réalité désignée par de tels syntagmes. À cet égard, il convient de remarquer que ces derniers reposent sur un certain nombre de sous-entendus. Afin d'illustrer ce point, on utilisera les exemples donnés en début de section.

Par « grande clientèle », il faut comprendre qu'il s'agit d'une clientèle dont les transactions portent sur des montants très importants (il s'agit d'investisseurs institutionnels, comme les compagnies d'assurance ou les fonds de pension de dimension internationale) ; la « consommation » dans « crédit à la consommation », fait allusion à l'achat des biens de consommation ; la « retraite » de l'« épargne retraite » est une forme d'épargne individuelle destinée à compléter la pension versée dans le cadre d'un régime de retraite collective par redistribution ; le « financement des entreprises » correspond aux crédits consentis par la banque aux entreprises et non aux différentes formes de financement utilisées par les entreprises ; l'« activité commerciale » se résume à la vente de crédits et à la collecte de dépôts ; les « produits d'assurance » renvoient à la vente de contrats d'assurance ; la « banque au quotidien », ce sont les chèques, l'argent liquide, les cartes bancaires, le compte courant utilisés dans la vie quotidienne ; les « marchés actions » sont les

bourses où sont achetées et vendues des actions ; enfin, le « secteur bancaire »...ce sont les banques.

À l'aide de ces exemples, une autre fonction assignée à la phraséologie des rapports d'activité se laisse appréhender : la mise à distance d'une réalité aux aspects souvent banals, par le recours à des notions dont la généralité doit beaucoup à leur imprécision. On peut émettre l'hypothèse selon laquelle la connivence recherchée par le locuteur (voir chapitre 3) repose probablement en partie sur l'effet déréalisant de cette phraséologie. Celle-ci sert de toile de fond, mais aussi de liant aux syntagmes qui circulent entre les deux pôles déjà identifiés de la langue commune et de la terminologie bancaire et financière. C'est ce *continuum* qu'il s'agit maintenant de préciser.

Pour analyser plus finement ce *continuum*, on reprendra les syntagmes dont le mot « banque » constitue le régissant, car, probablement du fait de la position centrale qu'il occupe au regard des variétés langue utilisées dans les rapports d'activité, il offre la plus grande richesse d'exemples. Cela permettra d'évaluer le spectre linguistique propre aux rapports d'activité qu'il sera nécessaire de prendre en compte pour les développements exposés dans les chapitres suivants.

*Tableau 40.
Répartition des syntagmes utilisant le mot « banque » entre différentes variétés de terminologies*

Variétés de langue	Syntagmes
Terminologie spécifique de l'émetteur	banque centrale ¹
Terminologie du <i>LBMF</i>	banque à distance, banque d affaires
Terminologie bancaire générale ²	banque privée, Banque privée, Banque de détail, banque de détail, banque de financement, Banque de financement, banque commerciale, Banque de grande clientèle, Banque d investissement, banque d investissement, banque de financement et d investissement
Vocabulaire du marketing	banque de proximité, Banque de proximité, Banque de proximité en France, banque de proximité en France, banque au quotidien, banque à accès multiples
Vocabulaire économique	banque universelle, banque française, banques à réseau

On voit à partir de cet exemple que le spectre de la phraséologie des rapports d'activité est loin de se réduire à un simple feuilletage de terminologie bancaire et de langue commune. Les syntagmes repris dans cette section suggèrent au contraire l'existence de nombreuses variétés terminologiques se répartissant sur un *continuum* allant d'un vocabulaire propre à l'émetteur jusqu'à la langue commune, en passant par les échelons intermédiaires de terminologies plus ou moins spécialisées relevant du domaine de l'économie au sens large. Or, sa présence massive et multiforme,

¹ La notion de banque centrale désigne d'une façon générale « (...) *la banque qui dans un pays assure l'émission de la monnaie fiduciaire et le contrôle du volume de la monnaie et du crédit.* » *LBMF*. C'est donc par assimilation que ce terme est utilisé pour caractériser la Caisse nationale de Crédit agricole puis Crédit Agricole s.a. dans le corpus. On fait l'hypothèse qu'il s'agit de la transposition de cette partie de l'article L 511-31 du *Code monétaire et financier* : « Art. L. 511-31. - *Les organes centraux représentent les établissements de crédit qui leur sont affiliés auprès de la Banque de France, du comité des établissements de crédit et des entreprises d'investissement et, sous réserve des règles propres à la procédure disciplinaire, de la commission bancaire. Ils sont chargés de veiller à la cohésion de leur réseau et de s'assurer du bon fonctionnement des établissements qui leur sont affiliés. A cette fin, ils prennent toutes mesures nécessaires, notamment pour garantir la liquidité et la solvabilité de chacun de ces établissements comme de l'ensemble du réseau. (...).* »

² Cette « terminologie bancaire générale » est évaluée à partir d'un corpus de rapports d'activité de trois autres établissements bancaires (Crédit Lyonnais, Société Générale, BNP Paribas) couvrant la période 1997-2003.

cette phraséologie semble également caractérisée par une relative instabilité sémantique. C'est fréquemment sur cet argument que les terminologues, et les lexicologues a fortiori, prennent appui pour écarter de leur collecte ce type d'unités. Or, cette instabilité a pour contrepartie une prise indéniable sur l'évolution de l'activité bancaire, telle qu'elle est restituée dans les discours de l'entreprise (et dont les textes du corpus constituent un échantillon). Elle constitue donc un matériau à retenir dans la perspective de l'élaboration de référentiels terminologiques adaptés au contexte. C'est pourquoi on peut étudier cette relative instabilité en observant la manière dont ces syntagmes lexicalisés peuvent évoluer d'une strate à l'autre dans l'espace de la dizaine d'années couverte par le corpus. On utilisera pour cela certains syntagmes lexicalisés donnés dans le tableau ci-dessous. Après avoir illustré le point relatif au caractère fluctuant de cette phraséologie, on tentera de montrer à l'aide d'un exemple comment elle joue aussi un rôle structurant dans la manière d'organiser la présentation de l'activité bancaire. Auparavant, il convient de donner les fréquences des syntagmes pour les différentes parties du corpus. On renverra à ce graphique chaque fois que cela sera nécessaire au cours de l'exposé qui suit.

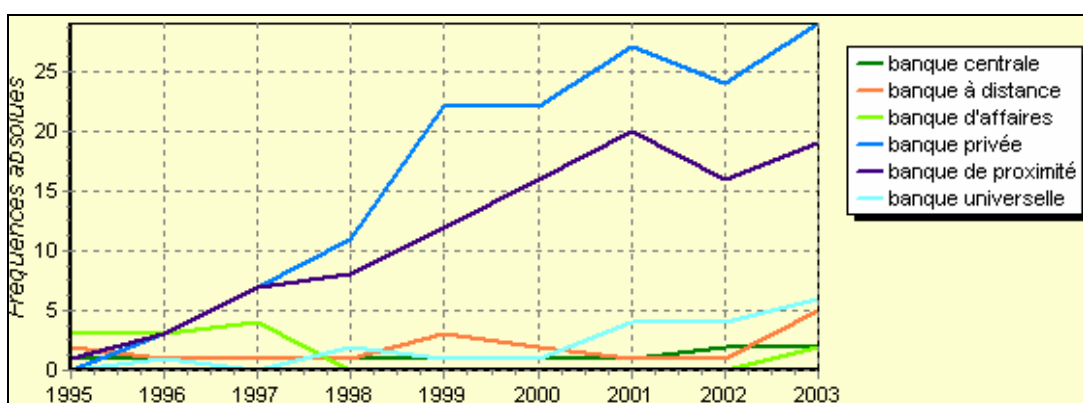


Figure 36.
Ventilation des occurrences de certains syntagmes lexicalisés ayant le mot
« banque » pour régissant

Lecture de la figure 36 : les courbes relient les points qui donnent pour un syntagme lexicalisé donné sa fréquence absolue observée dans une partie du corpus. Par exemple, le syntagme « banque de proximité » (courbe violette) compte une vingtaine d'occurrences pour l'année 2001.

B. Banque centrale : un usage spécifique de la terminologie standard

Comme on l'a vu précédemment, d'un point de vue strictement terminologique, la notion de « banque centrale » ne s'applique ni à la Caisse nationale de Crédit agricole ni à Crédit Agricole s.a. Si cet organisme est qualifié ainsi, c'est que pour jouer son rôle d' « organe central », il centralise une partie de l'épargne des banques qui lui sont affiliées (les caisses régionales), pour ensuite les aider à se refinancer, assurant ainsi leur liquidité par un mécanisme interne au Crédit agricole à la manière des banques centrales¹. Il s'agit dans tous les cas d'une fonction « régaliennne » consubstantielle au rôle d'un organe central tel que la loi l'a défini. À partir du moment où le locuteur collectif des rapports d'activité a estimé qu'il était nécessaire de rappeler cette spécificité du groupe bancaire, le statut d'organe central a en quelque sorte garanti la récurrence du terme « banque centrale », alors que sa fréquence est relativement peu élevée. On a avec cet exemple le cas d'un terme emprunté à la terminologie bancaire générale et qui a été « sur-spécialisé » dans la phraséologie propre à un groupe bancaire.

C. Banque à distance : le « corps à corps » des néologismes

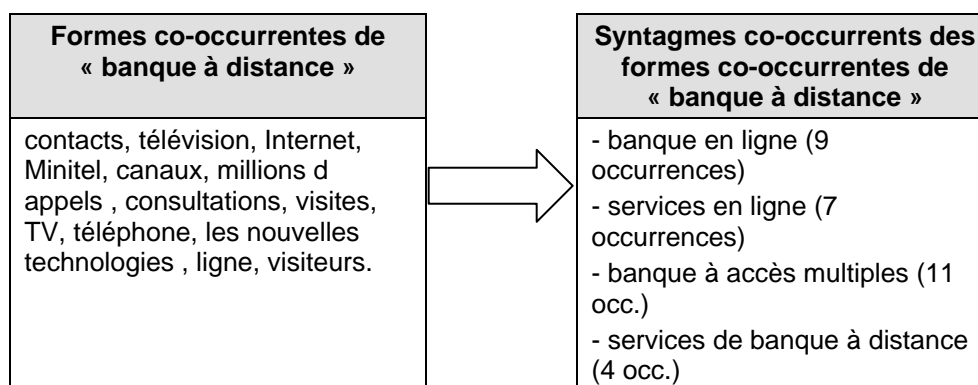
La « banque à distance » désigne d'une façon générale « *toute opération effectuée entre un établissement de crédit et son client sans la présence physique de ce dernier. Cependant, le plus souvent, l'expression est employée lorsque ces opérations sont effectuées par voie électronique*². » Il ne s'agit donc pas d'une sorte de banque mais d'une manière d'accéder aux services offerts par une banque en utilisant des moyens de communication à distance. Compte tenu de la dimension technologique associée à la notion, on peut faire l'hypothèse qu'il s'agit d'un néologisme. On sait que l'apparition d'un phénomène technique s'accompagne en

¹ Parmi les rôles joués par une banque centrale, il y a celui de « banque des banques » : « *le caractère universel de son mode monétaire entraîne l'utilisation de ce dernier comme monnaie de refinancement ; elle est le prêteur en dernier ressort qui permet l'ajustement des trésoreries des banques.* » Par ailleurs, « *elle possède un pouvoir d'autorité sur le système bancaire en contrôlant les opérations de création, de fusion, ainsi que les ratios prudentiels imposés aux banques.* » Y. Crozet et alii. (1993), *Dictionnaire de banque et bourse*.

² B. Sousi-Roubi (2001).

général d'un tâtonnement terminologique explorant diverses manières de désigner une même réalité¹. Qu'en est-il dans le corpus RA ? Pour tenter de le savoir, on dispose de la possibilité de rechercher les co-occurents de ce terme, afin d'identifier d'éventuelles reformulations ; par ailleurs, en étendant les investigations aux co-occurents des principaux co-occurents du terme, il est possible d'explorer des environnements aux caractéristiques lexicales voisines dans lesquels d'autres termes synonymes ou quasi-synonymes seraient susceptibles d'être employés². Dans les deux cas, on utilise le paragraphe comme fenêtre pour le calcul des co-occurents. Celui-ci permet l'identification des formes ou segments spécifiques³ apparaissant dans les mêmes contextes que la forme ou le groupe de formes étudiées. On préfère utiliser ici le découpage en paragraphes, car il permet de capter des reprises anaphoriques nominales susceptibles d'être pertinentes pour l'identification de synonymes ou quasi-synonymes. À partir d'une liste d'une dizaine de co-occurents du segment « banque à distance » dans le corpus RA, on constitue un groupe dont on va rechercher les co-occurents. On présente ci-dessous les résultats de cette exploration :

*Tableaux 41 et 42.
Principales formes co-occurentes de « banque à distance » et principaux syntagmes co-occurentes de ces formes*



¹ P.J. Wexler (1955) et L. Guilbert (1965 et 1975).

² Cette approche est décrite par W. Martinez (2000).

³ Le calcul des co-occurents repose sur le calcul des spécificités propres aux sections dans lesquelles les occurrences de la forme ou du segment étudiés apparaissent. Pour une succincte présentation de ce calcul on renvoie au chapitre 5.

La recherche des principaux co-occurents des co-occurents de « banque à distance » dans le corpus RA permet d'identifier des synonymes ou quasi-synonymes de fréquence moindre. La « banque à accès multiples » est un synonyme direct qui semble toutefois propre au Crédit agricole, qui a baptisé ainsi un programme général de diversification de ses canaux de distribution au milieu des années 1990. « Banque en ligne », « services en ligne » et « services de banque à distance » fonctionnent plutôt comme des spécifiques de « banque à distance ». En effet, ils ne portent que sur les accès via Internet, ou bien ils marquent une spécialisation de « banque à distance » dans cette direction, au détriment d'autres moyens d'accès à la banque (minitel, fax, téléphone, puis télévision interactive). On donne ci-dessous la ventilation de ces différents syntagmes sur la période couverte par le corpus :

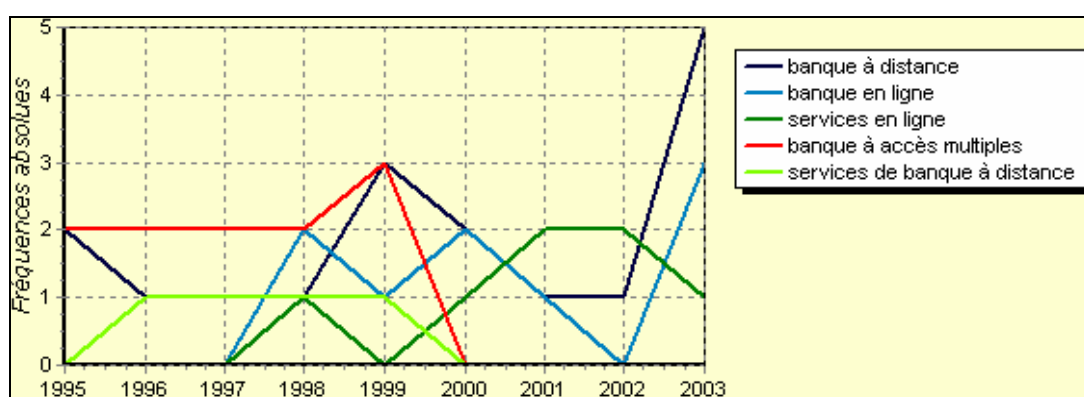


Figure 37.

Ventilation des syntagmes sémantiquement proches de « banque à distance » dans le corpus RA

Ce graphique met en évidence, d'une part, la concurrence des différentes expressions sur toute la durée du corpus, d'autre part, la sélection « naturelle » qui s'applique à ces néologismes. De ce point de vue, l'année 1999 semble marquer un tournant après lequel le terme de « banque à accès multiples » n'apparaît plus et celui de « banque à distance » est recentré autour de la notion de « on line », véhiculée par le développement d'Internet. Du fait de la concurrence des dénominations à laquelle ils sont fréquemment associés, les néologismes sont une autre source d'instabilité marquant la phraséologie des rapports d'activité. On constate, en effet, que durant la période 1995-1999, c'est la phraséologie propre au

groupe bancaire qui semble la plus visible ; après 1999, la terminologie bancaire « standard » du *LBMF* qui sort victorieuse de l'épreuve, semble talonnée de près par la banque et les services « en ligne » accessibles depuis Internet. La notice du *Grand dictionnaire terminologique* datée de 2004 ne reprend pas « banque à distance » mais enregistre en revanche de nombreux synonymes pour « banque en ligne » / « *virtual bank* », semblant ainsi avoir pris acte de la nouvelle tendance : « banque virtuelle », « cyberbanque », « banque sur Internet », « banque électronique » et, synonyme déconseillé « e-banque » (calque de *e-bank*). On va voir dans l'exemple qui suit une illustration du mouvement inverse.

D. Banque d'affaires : la patine d'un terme obsolète

La banque d'affaires s'opposait avant la loi bancaire de 1984 à la banque de dépôts (admise à recevoir du public des dépôts à vue ou à moins de deux ans), en ce que son « (...) *activité principale était la prise et la gestion de participations dans des entreprises dans lesquelles elle ne pouvait investir des fonds à vue ou à moins de deux ans de terme*¹. (...) » Le terme possédait un prestige n'était pas pour rien dans sa perpétuation. Ce point est illustré par une réapparition en 2003, où un prestige persistant est utilisé pour valoriser une offre à destination des PME (petites et moyennes entreprises) : « *Et, pour mieux répondre aux attentes des chefs d'entreprise, un service " banque d'affaires des PME " associant les opérations de haut de bilan et la gestion patrimoniale se déploie dans les Caisses régionales.* » (RA 2003). On notera les guillemets qui soulignent la référence, en même temps qu'ils permettent de citer la dénomination de cette nouvelle offre.

Il reste à expliquer pourquoi un terme si valorisé dans les rapports d'activité disparaît soudainement à partir de 1998. Il faut pour cela observer ce qui se passe dans les trois années précédentes. En effet, celles-ci sont le théâtre d'une restructuration profonde de l'activité de banque d'affaires jusqu'alors tenue par une filiale du Crédit agricole dénommée UI (Union d'études et d'Investissements). En 1995, lorsqu'il est question de banque d'affaires, c'est de l'activité de cette filiale

¹ B. Sousi-Roubi (2001).

que l'on parle. En 1996, avec l'achat de la banque Indosuez, l'appellation change de filiale et UI devient une simple « société d'investissement »¹. L'année 1997 sert de transition et la nouvelle répartition des rôles qui a été mise en place fait l'objet de développements spécifiques dans le rapport d'activité, d'où la légère augmentation de fréquence. Or, en fait, c'est dès 1996 que la phraséologie propre à la banque d'affaires récemment achetée, devenue entre-temps « Crédit Agricole Indosuez », est mise en avant : « *Les activités grandes clientèles, internationales et de marchés du Crédit agricole sont regroupées notamment au sein de Crédit Agricole Indosuez.* » On en a l'illustration ci-dessous avec le graphique présentant la ventilation des dénominations de secteurs d'activité bancaire ayant connu un accroissement de leur fréquence à partir de 1996.

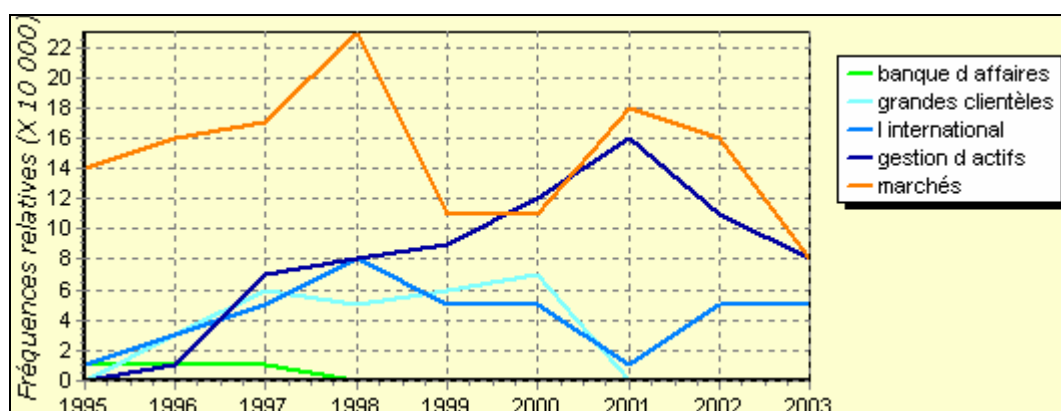


Figure 38.

Ventilation en fréquences relatives des dénominations de domaines d'activité bancaire dont la fréquence augmente à partir de 1996

La phraséologie des rapports d'activité enregistre donc un mouvement d'un autre type, lié cette fois à l'intégration de cultures différentes lors d'achats d'entreprises intégrées au groupe bancaire. Dans ce cas, ce n'est pas tant la prise en compte de la terminologie standard qui accompagne l'intégration ou le développement d'une nouvelle forme de l'activité bancaire, que l'absorption d'une terminologie propre à l'entreprise achetée avec son système de référence. Par

¹ Voir F. Erlos (2006) pour le détail de cette évolution.

ailleurs, on constate également qu'un terme standard obsolète peut être réutilisé seulement en vertu de ses connotations positives, ce qui a pour effet de l'intégrer à une phraséologie spécifique.

E. Banque privée : un parallèle possible entre variation de fréquence et importance d'un secteur d'activité

Le terme de « banque privée » apparaît en 1996, pour la même raison que celle qui explique la disparition momentanée du terme « banque d'affaires » : il s'agit d'une restructuration de l'activité ayant pour cause l'achat de la banque Indosuez. La « banque privée » ne s'oppose pas à la « banque publique » au sens où, par exemple, il y a eu en France à deux reprises des banques nationalisées. Il faut plutôt voir dans ce terme, non répertorié par le *LBMF*, un calque du *private banking* anglo-saxon. L'équivalent français disponible est normalement « gestion privée¹ », car il ne s'agit pas d'une sorte de banque mais d'une partie de l'activité bancaire qui correspond aux activités de gestion de patrimoine appartenant à de riches particuliers. La progression de la fréquence observée d'année en année semble parallèle au développement de ce secteur au sein du groupe bancaire (les actifs gérés passent de 28 milliards d'euros en 1997 à 88 milliards en 2003). De même, l'augmentation de la fréquence en 1999 doit être rapportée à la réorganisation du secteur qui a eu lieu cette année-là et qui occasionne des commentaires plus nombreux que dans la période précédente :

« L'année 1999 a également été marquée par la réorganisation des activités de banque privée du Crédit agricole. En France, le rapprochement des équipes de la Banque de Gestion Privée et de la Banque Privée en France de Crédit Agricole Indosuez a donné naissance à la BGP Indosuez, filiale de Crédit Agricole Indosuez. De même, en Suisse, la Banque du Crédit Agricole Suisse (BCAS) et les

¹ « gestion privée » possède quatre occurrences dans le corpus pour les années 1995 et 1997 puis 2003. L'explication de sa disparition par la prise en compte dans les rapports d'activité de terminologies propres aux cultures des entreprises achetées et intégrées au groupe bancaire vaut également dans ce cas. En effet, la réapparition de la forme en 2003 se fait dans un paragraphe où il est question de l'activité du Crédit Lyonnais récemment acquis. On peut ajouter que le calque « banque privée » préféré à « gestion privée » permet d'éviter un rapprochement gênant avec « gestion d'actifs », calque d' « *asset management* ».

succursales de Crédit Agricole Indosuez ont été regroupées au sein de Crédit Agricole Indosuez (Suisse) SA, filiale bancaire de Crédit Agricole Indosuez. » (RA 1999).

Il apparaît à cette occasion qu'un parallèle semble pouvoir être établi entre accroissement de fréquence et développement d'un secteur d'activité, tel qu'il est évoqué dans les rapports d'activité. Cependant, il faut se garder de voir une relation mécanique entre la ventilation des occurrences dans le corpus et l'importance grandissante prise par la réalité désignée par ce syntagme lexicalisé au sein du groupe bancaire.

F. Banque de proximité : les apports d'un suivi topographique et diachronique

En effet, le démenti est rapidement apporté sur ce point par l'expression « banque de proximité » qui connaît un accroissement de sa fréquence proche de l'expression précédente. En fait, tout se passe comme si le concept marketing de « banque de proximité » était mis en avant pour contrebalancer le développement du groupe bancaire et l'acquisition progressive d'une dimension internationale de plus en plus visible¹. Du point de vue des représentations de la banque que la communication institutionnelle cherche à promouvoir, la « banque de proximité » fonctionne peut-être comme un lointain écho de la campagne publicitaire de 1976 qui avait popularisée la marque avec le slogan « le bon sens près de chez vous ». Mais la réalité désignée n'a pas changé, il s'agit toujours dans les grandes lignes de la banque de détail...en France, la mention du pays étant devenue nécessaire entre-temps. Cependant, cette sorte de banque, qui n'en est pas une en réalité, est très présente dans les discours. Bien plus, elle est utilisée pour structurer la présentation de l'activité bancaire, ce qui est rendu perceptible par l'apparition de distinctions graphiques appliquées à certains termes ou expressions. La « banque de proximité » fournit de ce point de vue un exemple intéressant, mais il n'est pas unique : ce phénomène est même assez banal ; pour s'en tenir aux syntagmes lexicalisés ayant « banque » pour tête, on le constate pour « banque privée », « banque de détail »,

¹ Voir chapitre 4.

« banque de financement » et « banque d'investissement ». Afin de mesurer l'importance de ce phénomène, on s'attachera à présenter la répartition de cette dénomination sur la topographie des textes du corpus. On reprendra pour cela la distinction opérée au chapitre 3, entre contextes à dominante syntactique et non syntactique.

Lorsque l'on distingue les graphies de « banque de proximité », on peut voir que la répartition du syntagme semble dépendre de la nature des contextes. Ce phénomène présente un intérêt lorsque l'on s'intéresse à la manière dont l'activité bancaire est restituée. En effet, on l'a vu précédemment, les contextes non syntactiques sont plutôt réservés à la présentation des aspects structurels du groupe bancaire, qu'il s'agisse de structure financière, capitalistique ou organisationnelle. La présence du syntagme dans une zone de ce type est donc une indication du rôle que le locuteur collectif des rapports d'activité lui fait jouer. Grâce à la carte des sections de Lexico 3, il est possible d'adopter pour le corpus RA un point de vue à la fois diachronique et topographique, qui permet de restituer l'évolution de l'usage du syntagme. Pour cela, on projette séparément sur la carte des sections (un carré correspond ici à un paragraphe du texte original), les deux graphies « Banque de proximité » et « banque de proximité ».



Figure 39.

Occurrence du syntagme « Banque de proximité » de la partie 1995 du corpus RA

Lecture de la figure 39 : la carte des sections est découpée selon les paragraphes de la partie 1995 du corpus RA. La partie est divisée en deux ensembles regroupant les paragraphes relevant de contextes à dominante non syntactique (1995cdn) et syntactique (1995cda). Le carré coloré en orange dans la sous partie 1995cda correspond à l'unique occurrence du segment « Banque de proximité » relevée pour cette année.

L'unique occurrence du syntagme pour l'année 1995 figure dans un contexte à dominante syntactique. La majuscule s'explique ici par la position du syntagme en début de phrase : « Banque de proximité grâce à sa structure décentralisée et à la densité de son réseau, il est aussi le premier banquier des ménages. » On est encore à l'époque où le premier réseau bancaire, doté d'une soixantaine de banques de taille régionale, voire départementale, et de ce fait particulièrement bien implanté sur le territoire national, n'a pas besoin de souligner une proximité de la clientèle qui semble aller de soi. Ce n'est plus le cas cinq ans plus tard :

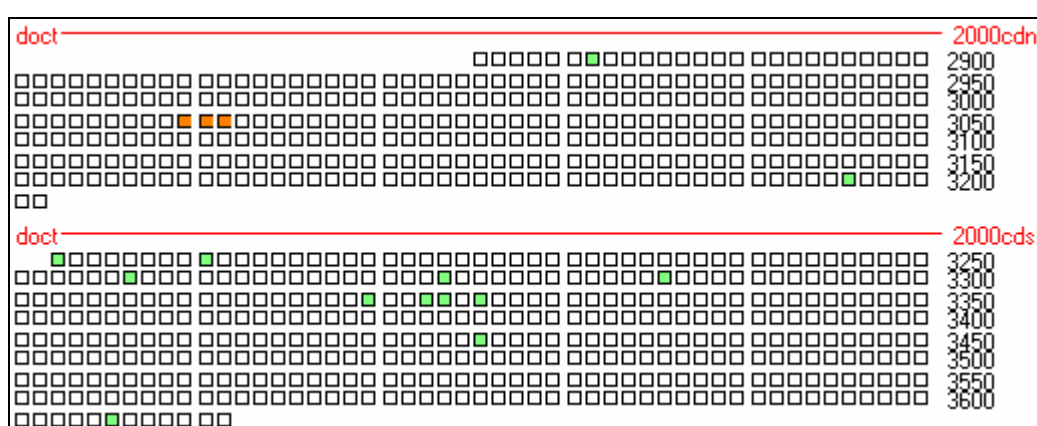


Figure 40.
Occurrences des syntagmes « Banque de proximité » (carrés orange) et « banque de proximité » (carrés verts) pour la partie 2000 du corpus RA

Pour l'année 2000, les contextes non syntactiques utilisent le syntagme « Banque de proximité » afin d'organiser une partie de l'activité du groupe représentée par ses filiales ou ses principales participations. La notion fonctionne en opposition avec d'autres regroupements, comme la « gestion d'actifs, l'assurance-vie et la banque privée », les « grandes clientèles et l'international » et les « participations et partenariats bancaires en Europe et dans le monde » :

« Banque de proximité

Banque de proximité France : Groupe Sofinco FRANCE, Cedicam, Transfact Foncaris Services centraux de titres Sud Mer Brunoy, Sociétés de financement régionales.

Banque de proximité ETRANGER : Groupe Sofinco ETRANGER : Allemagne, Espagne, Hongrie, Italie, Maroc, Pays - Bas, Portugal. Bankoa : Espagne. Banco Bisel : Argentine. Banque Libano-Française : Liban. »

Les Caisses régionales de Crédit agricole, comme dans les années précédentes, font l'objet d'un traitement à part dans les contextes non syntactiques. En revanche, dans les contextes syntactiques, la banque de proximité commence à chapeauter d'une manière encore assez indécise les Caisses régionales de Crédit agricole et certaines filiales spécialisées du groupe bancaire « *Leader de la banque de proximité en France grâce à sa structure décentralisée et à la densité de son réseau de 7679 agences, le Crédit Agricole est le premier banquier des particuliers avec 16 millions de clients. Financier privilégié de l'agriculture depuis l'origine, il a élargi au fil des années son domaine d'activité pour être, aujourd'hui, le partenaire de tous les acteurs de la vie économique : particuliers, entreprises, professionnels et collectivités locales. Outre les services bancaires classiques, le Crédit Agricole propose à ses clients une offre complète de services financiers et d'assurance conçus par ses filiales nationales spécialisées¹.* » La présence de nombreuses occurrences du syntagme sans majuscule initiale fonctionne comme un indice de la banalisation de la notion, mais cette banalisation peut être trompeuse. Là encore, les contextes non syntactiques permettent de suivre finement l'évolution de ce syntagme lexicalisé servant à désigner une notion aux contours imprécis. C'est ce que l'on peut voir à l'aide de la carte des sections pour la partie 2003 :

¹ *Rapport d'activité de l'année 2000.*

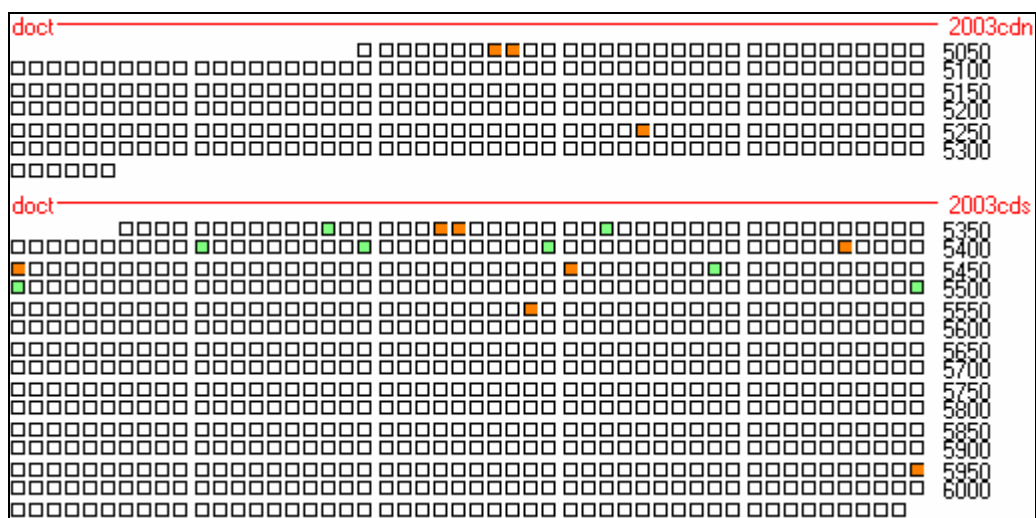


Figure 41.
 Occurrences des syntagmes « Banque de proximité » (carrés orange) et « banque de proximité » (carrés verts) pour la partie 2003 du corpus RA

Sur cette carte des paragraphes de la partie 2003, on note l'apparition, nouvelle par rapport à l'année 2000 de formes du syntagme avec majuscule initiale dans les contextes à dominante syntactique. Cette présence ne signale pas, comme pour 1995, des débuts de phrases classiques, mais des phrases nominales ou des intitulés de lignes de tableau. Ce phénomène a été identifié dans le chapitre 4 comme l'une des marques de l'apparition de contextes mixtes à partir de l'année 2001. En revanche, c'est la stabilité apparente des contextes non syntactiques qui est trompeuse. On retrouve en effet une configuration proche de celle de l'année 2000. Ce que dément l'observation du contenu des paragraphes ainsi isolés :

« Banque de proximité en France - Caisses régionales de Crédit Agricole.

Banque de proximité en France - Réseau du Crédit Lyonnais.

Banque de proximité en France : 43 Caisses régionales, 25% ; Crédit Lyonnais 94,82%. » (RA 2003). »

L'extension de la notion de « banque de proximité » a évolué en l'espace de trois ans, mais elle demeure, probablement pour perpétuer son rôle de valeur représentative d'une certaine stabilité du groupe bancaire. La cotation de 2001 et l'acquisition du Crédit Lyonnais en 2003 ont profondément modifié les équilibres au sein du Crédit agricole avec l'émergence d'un deuxième groupe constitué en son sein autour de Crédit Agricole s.a. Avec l'acquisition du Crédit Lyonnais, Crédit

Agricole s.a. s'est dotée d'une filiale au statut particulier puisqu'il ne s'agit plus, comme par le passé, d'une entreprise exerçant un métier bancaire ou parabancaire spécialisé, mais d'un réseau bancaire au même titre que celui des Caisses régionales de Crédit agricole. Tout se passe comme si, en quelques années, le centre de gravité de la notion de « banque de proximité » s'était déplacé de l'activité des seules Caisses régionales à la conjonction de cette activité avec celle des filiales spécialisées, pour englober finalement l'activité des caisses régionales de Crédit agricole (à hauteur des 25% de leur capital possédé par Crédit Agricole s.a.) et celle de la banque de détail du Crédit Lyonnais et, de manière plus accessoire, l'activité des filiales spécialisées : « *Crédit Agricole S.A. est le leader des activités de banque de proximité en France, grâce aux positions de premier plan acquises de longue date par les Caisses régionales. Le succès de l'offre publique de rapprochement avec le Crédit Lyonnais a encore renforcé cette prééminence, complétée par les positions très solides des filiales de gestion d'actifs, d'assurances et de services spécialisés*¹. » Comme le montre cet exemple, le flou entourant l'intension de certains syntagmes propres au marketing permet d'en modifier l'extension au fil du temps ; ce qui nécessite, toujours dans la perspective d'une collecte destinée à un référentiel terminologique adapté au contexte, de tenir compte des évolutions diachroniques affectant la phraséologie des rapports d'activité, surtout lorsque certaines notions jouent un rôle d'organisateur au sein d'un référentiel spécifique.

G. Banque universelle : un exemple de phraséologie en mutation

Une « banque universelle » est « *un établissement financier qui fournit toute la gamme des services bancaires à ses clients*². » On a vu dans le chapitre 2 que ce qualificatif n'a pu s'appliquer à la lettre au Crédit agricole qu'à partir de 1991, date à laquelle la banque a offert ses services à toutes les entreprises sur l'ensemble du territoire français. Cette réalité est rappelée à intervalles réguliers jusqu'en 2001, où elle connaît une inflation significative. Dans tous les cas, l'augmentation de fréquence est liée à un changement de dimension du groupe bancaire ou à son

¹ *Rapport d'activité de l'année 2003.*

² A. Silem et J.-M. Albertini (2006).

évocation : en 1996, l'augmentation de fréquence est couplée avec l'achat de la banque Indosuez ; en 1998, la célébration du dixième anniversaire de la privatisation de l'organe central du Crédit agricole est l'occasion de rappeler les étapes d'une transformation débouchant sur son universalité en matière bancaire ; 2001, enfin, date de transformation de la Caisse nationale de Crédit agricole en Crédit Agricole s.a., société cotée en bourse, est l'occasion de rappeler cette dimension : « *Depuis maintenant plus d'un siècle, l'histoire du Crédit Agricole est une belle aventure bancaire, au service de tous les acteurs économiques. Le Crédit Agricole a su se développer jusqu'à changer de dimension pour devenir une banque universelle.* » (RA 2001). Dans le rapport d'activité de 2003, le concept banal est promu au rang de fait caractéristique du groupe dans le paysage bancaire français. C'est que le rachat du Crédit Lyonnais et de certaines de ses filiales a renforcé le groupe sur tous ses métiers, ce qu'une visée de communication institutionnelle, hagiographique par essence, ne pouvait laisser échapper : « *Par ailleurs, fort des positions de premier plan qu'il occupe dans le crédit à la consommation et autres services financiers spécialisés, dans l'assurance vie et dommages, dans la gestion d'actifs et la banque privée, le Groupe propose aujourd'hui un nouveau modèle de banque universelle : la banque universelle de proximité, fidèle à sa vocation de service.* »

Il est frappant de constater que le défigement de « banque de proximité » n'affecte ni sa fréquence ni celle de « banque universelle » ; le terme hybride venant s'ajouter aux unités préexistantes, contribuant ainsi à complexifier la phraséologie des rapports d'activité. On voit aussi comment un terme banal du vocabulaire économique¹ peut être amené à jouer un rôle central dans la terminologie spécifique propre à groupe bancaire.

¹ Le syntagme lexicalisé « banque universelle » est repris dans *Le vocabulaire de l'économie et des finances* (2006) et a paru au Journal officiel du 22 septembre 2000.

En utilisant le vocabulaire commun pour explorer la phraséologie des rapports d'activité, on a voulu d'abord contrôler les observations réalisées dans les chapitres précédents. Cet objectif est atteint dans la mesure où l'on a retrouvé parmi les cinquante premières formes du vocabulaire commun thématique les variétés de langue qui avaient été précédemment identifiées. Il a également été possible d'observer le fonctionnement d'unités qui échappaient à la répartition entre langue commune et terminologie bancaire, telle qu'elle peut être représentée par la nomenclature du *Lexique de la banque et des marchés financiers*.

Il ressort de ces observations que la tendance générale caractéristique de l'usage de ces mots dans les rapports d'activité repose sur la sélection quasiment systématique de la partie de leur aire sémantique qui ressortit au domaine de l'économie compris au sens large, puisque la banque est concernée au même titre que le marketing et les activités commerciales. Le vocabulaire des rapports d'activité est au service d'une restitution de l'activité bancaire immergée dans les contextes économiques national et mondial. L'évocation de cette réalité s'appuie sur un référentiel spécifique dont les noms propres de personnes et de lieux constituent des représentants privilégiés relativement présents dans le vocabulaire commun thématique, puisqu'ils correspondent à un cinquième des effectifs.

Les variétés de langue jouent des rôles distincts dans la restitution de cette activité. La langue commune, utilisée pour proposer un survol général de l'activité (chapitre 4), fournit un vocabulaire commercial un peu passe-partout et facilement accessible ; elle semble particulièrement sollicitée dès lors qu'il s'agit de caractériser l'activité bancaire dans son ensemble. La terminologie bancaire est employée de préférence pour évoquer les opérations et les objets bancaires, ancrant ainsi le vocabulaire du corpus dans une réalité économique spécifique. Quelle que soit la variété de langue à laquelle les unités utilisées se rattachent, on constate qu'elles sont mises à contribution pour produire une phraséologie dans laquelle les syntagmes lexicalisés sont relativement abondants. En partant de douze unités destinées à repérer la présence de syntagmes lexicalisés dont la fréquence dans le corpus est égale ou supérieur à 10, on obtient plus d'une centaine de ces syntagmes. Certains d'entre eux sont répertoriés dans le *LBMF* ; cependant ils constituent un

groupe minoritaire représentant environ 6% de la collecte. La majorité de ces syntagmes lexicalisés se répartissent sur un *continuum*, allant de la terminologie propre à l'émetteur des rapports d'activité jusqu'à une phraséologie commerciale plutôt imprécise, en passant par une terminologie bancaire étendue et une terminologie économique relativement banale. Ces distinctions fines entre variétés de langue ont permis d'examiner la manière dont ces syntagmes sont amenés à évoluer au fil du temps ; car, outre leur abondance, ils sont caractérisés par une relative instabilité sémantique.

En effet, on a pu mettre en évidence à l'aide des syntagmes utilisant « banque » comme régissant un certain nombre de phénomènes relevant d'une sorte de flottement sémantique, résultat des interactions entre données extralinguistiques de la situation de communication et restitution de l'activité bancaire médiatisée par le genre discursif. Un détournement de la terminologie bancaire (« banque centrale »), s'explique par le rôle d'organe central joué par la Caisse nationale de Crédit agricole puis Crédit Agricole s.a. La terminologie « standard » du *LBMF* n'est pas la seule utilisée afin de caractériser l'activité bancaire : elle est, en fait, en concurrence permanente, d'une part, avec des terminologies *ad hoc* propres à l'organisme émetteur (« banque à accès multiples »), d'autre part, avec les néologismes qui la travaillent en quelque sorte de l'intérieur (« banque en ligne »).

Une terminologie approximative peut résulter également de la force d'attraction d'un calque du terme anglo-saxon préféré au terme en place : « *private banking* », traduit habituellement par « gestion privée », devient « banque privée » en conservant le même sens. Pour ce qui est du vocabulaire des rapports d'activité, ce changement est lié directement à la prise en compte d'une culture d'entreprise extérieure, à la suite de l'achat d'une banque ayant particulièrement développé ce type d'activité. Il en est de même pour « banque d'affaires », remplacé à une date précise par les syntagmes « société d'investissement », « banque de grande(s) clientèle(s) », « banque d'investissement », « banque de financement », etc.

Lorsque le syntagme relève plutôt d'une phraséologie commerciale, le flou caractérisant l'intension de la notion favorise un usage référentiel expansif et par conséquent difficile à cerner. La « banque de proximité » voit ainsi son champ s'accroître sensiblement en moins d'une dizaine d'années. Le moteur de cette

évolution réside dans le rôle que l'on fait jouer à la notion : par un jeu d'oppositions, elle sert à délimiter les pans d'activité relevant du national au fur et à mesure que l'activité internationale se développe.

Or, un autre aspect est également sous-jacent : c'est le souci de donner une image équilibrée de l'organisation du groupe bancaire. On rejoint ici des préoccupations évoquées dans le chapitre 2, relatives aux intentions de l'émetteur ; celles-ci sont guidées, entre autres, par la volonté de présenter un aspect rassurant du développement de l'activité, à l'opposé d'un développement opportuniste, désordonné et, par conséquent, difficilement prévisible. L'expansion du périmètre de la « banque de proximité » peut alors être interprétée comme l'expression du souci de l'émetteur de donner plus de poids à cette notion. C'est, semble-t-il, le même mécanisme qui produit cet autre résultat : lorsqu'il s'agit de prises de participations dans des banques étrangères, la « banque de proximité » devient la « banque de détail à l'étranger » ; une « banque de proximité à l'étranger » pourrait en effet laisser le lecteur perplexe. La « banque universelle », terme bancaire courant, hors *LBMF*, subit une évolution spécifique qui s'explique par le mouvement sous-jacent des tensions qui secouent le modèle de groupe bancaire auquel le Crédit agricole souhaite se rattacher. L'improbable expansion représentée par « banque universelle de proximité » témoigne de l'agitation intérieure qui anime toute phraséologie spécifique.

De façon plus générale, ce chapitre a permis de mettre en lumière l'utilité d'aborder les phraséologies à l'aide d'une série textuelle chronologique. Seule cette mise en perspective est capable de fournir les clés nécessaires pour saisir les évolutions propres à un parler d'entreprise. Ce parler, en prise sur l'activité concrète, évolue autant sous l'effet de causes extérieures que sous celui de sa logique propre. Certaines de ses unités, que l'on pourrait considérer sur des critères externes comme peu dignes d'intérêt, jouent au contraire un rôle structurant dans l'organisation de la restitution d'un référentiel. À partir de ces observations, il est possible d'aborder la question de la sélection lexicale dans le genre de discours des rapports d'activité sous les aspects de la diachronie et de la prise en charge énonciative.

CHAPITRE 7

UN VOCABULAIRE PEUT-IL ÊTRE TRANSPARENT ?

Les trois chapitres précédents ont permis d'affiner les connaissances relatives aux variétés de langue utilisées dans les rapports d'activité. On a vu quel rôle elles peuvent jouer dans la restitution d'une activité bancaire. Ces analyses ont été l'occasion de remarquer qu'à l'exception des termes récurrents absents des nomenclatures des dictionnaires de langue, une tendance dominante caractérise le vocabulaire des rapports d'activité : un parti pris de généralité qui se traduit pour les termes par un parti pris de généricité. L'expression « parti pris » suggère une intervention délibérée du locuteur, une certaine maîtrise de sa part dans l'usage qu'il fait d'un vocabulaire spécifique. Or, certains éléments semblent indiquer que ce n'est pas le cas : on a ainsi constaté la présence de termes dont on peut raisonnablement penser qu'ils sont abscons pour un large public ; par ailleurs, l'étude de certains syntagmes lexicalisés montre l'existence d'une phraséologie propre, dont on peut se demander parfois si sa prolifération est bien maîtrisée. En effet, on a pu constater que le caractère général de certaines notions repose sur un flou certain, que des manières différentes de parler d'une même réalité coexistent de façon un peu anarchique, qu'enfin, la production syntagmatique, animée par un double mouvement externe et interne, semble condamnée à proliférer sans autre limite que l'usure de certaines de ses unités.

Dans la mesure où l'on considère que l'explication de la sélection des unités lexicales chez un locuteur, qui plus est collectif, échappe encore largement à l'analyse, il reste néanmoins possible d'éclairer la caractérisation du vocabulaire des rapports d'activité en prenant en compte le rôle de filtre thématique et stylistique que peut jouer le genre de discours. On a rappelé plus haut ce qu'a permis de préciser la caractérisation du vocabulaire : thématique économique, restitution de l'activité bancaire en des termes assez généraux, usage de variétés de langue appelées à jouer un rôle différent dans cette restitution. On peut poursuivre l'enquête en se demandant si les finalités du genre de discours représenté par les rapports d'activité n'ont pas d'autres influences sur la manière d'utiliser ce vocabulaire. On peut alors chercher à interroger la limite étroite qui sépare ce qui est dit de ce qui est tu, ce qui est implicite de ce qui mérite une glose.

Pour répondre à la première interrogation, on utilisera des éléments tirés de la catégorisation des composants de l'activité bancaire, afin d'interroger certains contrastes de fréquence. Le deuxième point sera abordé par une enquête sur les éventuelles propriétés de discours seconds des rapports d'activité. On se livrera à une interrogation à partir d'un point de vue différent, des articulations entre les principales variétés de langue qui sont utilisées dans le corpus. Enfin, on se demandera si un vocabulaire de ce type peut être tout à fait transparent.

I. FILTRAGES THÉMATIQUES ET VARIATIONS DE FRÉQUENCES DANS LES RAPPORTS D'ACTIVITÉ

A. Effet de discrétion, ou de l'économiquement correct

Les variations de fréquence affectant certaines unités du vocabulaire et les événements liés à l'activité bancaire tels qu'ils sont relatés dans les textes du corpus sont susceptibles d'être articulés de diverses façons. L'étude de ces articulations a pour but d'interroger la sélection du lexique liée à la situation de communication particulière dans laquelle les rapports d'activité sont produits. On ne travaille que sur les traces laissées par les discours et, par conséquent, on n'accède jamais à une part du lexique, virtuelle ou passive, qui aurait pu être utilisée mais qui ne l'a pas été. Cela n'empêche pas d'interroger les variations de fréquence affectant les mots dans les textes d'un corpus, car de même que le lexique virtuel reste inaccessible, de même ces variations, lorsqu'il est possible de les observer sur la durée, obéissent-elles probablement à des motifs qu'il s'agit de cerner. Cette question peut être abordée à l'échelle d'un syntagme lexicalisé, comme dans le précédent chapitre. Dans cette première section, on cherchera à dégager les principales tendances observables à partir de l'étude des différentes variétés de langue rencontrées dans les textes du corpus.

On a eu l'occasion de voir dans les précédentes sections que les termes et la phraséologie présents dans les rapports d'activité connaissent des fréquences plus basses et des variations plus importantes que les formes simples identifiées comme appartenant à la langue commune. Parmi les cinquante-sept termes « récurrents », un tiers n'est pas présent dans toutes les parties du corpus. En revanche, pour les vingt-trois mots identifiés comme relevant de la langue commune, deux seulement ne font pas partie du vocabulaire de base. Avec l'étude des syntagmes lexicalisés caractéristiques de la phraséologie du corpus, on a déjà pu constater le fait que les variations de fréquence sont susceptibles, dans certains cas (le développement d'un secteur d'activité), d'interagir assez étroitement avec la restitution de l'activité bancaire. Encore faut-il préciser ce point : le fait que « banque privée » connaisse un

accroissement du nombre de ses occurrences entre le début et la fin de la période couverte par le corpus n'est pas sans rapport, non seulement avec l'importance que le locuteur cherche à donner à cette activité dans la restitution de l'activité du groupe bancaire, mais aussi avec l'importance prise par cette activité dans les résultats économiques du groupe. En revanche, si « banque de proximité » connaît une hausse similaire, c'est pour une raison bien différente qui relève plutôt d'une tentative d'organisation de l'activité bancaire telle qu'elle est restituée dans les rapports d'activité. De ce point de vue, l'analyse textométrique permet d'apporter aux extrapolations hâtives autant de démentis que nécessaire. Entre ces deux cas de figure archétypiques, il existe une série d'échelons intermédiaires dont il convient de rendre compte. Pour cela, on réutilisera le découpage de l'activité bancaire en composants de base (agents, objets, opérations et propriétés) proposé dans les précédents chapitres.

Un premier phénomène concerne ce que l'on pourrait appeler un « effet de discrétion » qui, lorsqu'il s'applique à certaines réalités désignées par les mots du corpus, a tendance à minorer la fréquence des expressions utilisées, non pas parce que la réalité en question a disparue ni qu'elle ait perdu de son importance, mais parce qu'il n'est plus très pertinent, compte tenu par ailleurs des finalités poursuivies à travers les rapports d'activité, de continuer à l'évoquer. On prendra, pour illustrer ce point, l'exemple de la baisse des taux d'intérêt durant la période couverte par le corpus. Il s'agit d'un phénomène économique général dont les répercussions sur l'activité bancaire sont directes. On peut résumer très succinctement les enjeux de cette baisse des taux pour les banques : les taux d'intérêt du marché permettent aux banques de déterminer à quel taux il est possible de prêter de l'argent tout en gagnant, et à quel taux il faut rémunérer l'épargne pour la capter au juste prix. Lorsque les taux sont bas, et ils n'ont pratiquement pas cessé de baisser sur la période, les marges des banques et la rémunération de l'épargne baissent également. On donne ci-dessous la tendance générale des taux à court et moyen terme sur la période 1995-2003.

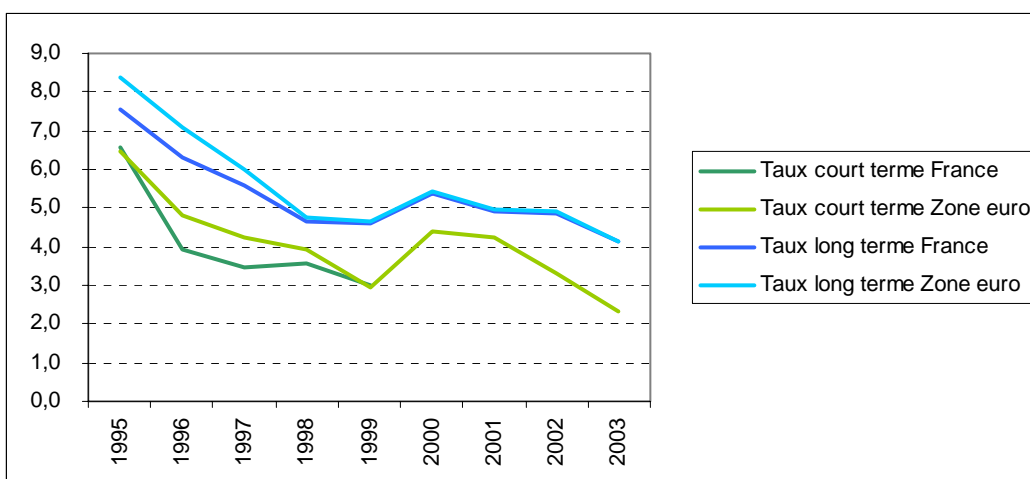


Figure 42.

En pourcentage (colonne des ordonnées), les principaux taux d'intérêt du marché pour la période 1995-2003¹

On peut alors comparer cette courbe des taux d'intérêts avec celle des fréquences du terme « taux d'intérêt » dans les différentes parties du corpus :

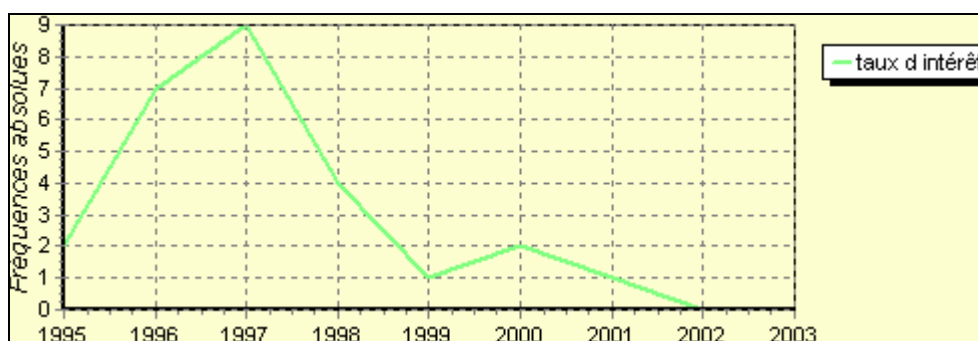


Figure 43.

Ventilation des fréquences du segment « taux d'intérêt » sur le Corpus RA

De la comparaison des deux courbes, il ressort que les années 1996-1997 correspondent, d'une part, dans le monde économique, au début de la baisse des taux, d'autre part, dans le corpus de rapports d'activité, à une augmentation de la fréquence du terme. Les deux phénomènes paraissent liés, lorsque l'on observe les

¹ Pour les taux court terme, la référence sont les taux du marché à trois mois si disponibles ou instruments financiers similaires. Pour les taux long terme, la référence sont les emprunts phare d'État à 10 ans si disponibles ou instruments financiers similaires. Dans les deux cas, il s'agit d'un pourcentage annuel. Source : *Base de données des Perspectives économiques de l'OCDE*, n° 79 et <http://www.oecd.org/eco/sources-and-methods>.

contextes d'apparition du terme dans les textes de 1996 et 1997 : il n'y est presque question que de la « baisse des taux d'intérêts ». Mais on peut se demander pourquoi, la baisse continuant, il est de moins en moins question de ce phénomène dans le corpus. C'est que la baisse des taux, qui est invoquée dans un premier temps pour justifier certaines contre-performances, devient très probablement un sujet économiquement incorrect sur lequel il n'est plus nécessaire de s'étendre. Cela se justifie d'autant plus que l'augmentation de l'activité économique favorisée par un crédit devenu meilleur marché peut être mise en avant. Par ailleurs, il n'est pas très surprenant que les rapports d'activité, centrés sur l'activité d'une entreprise en particulier, n'accordent pas de longs développements à un phénomène économique global. Le discours des rapports d'activité rétablit donc une sorte d'équilibre rompu par une crise des taux qui doit plutôt être présentée comme une parenthèse. On peut observer un phénomène inverse qui relève de la mise en valeur.

B. Effet de valorisation, ou l'actualité institutionnelle...en dépit du reste

Pour illustrer un autre aspect de la manière dont un vocabulaire travaille sous l'effet d'un genre, on prendra pour toile de fond le cours de la bourse en France, synthétisé par l'indice CAC 40¹. On donne ci-dessous les principaux cours de clôture de la période.

¹ « Indice calculé et publié par Euronext Paris à partir de 40 valeur cotées sur le Premier marché choisies pour leur forte capitalisation boursière, leur large diffusion dans le public, leur représentativité sectorielle et l'importance en volume des transactions les concernant. Cet indice exprime la capitalisation boursière instantanée de ces quarante valeurs par rapport à leur capitalisation boursière de référence fixée à 1000 au 31 décembre 1987. (...) » B. Sousi-Roubi (2001).

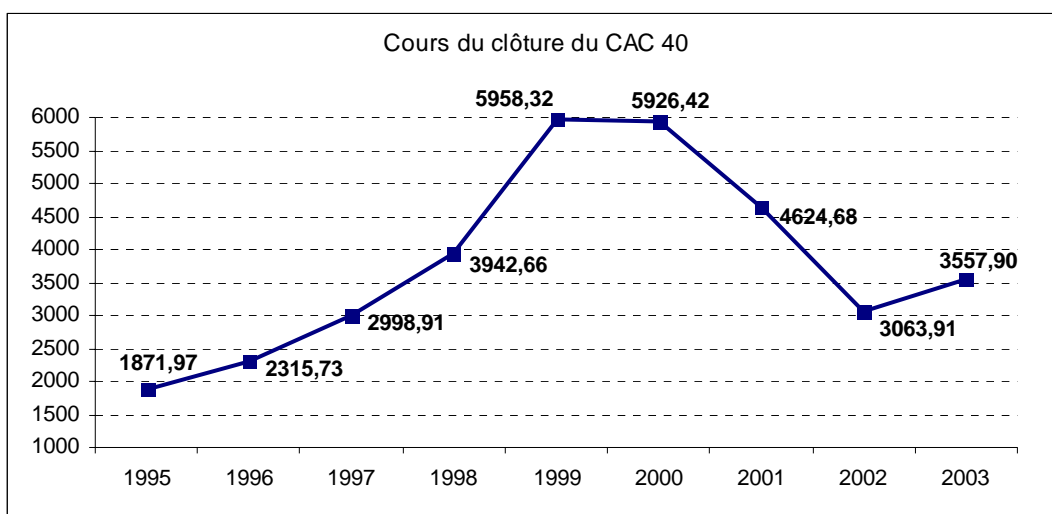


Figure 44.
Cours de clôture du CAC 40 sur la période 1995-2003¹

Afin de mettre en évidence le phénomène de valorisation de l'actualité institutionnelle du groupe bancaire dans les rapports d'activité, on reprend certains termes identifiés dans le chapitre 6, dont on observe la répartition des fréquences sur les différentes parties du corpus.

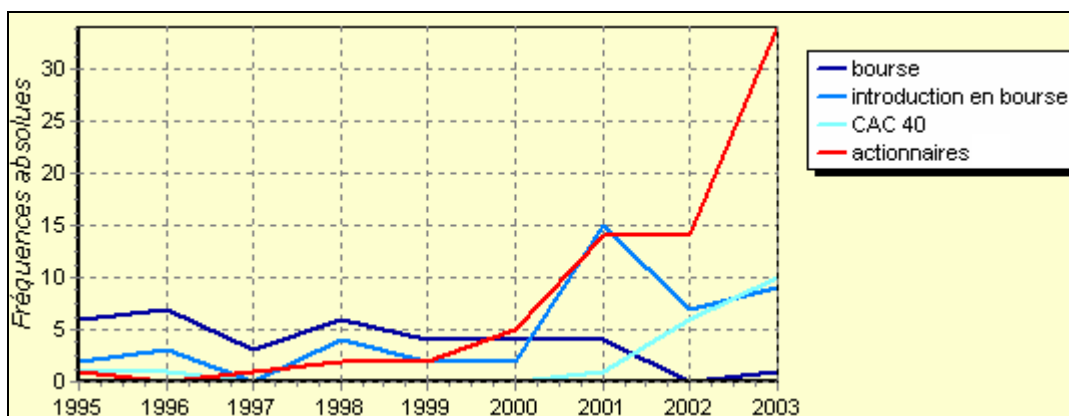


Figure 45.
Ventilation des occurrences des groupes de formes « bourse », « introduction en bourse », « CAC 40 » et « actionnaires » sur les neuf parties du corpus RA

Lecture de la figure 45 : quatre groupes de formes ont été constitués de manière à regrouper différentes variantes graphiques correspondant au même terme simple

¹ Source : Datastream.

ou complexe. Par exemple, le groupe « bourse » est composé des formes suivantes « Bourse », « BOURSE », « bourses », « Bourses » qui sont utilisées dans le corpus pour désigner des marchés réglementés d'instruments financiers, à l'exclusion de la forme « bourse » qui est majoritairement employée dans des syntagmes lexicalisés.

On observe que l'indice CAC 40 progresse sur la première période, atteint son sommet en 1999-2000 puis décline jusqu'en 2002 pour amorcer ensuite une reprise moins importante que sa précédente chute. Les quatre termes fluctuent légèrement durant la période allant de 1995 à 2000, montrant une relative insensibilité au développement de l'activité boursière qui semble jouer dans les rapports d'activité le rôle d'une toile de fond. En 2001, brusquement, les fréquences de deux de ces termes augmentent nettement : « actionnaires » et « introduction en bourse » voient en effet leur nombre d'occurrences tripler. Cet accroissement est lié à la restitution d'un événement majeur concernant seulement le Crédit agricole, la cotation en bourse de son organe central. La fréquence du terme « indice 40 » semble en léger décalage. C'est qu'en fait il est intégré comme élément de la communication financière relative au titre nouvellement coté (en décembre 2001) et qui apparaît dès 2002. Les rédacteurs du rapport ne cherchent pas à déplorer la baisse de l'indice boursier mais au contraire à louer la progression du titre en dépit de la mauvaise conjoncture boursière (on parle alors d'un cours qui a « surperformé » l'indice CAC 40).

Quant à la disparition du terme « bourse » en 2002, on peut être tenté de l'attribuer à l'effet de discrétion évoqué plus haut : la seule occurrence qui apparaît en 2003 concerne en fait une « Bourse d'Emploi en ligne ».

C. Variations mécaniques

Par ailleurs, les variations de fréquences peuvent enregistrer des phénomènes éloignés de toute explication thématique liée à la restitution de l'activité. On a vu dans le chapitre 3 que la distinction de deux types de contextes (contextes à dominante syntactique et non syntactique) reposait sur une sélection du vocabulaire. On peut voir aussi que cette sélection est elle-même soumise aux traitements différents qui sont imposés à chaque type de contexte dans les rapports d'activité. En

effet, les contextes non syntactiques proposent une information simplifiée, relativement récurrente et centrée sur les structures de l'organisation. Ces caractéristiques entraînent une plus grande sensibilité de ces contextes aux conséquences de certains remaniements formels. Ces derniers ne sont pas détachés de toute ramification thématique (on a vu que le traitement différent appliqué à la présentation des Caisses régionales de Crédit agricole était lié à l'apparition d'une nouvelle donne institutionnelle à l'intérieur du groupe bancaire), mais ils sont susceptibles de déclencher des modifications de fréquence sans commune mesure avec ce que l'on observe en général dans les contextes syntactiques. On peut illustrer ce point à l'aide d'un terme, « fonds propres » figurant dans les fiches signalétiques donnant les caractéristiques complètes des Caisses régionales jusqu'en 2000.

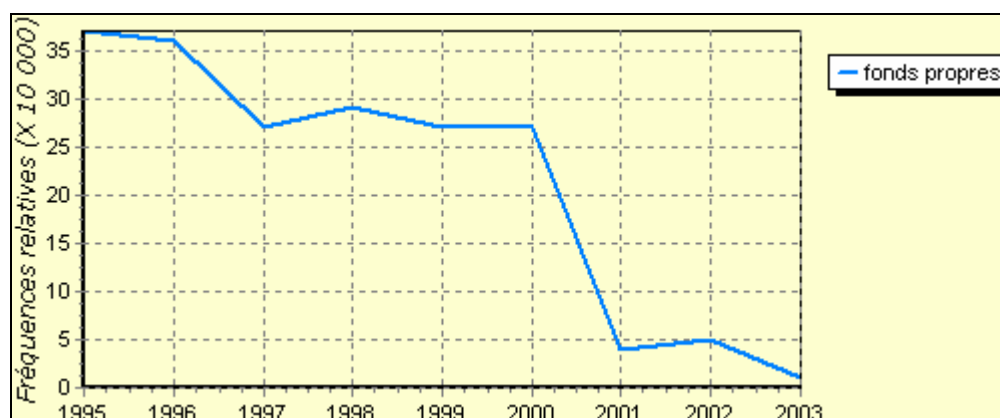


Figure 46.
Ventilation des occurrences de « fonds propres » dans le corpus RA

Lecture de la figure 46 : le groupe de formes « fonds propres » est composé des deux segments répétés « fonds propres » et « Fonds propres ». Le graphique représente les occurrences des deux segments données en fréquences relatives. Celles-ci, contrairement aux fréquences absolues qui correspondent aux données brutes du tableau lexical entier du corpus, sont relatives à la taille de chaque partie rapportée au corpus. De cette manière, les fréquences relatives permettent d'avoir une image de la ventilation des segments affranchie des variations qui pourraient être liées de façon mécanique à la taille du texte de chaque partie.

La chute de fréquence que l'on constate à partir de 2001 est directement liée à la modification d'un dispositif dans les contextes à dominante non syntactique (fiches signalétiques des Caisses régionales de Crédit agricole), dans lesquels l'expression « fonds propres » apparaissait de façon récurrente avec pour chaque partie plusieurs dizaines d'occurrences.

Les rapports d'activité peuvent également enregistrer mécaniquement d'autres phénomènes indépendants d'une manière particulière de caractériser l'activité bancaire. C'est le cas durant la période couverte par le corpus pour les unités monétaires. Le passage du franc à l'euro entraîne le remplacement des unités monétaires, mais aussi une diminution des montants liés à la valeur moindre du franc. On observe, par conséquent, une transformation des milliards en millions, ce phénomène répercutant la conversion d'une monnaie dans l'autre.

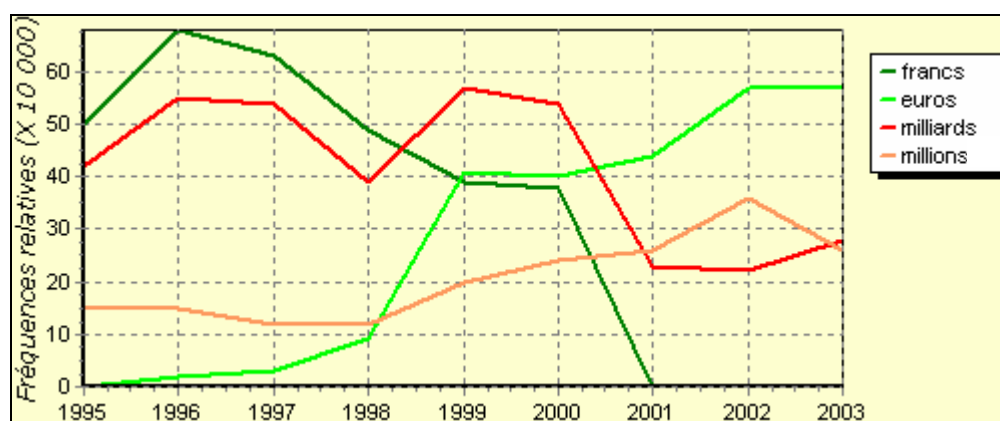


Figure 47.

Fréquences relatives de « francs », « euros », « milliards » et « millions » dans le corpus RA

II. VARIATIONS DE FRÉQUENCE ET VARIÉTÉS DE LANGUE

A. Répartition des usages du vocabulaire commun

Dès lors, on peut analyser comment les différentes variétés de langue identifiées dans le vocabulaire commun contribuent à la restitution de l'activité bancaire. Comme on l'a vu précédemment¹, les cinquante premières formes du vocabulaire commun peuvent être regroupées en quatre ensembles. La langue commune est la mieux représentée avec 19 mots², les trois autres groupes sont

¹ Voir chapitre 6.

² Il s'agit de « Directeur, Général, Président, activités, année, activité, bancaire, clients, croissance, clientèle, développement, ensemble, milliards, millions, offre, produits, place, progression, services ».

relativement équilibrés ; 11 mots pour la terminologie répertoriée dans le *LBMF*¹, 8 pour les noms propres ou parties de dénominations complexes² et 12 pour les mots dont l'usage du corpus révèle une sélection spécifique de leur aire sémantique³.

À partir des constats établis dans les précédents chapitres, on pourrait s'attendre à voir la langue commune favorisée au détriment de la terminologie bancaire et financière à partir de 2001. On aurait ainsi un indice d'une adaptation de son vocabulaire par le locuteur des rapports d'activité, compte tenu de l'élargissement de son public. On pourrait également s'attendre à une relative neutralité des noms propres. En revanche, le quatrième groupe, peut-être à cause de sa relative hétérogénéité, laisse les pronostics ouverts. Afin de vérifier ces points, on commentera les résultats de la ventilation de chaque variété sur les différentes parties du corpus.

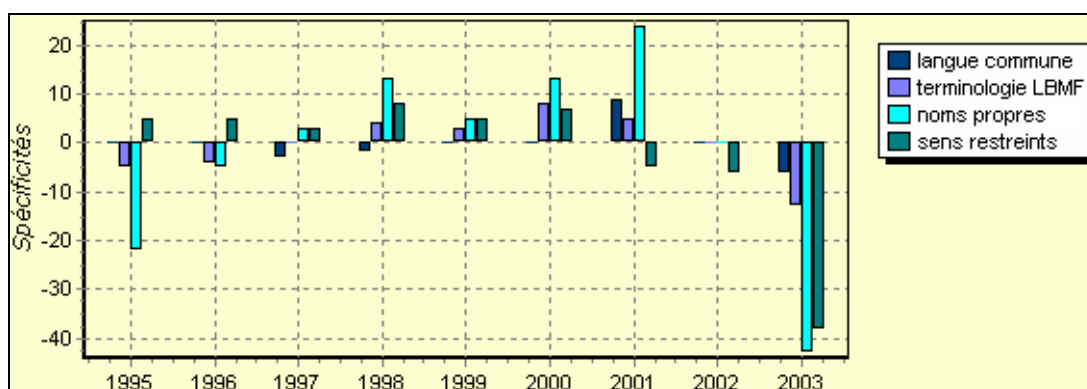


Figure 48.

Spécificités des quatre groupes de formes « langue commune », « terminologie LBMF », « noms propres » et « sens restreints » dans le corpus RA

Lecture de la figure 48 : quatre groupes de formes rassemblant des mots relevant de la langue commune, de la terminologie au sens du *LBMF*, correspondant à des noms propres ou à des mots dont l'usage dans le corpus restreint de manière récurrente leur aire sémantique, ont été constitués. Ils ont ensuite été projetés sur un graphique affichant pour chaque groupe de formes ses spécificités positives, négatives ou sa banalité dans chacune des parties du corpus RA. La méthode des

¹ Il s'agit de « Crédit, actifs, actions, banque, crédit, capital, encours, fonds, gestion, propres, épargne ».

² Il s'agit de « Agricole, Banque, Caisse, Europe, France, Indosuez, Jean, agricole ».

³ Il s'agit de « Caisses, Groupe, Total, assurance, bilan, entreprises, filiales, groupe, locales, marchés, marché, réseau ».

spécificités permet de porter un jugement sur la fréquence d'une forme ou, comme ici, d'un groupe de formes, compte tenu des paramètres suivants : occurrences des formes de chaque groupe dans le corpus, occurrences dans chacune des parties du corpus, taille de chaque partie du corpus et taille du corpus. L'axe des ordonnées fournit les valeurs positives ou négatives prises par le coefficient de spécificité calculé pour chaque groupe. Une spécificité positive indique un suremploi du groupe de formes, une spécificité négative, un sous emploi, un coefficient égal à zéro indique que les fréquences observées pour telle partie sont banales.

Les résultats du calcul des spécificités présentés dans le graphique semblent infirmer les hypothèses formulées plus haut concernant les utilisations probables des variétés de langue dans le corpus, compte tenu de ce qui avait déjà été vu par ailleurs. Hormis pour 2001, la langue commune ne semble pas avoir fait l'objet d'un usage particulièrement renforcé ; l'usage de la terminologie, au sens du *LBMF*, est déficitaire à chaque extrémité du corpus et connaît un excédent durant la période 1998-2001 ; les noms propres font l'objet des variations les plus importantes et s'alignent sur la terminologie pour ce qui est de la répartition des excédents et des sous-emplois ; enfin, les mots dont l'aire sémantique est restreinte dans le corpus sont les seuls à marquer la coupure attendue avant et après 2001. Avant de rejeter les hypothèses formulées plus haut, il convient d'examiner de plus près les éléments dont chaque groupe de formes se compose.

B. Une restitution commerciale de l'activité bancaire accentuée à partir de 2001

Une manière de démêler l'écheveau des fréquences dans chaque groupe consiste à observer les déficits signalés par le calcul des spécificités dans les parties du corpus sur lesquelles on veut procéder à certaines vérifications. Pour le groupe « langue commune », la partie du corpus pour laquelle le résultat du calcul des spécificités va le plus à l'encontre de l'hypothèse d'un renforcement de la langue commune est l'année 2003. On donne ci-dessous les principales spécificités négatives de la partie en question :

*Tableau 43.
Principaux coefficients de spécificité négative pour la partie 2003 du corpus RA*

Formes	Frq. Tot.	Frq. 2003	Coef. de sp.
agricole	1108	7	***
Indosuez	827	22	-49
Total	688	15	-44
Général	485	7	-36
propres	453	16	-24
locales	484	23	-21
bureaux	223	1	-20
bilan	394	15	-20
Caisses	1047	107	-16
M	212	3	-16
1999	155	2	-13
Carr	138	2	-11
nationale	121	2	-10
Président	724	80	-10
fonds	567	61	-9
Banco	167	7	-9
Crédit	2992	461	-9
2000	207	12	-9
clientèles	107	4	-7
Tél	449	48	-7
I	106	4	-7
milliards	828	113	-6
Cheuvreux	114	5	-6
Directeur	724	103	-5

Lecture du tableau 43 : pour chaque ligne, ce tableau donne une forme, sa fréquence totale dans le corpus RA, sa fréquence dans la partie 2003 et le coefficient de spécificité qui est associé à cette fréquence dans la partie 2003. Plus le coefficient de spécificité a une valeur absolue élevée, plus le déficit d'occurrences de la forme est important dans cette partie, compte tenu du fait qu'il s'agit ici de coefficients de spécificité négatifs. Les trois astérisques donnés pour la première forme indiquent un coefficient dont la valeur absolue est supérieure à 50.

À la lecture du tableau, il apparaît que quatre formes au moins appartenant au groupe « langue commune » font partie des plus forts déficits pour la partie 2003. Il s'agit de « Directeur », « Général », « Président » et « milliards ». Pour ce qui est du dernier vocable, on a vu plus haut que sa baisse de fréquence doit être imputée à la conversion des résultats du groupe bancaire en euros. Les trois autres formes ressortissent au phénomène évoqué au sujet de « fonds propres », c'est-à-dire une recomposition des contextes non syntactiques entraînant une chute brutale de fréquence pour certains vocables qui y étaient regroupés en grand nombre. En 2002,

la présentation des noms des dirigeants des Caisses régionales et de leurs titres (Président et Directeur Général) n'a pas été reprise dans le descriptif des banques affiliées ; en 2003, ces titres sont repris mais la majuscule de « Général » est tombée, ce qui perpétue le déficit. Lorsque l'on réorganise le groupe de formes langue commune en deux sous ensembles, l'un réservé aux déficits (les quatre vocables dont il vient d'être question) et l'autre aux excédents (les quinze vocables restants de la liste initiale), on obtient des spécificités différentes :

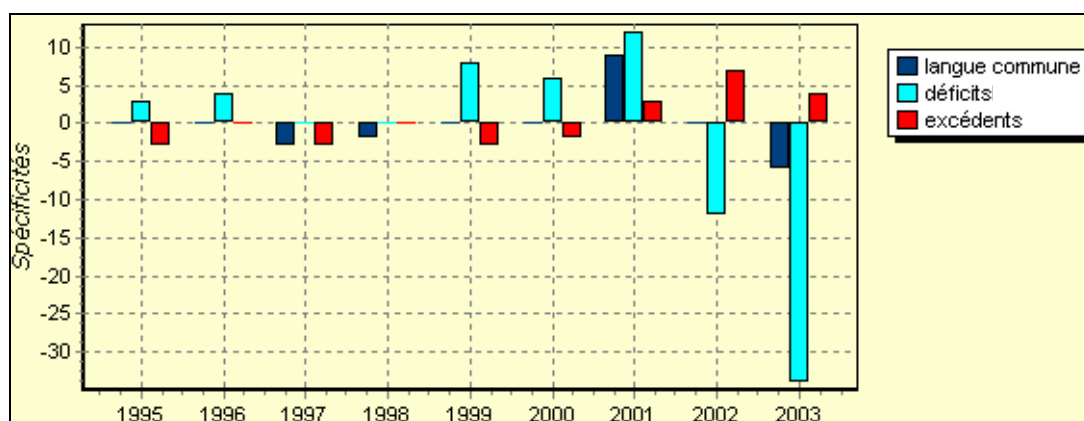


Figure 49.

Spécificités du groupe de forme « langue commune » et des deux sous ensembles « déficits » et « excédents » dans le corpus RA

Lecture de la figure 49 : ce graphique met en évidence les spécificités de trois groupes de formes. Il s'agit du groupe « langue commune » et de deux sous-ensembles reprenant pour l'un (« déficits »), les formes en fort déficit sur l'année 2003, et pour l'autre (« excédents »), les quinze autres formes du groupe initial, qu'elles soient en excédent ou qu'elles aient une fréquence banale dans la partie.

Ce graphique éclaire différemment les spécificités du groupe « langue commune ». En effet, on constate que lorsque l'on retire du groupe les formes déficitaires en 2003 pour des raisons particulières (modifications intervenues dans les contextes non syntactiques, conversion des francs en euros), un ensemble « langue commune » se détache comme étant caractéristique de la période 2001-2003. Le groupe de formes « langue commune » ramené de 19 à 15 formes se présente comme un ensemble assez cohérent dont la thématique commerciale ne laisse pas de doute : « millions, année, développement, activité, clients, offre, produits, services, clientèle, croissance, place, progression, activités, bancaire, ensemble ». Ce résultat est d'autant plus éloquent que le groupe ne rassemble que

des vocables appartenant au vocabulaire commun. Contrairement à ce qui a été affirmé plus haut, il semble confirmer que la période s'ouvrant en 2001 se caractérise bien par l'utilisation plus fréquente de la langue commune afin de présenter les résultats du groupe bancaire dans leurs aspects généraux les plus facilement accessibles pour un public élargi. On dispose ainsi d'un indice étayant l'hypothèse selon laquelle le vocabulaire du genre aurait subi une inflexion destinée à prendre en compte l'élargissement du public des rapports d'activité.

C. Terminologie bancaire et valorisation de certains domaines d'activité

Lorsque l'on applique la même démarche aux trois autres groupes, on obtient des résultats différents des premiers. Pour ce qui est de la terminologie au sens du *LBMF*, si l'on retire « Crédit », « fonds » et « propres » pour des raisons similaires à celles qui ont conduit à alléger le groupe « langue commune » de quatre de ses vocables, les spécificités du groupe terminologique sont nettement modifiées. Les spécificités négatives n'apparaissent plus qu'en début de période, et un excédent est signalé pour 2001 et 2002 puis une banalisation en 2003. Parmi les vocables restants (« actifs, actions, banque, crédit, capital, encours, gestion, épargne »), on distingue des situations contrastées mais éclairantes. Si « encours¹ » semble jouer un rôle relativement neutre, les autres vocables apparaissent comme en prise avec l'actualité de l'activité bancaire. En tenant compte des enseignements du chapitre 6, on constate un excédent de « banque » pour la période 1999-2001 qui correspond au réagencement de l'activité du groupe en grandes familles d'activités bancaires (banque de proximité, d'investissement, de grandes clientèles), en même temps que l'accent est mis sur la banque privée.

Le vocable « capital » a un excédent en 2000 auquel correspond un déficit pour « épargne » : la bulle financière n'a pas encore éclaté et les prises de participation dans le capital de diverses entreprises bat son plein alors que l'épargne bancaire classique est éclipsée par les achats d'actions ou de parts de Sicav.

¹ Sur ce vocable, voir *infra*.

Les effets de discrétion et de mise en valeur fonctionnent de manière assez prévisible. Les excédents réels pour la période se concentrent sur quatre vocables : « gestion », « actifs », « crédit » et « actions ». Pour « actions », l'explication est institutionnelle : il s'agit principalement de la communication financière autour des actions cotées de Crédit Agricole s.a. et du rappel des interventions du groupe bancaire relevant du développement durable (« actions de développement local », « actions de sensibilisation », « actions en faveur de... »). Quant à « gestion » et « actifs », ils forment le syntagme lexicalisé « gestion d'actifs » (*asset management*) qui sert à désigner les activités de gestion de portefeuille. Comme pour la banque privée, il s'agit d'un secteur dont la croissance est mise en valeur dans les rapports d'activité. La même explication tient pour les secteurs du crédit-bail et du crédit à la consommation, pour la désignation desquels le vocable « crédit » est sollicité.

On a donc pour la plupart de ces vocables appartenant au groupe « terminologie *LBMF* » la confirmation d'une certaine corrélation entre les variations de fréquences et l'augmentation ou la diminution de l'importance accordée à certains aspects de l'activité bancaire dans les rapports d'activité.

D. Les noms propres, une prise indirecte sur le référentiel du groupe bancaire

Le groupe des noms propres, que l'on avait crédité a priori d'une relative neutralité, est celui qui accuse les plus forts contrastes. Afin d'expliquer ces importantes variations de fréquence, on propose de décomposer le groupe des « noms propres » en deux sous-ensembles. Le premier regroupe les vocables « Caisse », « Indosuez », « agricole » et « Agricole ». On a rappelé également dans le graphique les spécificités du groupe pris dans son ensemble.

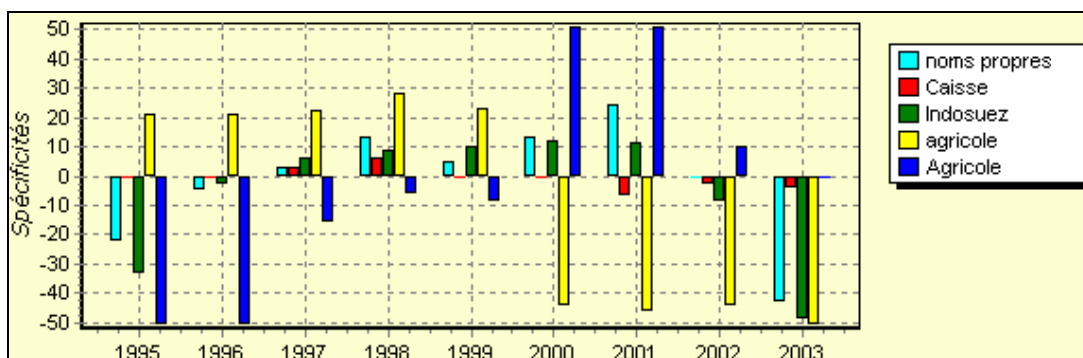


Figure 50.
Spécificités d'un sous-ensemble de formes du groupe « noms propres » dans le corpus RA

On voit sur ce graphique que l'allure générale des spécificités du groupe « noms propres » est commandée par les spécificités de trois formes : « agricole », « Agricole » et « Indosuez ». Pour les deux premiers vocables, il s'agit de la modification de la graphie de « agricole » dans « Crédit agricole » : à partir de 2000, le « a » de « agricole » porte une majuscule. On a vu dans le chapitre 4 que l'on peut considérer ce changement de graphie comme un signe annonciateur des modifications importantes que le groupe va connaître en 2001, avec la cotation de son organe central.

De son côté, le vocable « Indosuez » entre dans la dénomination « Banque Indosuez », remplacée à partir de 1997 par « Crédit Agricole Indosuez ». Le fait que ce nom ne se banalise pas durant une assez longue période courant de 1997 à 2001 s'explique par un phénomène relevant de la signalétique des dénominations propres¹. En effet, la dénomination propre est réutilisée comme tête pour former de nouvelles dénominations attribuées aux filiales de la banque de financement et d'investissement (« Crédit Agricole Indosuez Cheuvreux », « Crédit Agricole Indosuez Lazard Financial Products », etc.). La baisse brutale constatée en 2003 correspond à l'abandon de cette dénomination, remplacée à son tour par « Calyon » à la suite du rapprochement de la banque de financement et d'investissement du Crédit Lyonnais et de Crédit Agricole Indosuez en 2003. Enfin, la variation affectant

¹ Ce point est abordé en détail dans les chapitres suivants.

« Caisse » est liée, elle aussi, à un changement de dénomination lié à la cotation, « Caisse nationale de Crédit agricole » devenant à partir de 2001 « Crédit Agricole s.a. ».

On a regroupé dans le graphique ci-dessous les quatre autres vocables appartenant au groupe « noms propres », dont les spécificités ont été rappelées.

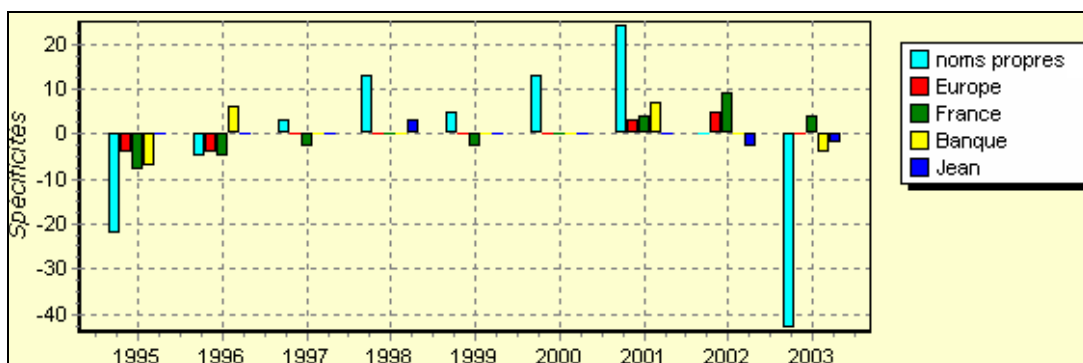


Figure 51.
Spécificités d'un sous-ensemble de formes du groupe « noms propres » dans le corpus RA

On peut voir que ce groupe de formes s'inscrit de façon plus discrète dans le mouvement général caractérisant les spécificités du groupe de formes « noms propres » pris dans leur ensemble. Les variations affectant le prénom « Jean » sont directement liées aux modifications touchant certains contextes non syntactiques. L'excédent de 1998 est à attribuer au fait que, cette partie étant plus petite que celles qui l'encadrent sous le rapport des contextes syntactiques, une fréquence équivalente attribuée à une forme apparaissant majoritairement dans des contextes non syntactiques s'en trouve valorisée. Pour le vocable « Banque », deux excédents attirent l'attention. Ils ne correspondent pas au même phénomène. En 1996, l'excédent est lié à la présentation de la Banque Indosuez récemment acquise. En 2001, il s'agit de la réorganisation de l'activité bancaire du groupe, déjà entamée à la fin des années 1990. Celle-ci est en quelque sorte réaffirmée – rendue plus visible pour un public nouveau? –, par une évolution du statut sémantique des dénominations des grands pôles de l'activité bancaire qui accèdent au statut de dénominations propres. Le vocable « Banque » y est toujours utilisé comme tête, mais porte maintenant la majuscule : « Banque de détail à l'étranger », « Banque de

financement », « Banque de grande clientèle », « Banque de proximité en France », « Banque privée », etc. Ce dispositif se banalise dans les années suivantes. Enfin, les augmentations de fréquence de « France » est « Europe » semblent liées. En effet, durant la dizaine d'années couverte par le corpus, le groupe bancaire, aux assises encore majoritairement nationales à la fin des années 1980, a pris une dimension internationale incontestable. Or, comme on l'a déjà noté plus haut au sujet de la « banque de proximité », la mise en valeur de ce développement s'accompagne d'une compensation destinée à valoriser également ce qui autrefois allait de soi ; d'où la montée parallèle de ces deux fréquences de noms propres de lieux associés étroitement à l'évolution du référentiel du groupe. On voit, ainsi, que les variations de fréquence des noms propres, à condition d'être analysées finement, renseignent de manière indirecte mais assez fiable sur les modifications qui affectent le référentiel du groupe bancaire tel qu'il est véhiculé par les rapports d'activité.

E. Usages sémantiques restreints

Le quatrième groupe de formes rassemble des vocables correspondant à des mots relevant de la langue commune, mais dont l'usage qui en est fait dans le corpus sélectionne de manière récurrente une partie seulement de leur aire sémantique. Les variations de fréquence de ces vocables dans les différentes parties du corpus peuvent s'expliquer dans la plupart des cas par des phénomènes déjà évoqués précédemment. Ainsi, les variations du groupe « Caisses », « locales », « bilan » et « Total » s'expliquent-elles par la disparition des informations relatives aux « Caisses locales » et au « bilan » des caisses régionales de Crédit agricole à partir de 2001.

De même, la répartition entre « Groupe » et « groupe » illustre-t-elle le bouleversement intervenu dans le groupe bancaire avec la cotation de Crédit Agricole s.a. L'usage ayant en quelque sorte réservé le vocable « Groupe » à l'individuation par catachrèse de « groupe Crédit agricole », le groupe constitué par les filiales rassemblées autour de Crédit Agricole s.a. ne bénéficie que timidement du même traitement dans les textes du corpus. Les vocables « marché », « marchés »

et « entreprises » ont des fréquences qui fluctuent peu. Sans être tout à fait banales, ces formes semblent d'abord au service de syntagmes lexicalisés¹, ou de rôles différents (entreprises clientes ou entreprises appartenant au groupe bancaire ?), dont on peut supposer que les variations de fréquence s'annulent.

Enfin, « assurance », « réseau » et « filiales » sont en excédent en 2001 et 2002. Les filiales, composantes à part entière du groupe Crédit Agricole s.a. voient leur rôle valorisé à partir de 2001, tout comme le secteur en développement de l'assurance-vie. Quant à « réseau », il est utilisé à partir du rapport 2001 pour valoriser le « premier réseau bancaire français » et présenter les caisses régionales comme relevant de la « banque de proximité ». La mise en avant du réseau est à mettre en relation avec, d'une part, le fait que Crédit Agricole s.a. détient 25% du capital de son premier actionnaire, et d'autre part, l'appauvrissement des renseignements signalétiques relatifs aux caisses régionales : à partir de 2001, les rapports d'activité du Crédit agricole sont d'abord ceux du groupe Crédit Agricole s.a.

Le quatrième groupe de vocables possède donc des propriétés proches à la fois de la langue commune (unités appartenant aux contextes non syntactiques) et de l'ensemble termes-noms propres, qui entretient des relations plus ou moins directes avec la restitution de la dynamique de l'activité du groupe bancaire dans les textes du corpus.

En résumé, les variations de fréquence affectent indifféremment tous les groupes de vocables du vocabulaire commun. L'interprétation de ces variations doit toutefois être soumise à un examen minutieux, au cours duquel on doit prendre soin de distinguer ce qui relève de mécanismes de mise en valeur ou de dissimulation, et ce qui doit être attribué à des variations mécaniquement enregistrées par les textes du corpus. Ainsi, les remaniements formels affectant les contextes non syntactiques engendrent-ils des fluctuations directement liées à l'utilisation de cadres répétitifs mobilisant de nombreuses occurrences. De même, un changement de graphie, qui entretient toujours un lien au moins indirect avec une évolution thématique, ne peut-

¹ Voir chapitre 6.

il pas être mis sur le même plan qu'une variation de fréquence touchant une forme à la graphie inchangée.

Lorsque cet examen préalable a été réalisé, il apparaît que chaque groupe est susceptible d'apporter un éclairage spécifique dans l'explication des mécanismes de sélection du vocabulaire des rapports d'activité. Les noms propres révèlent de façon indirecte, à travers des disparitions ou des changements de dénominations, l'utilisation de la majuscule pour certains syntagmes lexicalisés, et enfin, à travers la présence de toponymes largement répandus, des modifications intervenant dans le référentiel du groupe bancaire. De même, les termes donnent-ils accès aux mouvements touchant l'activité à travers la valorisation de certaines caractéristiques de l'institution ou en se faisant l'écho du développement de certains secteurs d'activité.

Enfin, on a pu identifier un renforcement du recours à la langue commune dans le but de promouvoir une restitution commerciale de l'activité bancaire traduisant le souci du locuteur collectif des rapports d'activité de s'adapter à un public élargi. Cette intervention relativement discrète est-elle la seule que ce locuteur ait envisagée pour s'adapter à ses nouveaux lecteurs ? Comment cette utilisation renforcée de certains vocables appartenant à la langue commune s'articule-t-elle avec d'autres variétés de langue dont les unités sont peut-être moins accessibles ?

III. LES RAPPORTS D'ACTIVITÉ CONSTITUENT-ILS UN GENRE DE DISCOURS SECOND ?

A. Reformulation et discours de vulgarisation

Dans le chapitre 2, on a esquissé un rapprochement entre discours de communication institutionnelle et discours de vulgarisation, car dans les deux cas il s'agit de produire un discours véhiculant des connaissances en direction d'un public de non-initiés. Si les finalités poursuivies sont naturellement différentes (communication économiquement finalisée dans le premier cas, et accroissement des connaissances du lecteur dans le second), il restait à établir les répercussions de cette différence de visées perlocutoires sur le vocabulaire. Avec la liste des dix termes

attestés dans le *LBMF*, présents dans le corpus, mais absents de deux importants dictionnaires de langue, on dispose d'un angle d'attaque pour repérer les modalités selon lesquelles l'utilisation des termes a pu faire l'objet d'une prise en charge énonciative dans les rapports d'activité.

Dans le cadre de cette exploration, il ne s'agit pas d'envisager le mécanisme de la reformulation sous tous ses aspects, comme l'a fait M. C. Conceição pour un corpus composé d'un ouvrage de pharmacologie¹. Parmi les indices, et les marqueurs de reformulation, on s'en tiendra aux plus caractéristiques. Les dix termes isolés dans le chapitre 6 serviront de fil conducteur pour la recherche d'éventuels reformulants. Ceux-ci comprennent les syntagmes et la phraséologie utilisés pour expliciter des termes qui risquent de ne pas être compris par les destinataires. Daniel Jacobi définit ainsi le cadre de ce type d'investigation : « *Dans un texte de vulgarisation, on peut donc rechercher au moins trois catégories de mécanismes de reformulation : ceux qui relèvent du paradigme désignationnel, par opposition à ceux qui s'inscrivent dans le paradigme définitionnel. Enfin, une troisième famille peut être distinguée à partir de l'axe dit métaphorique.*² »

M.-F. Mortureux propose quant à elle de distinguer plusieurs moyens linguistiques sollicités par la mise en place de tels paradigmes dans les textes.

En effet, le paradigme désignationnel est ainsi défini : « *Il s'agit, en première approximation, de listes de syntagmes (en général nominaux, parfois verbaux) fonctionnant en coréférence avec un vocable initial dans un discours donné*³. » Parmi les reformulants, le paradigme définitionnel permet de faire la distinction – jugée secondaire par M.-F. Mortureux –, « *entre syntagmes à valeur désignative et périphrases définitionnelles (paraphrases)*⁴. » Le repérage des reformulants doit être conduit en tenant compte des procédés disponibles qui sont susceptibles d'avoir été utilisés par l'énonciateur dans ses discours.

M.-F. Mortureux distingue quatre grands types de procédés. Il s'agit, tout d'abord, des opérations métalinguistiques reposant sur l'usage de substantifs –

¹ M. C. Conceição (2005).

² D. Jacobi (1999 : 144).

³ M.-F. Mortureux (1993 : 124).

⁴ *Ibidem* p. 124 et 131.

« mot », « terme », etc. –, et de verbes – « signifier », « désigner », « nommer », etc. –, utilisés pour parler des mots considérés en eux-mêmes. On note que ces équivalences sont en général livrées à l'aide de procédures moins visibles s'appuyant, par exemple, sur le verbe être, la coordination ou la simple juxtaposition. Une opération de reformulation peut également être signalée par un décrochage typographique recourant aux guillemets, aux parenthèses, ou par l'utilisation de caractères en gras ou en italiques. L'anaphore et la cataphore, en construisant la topicalisation du texte, sont également susceptibles d'établir des liens entre reformulé et reformulant. Enfin, M.-F. Mortureux indique que la paraphrase *in absentia* peut également créer ce type de relation mais avec une plus longue portée et en opérant une substitution complète du reformulé. Dans ce cas, la relation entre reformulant et reformulé peut être établie à partir de l'examen des équivalences distributionnelles identifiées dans le texte. Cependant, outre la difficulté prévisible du repérage – signalée par l'auteur –, l'absence de signalement de la relation de reformulation rend ce cas de figure plus difficile à identifier. On tentera néanmoins d'y apporter une réponse en abordant la question de l'anaphore nominale.

D'une manière générale, il n'y a pas de critère certain garantissant que l'on a affaire à une reformulation susceptible d'être intégrée au paradigme. Il s'agit plutôt d'indices qui s'échelonnent sur un *continuum*. M.-F. Mortureux convient que le repérage d'un paradigme est minutieux et qu'il doit reposer sur une observation attentive des reformulations. Il convient donc de distinguer parmi les reformulants, ceux qui tirent leur origine de relations inscrites en langue (relations lexicales entre synonymes/antonymes, hypo/hyperonymes et relations métonymiques), et ceux qui sont issus de relations non codées liées au cadre énonciatif (métaphores, « qualifications » et emprunts à d'autres systèmes conceptuels également présents dans le discours).

Comme on l'a indiqué plus haut, ce sont les termes absents des nomenclatures des deux dictionnaires de langue utilisés dans la section précédente qui seront utilisés comme fil conducteur. Il s'agit de « moyen de paiement », « arrangeur », « encours [bancaires] », « futures », « OPCVM », « organe central »,

« PEA », « gestion alternative », « trading » et « LBO ». Ces termes ont été utilisés pour vérifier la présence d'opérations de reformulation dans l'ensemble du corpus¹.

B. Les indices typographiques de la reformulation, ou la banalité du LBO

Les indications fournies par la typographie ayant disparu lors de la transposition des fichiers d'origine au format texte, un format intermédiaire a été utilisé pour le repérage de l'italique et des caractères gras (il s'agit du format Rich Text Format). Aucun des termes de la liste n'est signalé de cette manière à l'aide de l'italique. L'exploration du corpus a montré qu'il en est ainsi pour les autres termes présentés dans les sections précédentes, l'italique étant réservé en général à la mise en exergue de blocs de textes introduisant les rubriques. Les caractères gras, destinés à renforcer la présence visuelle de la police, répondent aux mêmes usages que l'italique et ne sont donc pas plus utilisés pour mettre en exergue les dix termes du test que les autres mots. Les intertitres constituent une exception, mais, comme pour les paragraphes introductifs, les caractères gras ne servent pas dans ce cas à isoler un mot en particulier.

Deux sortes de signes typographiques pouvant signaler l'usage d'un terme ou de l'un de ses reformulants ont été étudiés ici. Il s'agit des guillemets et des parenthèses. Pour ce qui est des guillemets, on recense dans le corpus 159 mots simples et expressions, lexicalisées ou non, encadrés par une paire de guillemets, pour 203 occurrences². Les principaux usages constatés sont classiques et se répartissent de cette manière : dans les deux tiers des cas (111 mots ou expressions), les guillemets servent à citer la dénomination exacte d'une entité. On trouve, par exemple : « *Partenaire du Musée Guimet depuis 1990, CAI a parrainé l'exposition " Rituels tibétains " en 2002 - 2003.* » (2003) ; ou encore : « *Le Crédit agricole s'est notamment associé, au côté de la Fédération française de football, au programme "*

¹ Les différentes graphies possibles ont à chaque fois été sollicitées. On en donne ici la liste : « arrangeur » (29 occ.), « arrangeurs » (11 occ.), « encours » (348 occ.), « Encours » (13 occ.), « futures » (15 occ.), « Futures » (70 occ.), « moyens de paiement » (25 occ.), « Moyens de paiement » (14 occ.), MOYENS DE PAIEMENTS (2 occ.), « moyens de paiements » (3 occ.), « OPCVM » (78 occ.), « organe central » (20 occ.), « Organe » (9 occ.), « PEA » (25 occ.), « trading » (23 occ.), « Trading » (1 occ.), « LBO » (21 occ.), « gestion alternative » (23 occ.).

² Cette liste est donnée en annexe.

T'es jeune, t'es foot " (...) » (1997). Comme ces deux exemples le suggèrent, les entités ainsi dénommées sont très diverses : il peut s'agir de banques, de marques commerciales, de dispositifs divers, de distinctions internationales, d'opérations de communication, de sites, d'ouvrages, etc. On doit noter leur faible fréquence, la majorité de ces dénominations correspondant soit à des hapax, soit à des fréquences faibles de deux ou trois occurrences. On en donne la liste complète en annexe.

Un tiers des mots et expressions (ils sont au nombre de 49), semblent avoir été mis entre guillemets afin de marquer une prise de distance de l'énonciateur, soit qu'il veuille signaler ainsi la reprise d'un terme relevant de la terminologie bancaire – souvent anglo-saxonne –, soit qu'il souhaite signifier qu'il ne reprend pas l'expression à son compte. C'est ainsi que l'on trouve dans le corpus : « *Parmi les réalisations de l'année 1995, figurent aussi la participation d'Unicredit en tant que "lead manager" (...)* » (1995) ; ou bien encore : « *Pour aborder la "révolution Internet", avec volonté et pragmatisme (...)* » (1999). D'une manière générale, les guillemets ne sont pas utilisés pour marquer une opération de reformulation. La seule occurrence identifiée pour l'un des termes du test (« futures »), correspond à une prise de distance sans apport cognitif.

On recense dans le corpus 1762 cas d'utilisation de parenthèses¹. Un peu plus de 40% de ce total (751 occurrences), concerne des données chiffrées venant préciser les propos tenus dans les textes. Ces éléments ont été éliminés de la liste de ce que l'on considère ici comme des expressions entre parenthèses. Celles-ci représentent donc un millier d'occurrences pour 671 unités simples ou complexes. On sait que les parenthèses marquent l'insertion d'un élément détaché et isolé par rapport à la phrase. Cet élément a généralement une importance secondaire, les parenthèses servant à insérer des réflexions incidentes, des commentaires, des rectifications, etc. La *Grammaire méthodique du français* signale, en outre, que « *dans un emploi spécialisé, les parenthèses peuvent aussi encadrer une référence (titre, page, indication d'un passage...) ou toute indication qui équivaut à une note*

¹ Le repérage des expressions entre parenthèses s'est fait à l'aide de l'expression régulière '(.*)' recherchant toute suite d'un nombre quelconque de caractères compris entre deux parenthèses. Une macro-commande a ensuite permis l'extraction automatisée de la liste complète.

*en bas de page ou en fin de volume*¹. » C'est ce dernier usage que l'on rencontre le plus fréquemment dans le corpus de rapports d'activité.

Trois types de relations entre termes et parenthèses ont été envisagés : la notion que recouvre le terme a été précisée par une périphrase mise entre parenthèses ; le terme lui-même a été mis entre parenthèses et c'est sa périphrase définitionnelle qui appartient au texte de base ; enfin, le terme n'apparaît dans aucune de ces deux situations². Cette dernière possibilité se rencontre deux fois pour les termes « organe central » et « gestion alternative ». Le cas de figure où le terme est défini à l'aide d'une paraphrase encadrée par des parenthèses ne se rencontre pas. Il reste donc le cas où le terme apparaît entre parenthèses. Cependant, que les parenthèses soient utilisées pour donner un sigle ou son développé, des précisions sur le domaine de rattachement d'une entité, ou des exemples, elles ne servent jamais à préciser le sens des notions désignées par les termes (ou bien de manière indirecte et imprécise).

Au contraire, les parenthèses semblent plutôt destinées à ancrer le discours des rapports d'activité dans une réalité économique-bancaire qui est supposée connue. Le terme « LBO » en fournit un bon exemple avec ces contextes typiques :

C1 : « *Dans le financement d'acquisitions (LBO), (...)* » (2001)

C2 : « *(...) les financements de rachat d'entreprises à effet de levier (LBO) (...)* » (2003)

C3 : « *En 2003, les opérations à effet de levier (LBO) de grande ampleur (...)* » (2003)

¹ M. Riegel, J.-C. Pellat, R. Rioul (2004 : 96).

² L'exploration des contextes autour d'un terme a été effectuée de cette manière : pour le contexte avant la parenthèse, l'expression régulière '^.* (*TERME' recherche et sélectionne (affiche en surbrillance), tout ce qui est compris entre le début d'une ligne et la parenthèse comprenant le terme puis tout ce qui est compris entre la parenthèse ouvrante et le terme ; l'expression est répétée en changeant ce qui doit l'être pour explorer le contexte droit jusqu'à la fin d'une ligne. Le terme fait lui-même l'objet d'un traitement pour tenir compte des variantes. Par exemple, l'expression '^.* (*[Mm]oyen. de paiement.' permet de retrouver tous les contextes précédant l'apparition d'une parenthèse ouvrante derrière laquelle figure le terme « moyens de paiement » avec les variantes attestées suivantes : « moyens de paiement » (25 occ.), « moyens de paiements » (3 occ.) et « Moyens de paiement » (14 occ.). L'intérêt de cette approche par rapport aux concordances traditionnelles (terme pivot et contexte gauche et droit calibrés pour toutes les occurrences de la même manière), est de permettre une saisie du contexte adaptée à l'objet précis de la recherche et aux configurations exactes dans lesquelles le terme apparaît. On donne en annexe un exemple pour « moyens de paiement ».

C4 : « (...) *une opération de transmission (LBO) dans Fraikin (...)* » (2003)

C5 : « (...) *les segments de marché du capital-risque, capital-développement, transmission (LBO) (...)* » (2003).

Le *LBMF* qui a été retenu pour servir de référence terminologique, donne cette définition du LBO : « **Leverage Buy Out** : *terme utilisé pour toutes les opérations d'acquisitions du capital d'une société qui comportent un effet de levier par l'endettement des acquéreurs, le remboursement de la dette contractée par les acquéreurs s'effectuant par les bénéfices tirés de la société achetée.* » Compte tenu de la diversité des définisseurs possibles proposés dans les exemples (financement d'acquisition, rachat d'entreprise, transmission), il ressort de la comparaison entre les contextes d'apparition de « LBO » et la définition du terme, que les parenthèses servent plutôt à préciser des périphrases approximatives par le recours au terme exact. L'apparition de « LBO » dans les parenthèses pouvant se paraphraser à son tour, dans l'exemple C1 « Dans le financement d'acquisitions réalisées sous la forme de LBO », ou dans l'exemple C4 « (...) transmission réalisée par LBO ». Autrement dit, le terme de « LBO » est supposé connu et ne fait l'objet d'aucune explicitation, contrairement aux apparences. On est d'autant plus enclin à admettre cette interprétation que le corpus donne par ailleurs de nombreux exemples dans lesquels les termes sont utilisés de façon elliptique pour ramener à plus de précision un discours qui semble pour ainsi dire, vulgariser malgré lui.

C. Indices métalinguistiques de la reformulation

Si les marques typographiques étudiées¹ ne sont pas utilisées pour expliciter certains termes, on peut se demander ce qu'il en est pour d'autres moyens métalinguistiques susceptibles de jouer le même rôle. Dès lors, dans cette sous-section on doit rechercher si les rapports d'activité recèlent des traces

¹ D'autres signes conventionnels sont présents dans le corpus, comme l'astérisque (un peu plus de 70 occurrences, compte tenu du fait que ce signe fonctionne par doublon, l'appel et la note de renvoi) ou les crochets (52 occurrences). Dans le premier cas, l'usage correspond à celui de la note de bas de page, dans le second à la mention de la source émettrice d'une information citée. Ces deux utilisations sont identiques à celles prises en charge par les parenthèses.

métalinguistiques de prise en charge énonciative de la signification des termes. Comme les traces laissées par ces opérations s'échelonnent sur un *continuum*, on suivra pour leur présentation une progression allant du plus au moins visible. Du point de vue de la méthode d'exploration adoptée, il s'agit, comme pour le développement précédent, d'un repérage visuel effectué à la surface du texte. Plusieurs techniques sont utilisées : les expressions régulières, afin de repérer directement dans le texte des contextes particuliers ; les concordances, qui permettent d'afficher le cotexte de part et d'autre d'une forme, ou d'un segment répété, choisis pour pivot ; la carte des sections, qui restitue le texte du corpus sous la forme d'une carte composée de carrés correspondant chacun à une unité textuelle délimitée par paramétrage, et qui permet l'affichage des occurrences d'une ou de deux formes ou groupes de formes.

J. Rey-Debove indique que de ce point de vue « *le type le plus intéressant de noms métalinguistiques à forte densité est le nom d'unité, comme **mot, expression, phrase, syllabe, lettre, etc.**¹ » On a donc recherché les mots qui paraissaient les plus susceptibles d'être utilisés compte tenu des thèmes abordés dans le corpus : « mot » (est absent du corpus), « terme » (134 occurrences au singulier et 23 au pluriel, toutes les occurrences renvoyant non pas au « terme » – sorte de mot –, mais au « terme » –, synonyme d'échéance), « expression » (9 occurrences désignant toutes le fait d'exprimer par le langage et non la suite de mots qui peut résulter de cette activité), et « nom ». Ce dernier est présent avec certains de ses dérivés : « nom » (7 occurrences), « nommé » (7 occurrences), « dénomination » (5 occurrences) et « dénommées » (4 occurrences). On ajoutera à la liste le participe passé employé comme adjectif « baptisé » (5 occurrences) et « baptisée(s) » (3 occurrences). Ces deux groupes semblent être les seuls à être utilisés à des fins métalinguistiques dans le corpus. La majorité des occurrences jouant un rôle métalinguistique (on en dénombre 20 sur 24), sont associées à des noms propres dans une situation bien particulière : celle d'un changement de nom commercial (nom propre). Les 4 occurrences restantes sont utilisées à chaque fois dans le même contexte pour*

¹ J. Rey-Debove (1986 – 1997 : 35). Les verbes comme « appeler » dans sa forme pronominale, ou « signifier » sont absents du corpus.

présenter un terme : « [les] sociétés de crédit agricole, dénommées Caisses locales (...) ». On apprend donc quelque chose sur l'entité désignée par le nom propre ou le terme, mais rien sur le concept désigné par les unités terminologiques.

Les expressions « est un/une », « est le/la/l' » et « sont des/les », présentes dans le corpus, sont susceptibles de mettre en relation un terme et un reformulant ou une périphrase. On recherche donc à l'aide de la carte des sections, tous les contextes dans lesquels apparaissent simultanément l'une de ces expressions et l'un des termes du test (sous ses différentes graphies). On donne ci-dessous un aperçu de la carte obtenue :

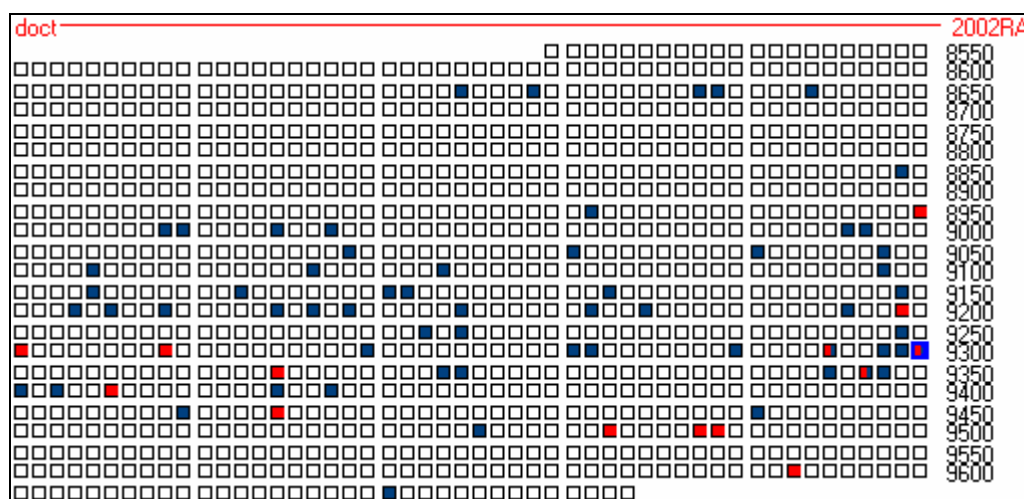


Figure 52.

Projection des groupes de formes « EST UN » et « TERMES TEST » sur la carte des sections correspondant aux phrases de la partie 2002 du corpus

Lecture de la figure 52 : les sections de la carte (carrés) correspondent globalement aux phrases de la partie 2002 du corpus (les phrases étant délimitées à partir de la présence d'un point). Le groupe de forme « TERMES TEST » regroupe les 10 termes du test et leurs variantes graphiques ; le groupe de formes « EST UN » regroupe les expressions présentées plus haut. Les carrés colorés en bleu contiennent au moins l'un des termes de la liste de test ; les carrés colorés en rouge contiennent au moins l'une des expressions étudiées. Lorsqu'un carré est bicolore, c'est qu'il contient au moins l'un des membres de chaque liste.

L'étude de tous les carrés bicolores où se trouvent être co-occurents au moins l'un des membres de chaque groupe de formes et d'expressions montre que les expressions du type « est un » ne sont jamais utilisées pour introduire un

synonyme ou une périphrase définitionnelle de l'un des termes de la liste. La rencontre d'un membre de chaque groupe dans la même phrase se produit onze fois, et dans tous les cas, la relation est établie est du même genre : il s'agit de la catégorisation d'une entité dénommée à l'aide d'un nom propre, comme dans l'exemple ci-dessous (qui correspond au contenu du carré bicolore encadré de bleu dans la figure 52) :

« Aux Etats-Unis, Carr **Futures est le** sixième courtier compensateur **Futures** en terme d'**encours** clients, et le premier détenu par des capitaux étrangers (source CFTC, décembre 2002)¹. »

L'examen des 155 sections comportant au moins l'une des expressions du groupe de formes « EST UN » montre qu'à une exception près², le fonctionnement est le même que celui observé à partir des formes du test. Le même type d'exploration a été conduit pour repérer d'éventuels contextes dans lesquels la conjonction de coordination « ou » pourrait introduire une reformulation. Sur les dix-neuf rencontres au sein de la même phrase entre la coordination et l'un des termes du test, il n'y en a aucune qui soit au service d'une reformulation. Il en va de même pour les 182 sections dans lesquelles la conjonction apparaît.

Enfin, on peut rechercher des traces encore plus ténues laissées par un processus de reformulation, comme la simple apposition. Dans ce cas, la position détachée du groupe nominal apposé est en général matérialisée par un encadrement entre deux virgules. On vérifie donc toutes les suites dans lesquelles un terme ou l'une de ses variantes graphiques apparaissent avant ou après une virgule³. On constate alors qu'aucun des dix termes du test ne bénéficie d'une reformulation introduite par un tel procédé.

¹ *Rapport d'activité de l'année 2002.*

² Il s'agit de la phrase suivante qui correspond à une note de bas de page reproduite à l'identique entre 1995 et 2000 : « (1) *Les certificats coopératifs d'investissement **sont des** titres émis par les Caisses régionales et cotés en bourse.* »

³ On utilise pour cela une expression rationnelle du type « arrangeur.\{0,60\},.\{30\} » qui recherche et sélectionne tout ce qui commence par le terme « arrangeur », se poursuit de zéro à soixante caractères jusqu'à une virgule et se continue au-delà de celle-ci sur trente caractères. Cette expression permet de retrouver des cotextes dans lesquels le terme a été complété par une expansion. Pour le terme « arrangeur » on a par exemple : « *arrangeur, avec des positions solides à* », mais aussi « *arrangeur de crédits syndiqués sur la zone euro, avec un total de 13,2 milliards d'euros (...)* ».

D. Utilisations de l'anaphore comme procédé de reformulation

1. Les anaphores pronominales et la définition déguisée d'« organe central »

Si les traces de reformulations explicitement marquées sont pratiquement inexistantes dans le corpus, on peut néanmoins rechercher d'éventuels reformulants dans le cadre des reprises anaphoriques et cataphoriques. Après avoir examiné les signalements directement liés aux mots graphiques, puis ceux qui s'expriment dans leur environnement immédiat, l'étude des reprises anaphoriques au sens large suppose que l'on se dégage des limites de la phrase pour envisager des phénomènes de reprise susceptibles d'accéder à une portée plus longue. C'est pourquoi, on a retenu le paragraphe comme cadre pour étudier ce type de relation. Comme on le verra plus loin avec les chaînes de référence incorporant un nom propre, le paragraphe présente a priori un espace thématique homogène dont la cohérence est susceptible de reposer au moins partiellement sur des reprises anaphoriques.

Pour la clarté du propos, on distinguera les anaphores pronominales des anaphores lexicales. En effet, si les premières ne proposent pas de reformulant nominal, elles peuvent être utilisées pour décrire l'activité ou l'usage d'une entité, ce qui est une manière indirecte de la caractériser. Quant aux anaphores nominales, elles seront envisagées sous leurs différentes formes (selon la gradation classique, anaphore fidèle, infidèle, conceptuelle et associative)¹, y compris donc lorsque la coréférence entre groupes nominaux se fait de manière indirecte (anaphore associative).

Le repérage de l'anaphore pronominale consiste à isoler, dans un premier temps, les contextes dans lesquels les phénomènes de reprise pronominale sont particulièrement présents. Afin de durcir ce test de premier niveau, on ne retient que les pronoms susceptibles de reprendre entièrement le groupe nominal antécédent : il s'agit des pronoms de la troisième personne et du pronom relatif « qui »². Ceux-ci sont rassemblés dans un groupe de formes « PRONOMS ANAPHORIQUES » et

¹ M. Riegel, J.-C. Pellat, R. Rioul (2004 : 614-615).

² Il s'agit des formes « il » (230 occ.), « Il » (189 occ.), « elle » (103 occ.), « Elle » (106 occ.), « ils » (21 occ.), « Ils » (16 occ.), « elles » (46 occ.), « Elles » (38 occ.) et « qui » (643 occ.).

projetés sur la carte des sections délimitant des paragraphes. On donne ci-dessous un extrait de la carte du corpus résultant de cette projection :

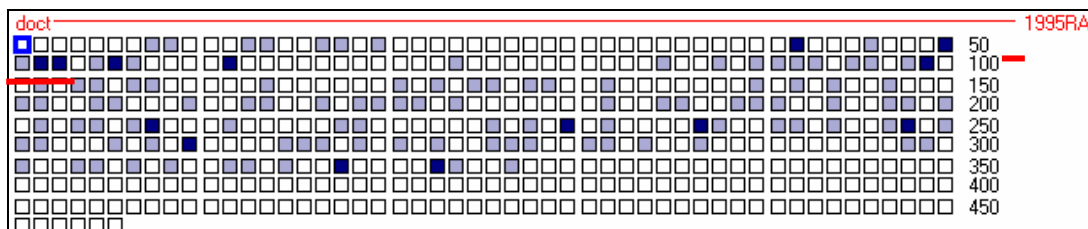


Figure 53.

Carte des sections représentant des paragraphes de la partie 1995 sur laquelle ont été projetées les formes du groupe « PRONOMS ANAPHORIQUES »

Lecture de la figure 53 : sur cette carte sont représentés les paragraphes de la partie 1995 du corpus RA. Les formes du groupe « PRONOMS ANAPHORIQUES » ayant été projetées, les carrés blancs indiquent l'absence, les carrés colorés en bleu pâle indiquent une faible concentration des membres du groupe et les carrés colorés en bleu foncé signalent les paragraphes dans lesquels les membres du groupe de formes sont présents de manière plus significative. En effet, un calcul des spécificités a été appliqué par le logiciel aux paragraphes sélectionnés. Le seuil de spécificité retenu est supérieur ou égal à 2, ce qui n'est pas très élevé, mais suffisamment discriminant ici. La série de carrés soulignée en rouge correspond au seul exemple identifié comportant l'un des termes du test (cet exemple est décrit plus bas).

L'examen de la carte porte sur les carrés colorés en bleu foncé qui signalent des paragraphes où les pronoms sont présents de manière significative. Lorsqu'un carré de cette sorte apparaît sur la carte, les carrés limitrophes sont également consultés de manière à pouvoir suivre les chaînes de reprise dont la portée dépasserait le paragraphe. On dénombre plus d'une centaine de carrés colorés en bleu foncé, répartis sur l'ensemble de la carte, et dans lesquels apparaissent une quarantaine d'antécédents différents, dont certains se retrouvent d'une année sur l'autre¹. Parmi ceux-ci ne figurent aucun des dix termes du test. On observe, d'autre part, que les entités nommées y sont majoritaires (une trentaine) et correspondent à des personnes, des entreprises, des unités, des produits et services ainsi qu'à des

¹ La liste des antécédents est donnée en annexe.

dispositifs particuliers. Lorsque les termes du test sont présents dans ces contextes de reprise (il s'agit d' « organe central », de « moyens de paiement », de « trading », de « gestion alternative », d' « arrangeur » et d' « encours »), c'est en général à titre auxiliaire et sans qu'ils fassent l'objet d'une reformulation. Par ailleurs, comme cette cooccurrence se reproduit rarement d'une année sur l'autre, on peut considérer ces apparitions comme étant peu significatives, à une exception près.

En effet, la présence d' « organe central » signale une configuration récurrente qui apparaît plusieurs années de suite. Cette configuration se présente sous la forme d'une série de carrés contigus contenant au moins l'un des pronoms du groupe de formes dans des proportions inhabituelles. On distingue une telle série sur la deuxième ligne de la figure 53 (elle est soulignée en rouge). Les paragraphes de cette série se succèdent ainsi :

P1 : « *La Caisse nationale assume trois missions majeures : Banque centrale du Crédit agricole, **elle** en garantit l'unité financière. **Elle** gère la trésorerie du Groupe et assure la péréquation nationale des trois quarts des ressources collectées par les Caisses régionales. Pour ce faire, **elle** leur distribue des avances pour financer leurs prêts à moyen et long terme. »*

P2 : « ***Elle** assure la cohérence du développement commercial du Crédit agricole.*

P3 : « ***Elle** conçoit et fabrique les produits d'épargne commercialisés par les Caisses régionales. » Responsable du développement international et des filiales nationales, **elle** appuie l'action des Caisses régionales sur leurs marchés. »*

P4 : « ***Organe** central du Crédit agricole, **elle** veille à la cohésion du réseau, à son bon fonctionnement et au respect des normes de gestion garantissant la liquidité et la solvabilité des Caisses régionales. A ce titre, **elle** représente le Crédit agricole auprès des autorités monétaires, bancaires et financières. » (1995)*

Visible sur la carte dans le prolongement de cette série, une deuxième série offre une description sur le même modèle des missions d'une autre entité, la Fédération Nationale de Crédit Agricole. Dans la série de paragraphes donnée à titre d'illustration, c'est l'antécédent « Caisse nationale de Crédit agricole » qui fait l'objet d'une reprise et non le terme « organe central » (« **Organe** central du Crédit agricole, **elle** veille (...)). C'est donc indirectement que l'on apprend quelque chose

sur les notions d' « organe central » ou de « banque centrale » ; il s'agit en quelque sorte de définitions déguisées. On s'attachera ici à la notion d'« organe central » qui fait partie des termes utilisés pour le test. On pourrait paraphraser le texte du rapport d'activité de 1995 en disant que la Caisse Nationale de Crédit Agricole est un organe central et qu'à ce titre elle en exerce les missions. De même, en empruntant la forme des « énoncés définitoires ordinaires¹ » on pourrait dire qu'un « organe central », « c'est une entité qui veille à la cohésion d'un réseau, etc. ». Le discours des rapports d'activité propose bien dans ce cas une reformulation d'un discours premier, en l'occurrence législatif.

En effet, la caractérisation de la notion d'organe central dans le corpus est tirée des articles L. 511-30 et L. 511-31 *Code monétaire et financier*². Elle procède à un découpage et à une reformulation visant à produire une sorte de résumé de l'article L. 511-31. Dans le texte donné en note³ on a indiqué en caractères gras les

¹ M. Riegel (1990 : p. 97-109).

² Article L. 511-30. du *Code monétaire et financier* : « Pour l'application des dispositions du présent code relatives aux établissements de crédit, sont considérées comme organes centraux : la Caisse nationale de crédit agricole, la Banque fédérale des banques populaires, la Confédération nationale du crédit mutuel, (la caisse centrale de crédit coopératif supprimé loi 2003-706 du 1 août 2003, art. 93-III), la Caisse nationale des caisses d'épargne et de prévoyance ainsi que la Chambre syndicale des sociétés anonymes de crédit immobilier. »

³ « Art. L. 511-31. - **Les organes centraux représentent les établissements de crédit qui leur sont affiliés auprès de la Banque de France, du comité des établissements de crédit et des entreprises d'investissement et, sous réserve des règles propres à la procédure disciplinaire, de la commission bancaire.**

Ils sont chargés de veiller à la cohésion de leur réseau et de s'assurer du bon fonctionnement des établissements qui leur sont affiliés. A cette fin, ils prennent toutes mesures nécessaires, notamment pour garantir la liquidité et la solvabilité de chacun de ces établissements comme de l'ensemble du réseau. Ils peuvent également décider d'interdire ou de limiter la distribution d'un dividende aux actionnaires ou d'une rémunération des parts sociales aux sociétaires des établissements de crédit ou des entreprises d'investissement qui leur sont affiliés. (...)

Ils veillent à l'application des dispositions législatives et réglementaires propres à ces établissements et exercent un contrôle administratif, technique et financier sur leur organisation et leur gestion. Les contrôles sur place des organes centraux peuvent être étendus à leurs filiales directes ou indirectes, ainsi qu'à celles des établissements qui leur sont affiliés.

Dans le cadre de ces compétences, ils peuvent prendre les sanctions prévues par les textes législatifs et réglementaires qui leur sont propres. (...)

Après en avoir informé la commission bancaire et sous réserve des compétences du comité des établissements de crédit et des entreprises d'investissement, les organes centraux peuvent, lorsque la situation financière des établissements concernés le justifie, et nonobstant toutes dispositions ou stipulations contraires, décider la fusion de deux ou plusieurs personnes morales qui leur sont affiliées, la cession totale ou partielle de leur fonds de commerce ainsi que leur dissolution. Les organes dirigeants des personnes morales concernées doivent au préalable avoir été consultés par les organes centraux. Ces derniers sont chargés de la liquidation des établissements de crédit qui leur sont affiliés ou de la cession totale ou partielle de leur fonds de commerce. »

passages concernés par ce travail de reformulation et d'adaptation que l'on trouve dans les rapports d'activité de la première période (1995-2000). D'une manière générale, on peut voir que la reprise des caractéristiques de l'organe central se cantonne aux rôles positifs de chef d'orchestre et de garant. Elle évite les précisions relatives à l'exercice des fonctions régaliennes, telles que les décisions relatives à la suppression ou à la limitation du versement des dividendes, les contrôles administratifs, techniques et financiers portant sur l'organisation et la gestion des établissements affiliés, enfin, les sanctions – pouvant aller jusqu'à la liquidation –, que l'organe central peut prendre à l'encontre des établissements fautifs. On relève à cette occasion une caractéristique typique du travail de communication institutionnelle, qui vise à édulcorer les aspects de la réalité les moins « vendeurs ».

Comme la carte complète des sections permet de le voir¹, ces contextes particuliers se retrouvent d'une année sur l'autre de 1995 à 2000 puis semblent disparaître brutalement. Dans la mesure où, par ailleurs, on a pu voir que le terme « organe central » est présent dans toutes les parties du corpus, la disparition d'un contexte de glose est un phénomène qui mérite attention. En effet, on a ici une illustration du caractère volatil de ce type de contexte dans lequel le locuteur prend soin de fournir à son lecteur des informations destinées à décrire la réalité évoquée. Le fait que ces explications soient jugées superflues à partir de 2001 confirme d'abord la primauté d'une logique de communication sur des préoccupations relevant de la vulgarisation. On voit ensuite que le locuteur anticipe ainsi le peu d'intérêt manifesté par son lectorat pour ce genre de précision. L'élargissement du public semble se traduire par un appauvrissement de l'image du lecteur modèle, pour lequel les spécificités institutionnelles importent moins que les résultats chiffrés. Le phénomène de discrétion, déjà évoqué plus haut, s'applique également aux descriptions des réalités évoquées, que ce soit pour édulcorer ou pour simplifier le contenu du message institutionnel.

Il reste que ce type de contexte à portée définitoire est exceptionnel dans le corpus (parmi la quarantaine d'exemples recensés, une demi-douzaine seulement peut prétendre à ce statut). Dans tous les cas, il s'agit moins d'une reformulation que

¹ Elle est donnée en annexe.

de données encyclopédiques fournies au sujet d'une entité nommée à l'aide d'un nom propre et non d'un terme. Plus précisément, ce type de contexte correspond à la description d'une entité, de son activité¹ ou de son usage. Avec « organe central » on dispose d'une indication sur le fait que ce type de glose dépend étroitement de l'orientation communicationnelle des rapports d'activité. À côté de ce cas de réduction d'une glose récurrente, on trouve des exemples d'expansions ponctuelles. Ceux-ci fonctionnent bien comme des indicateurs de l'importance attachée à l'entité décrite : plus la description est étoffée, plus il y a de chances pour qu'il s'agisse d'un sujet nouveau ou particulièrement mis en valeur dans un rapport d'activité. Toutefois, contrairement à ce qui a été constaté pour « organe central », ces descriptions sont soumises au principe de mise en valeur, qui dépend étroitement de l'actualité du groupe bancaire. En effet, il y a plus de sujets nouveaux qui apparaissent chaque année que de sujets reconduits d'année en année. On trouve, par exemple, la description d'une offre nouvelle, celle de l'activité d'une filiale, le parcours d'un nouveau dirigeant ou le dispositif mis en place par les ressources humaines à la suite du rachat du Crédit Lyonnais. Cependant, ces dispositifs textuels ne permettent pas d'accéder à la complexité de la réalité décrite.

2. Les anaphores nominales et le cas des « encours », entre jargon et banalisation

Afin d'examiner les différentes sortes d'anaphores nominales dans lesquelles les termes du test sont susceptibles d'être encapsulés, deux possibilités se présentent. Tout d'abord, on peut rechercher les contextes dans lesquels toutes les sortes d'anaphoriques sont susceptibles de se rencontrer : outre les pronoms relatifs de la troisième personne et du relatif « qui », il faudrait inclure également les pronoms démonstratifs – susceptibles de pointer sur un référent différent de leur antécédent, et à ce titre éliminés pour le premier test –, ainsi que tous les déterminants définis pouvant entrer dans la composition de groupes nominaux anaphoriques (articles définis, déterminants démonstratifs et possessifs à la troisième personne). La liste des formes attestées dans le corpus, toutes graphies confondues, représente une

¹ Test sur l'ensemble des paragraphes sélectionnés : arrive en tête « Crédit agricole »...

quarantaine d'unités¹. Lorsqu'elle est projetée sur la carte des sections (paragraphes), elle sélectionne 3345 sections sur 6000, soit plus de la moitié de l'ensemble. On obtient de la sorte un clivage du corpus qui se superpose approximativement avec celui des contextes syntactiques et non syntactiques. Dans la mesure où la poursuite de l'enquête dans cette direction va au-devant de résultats difficiles à exploiter, on a opté pour une deuxième possibilité.

Celle-ci consiste à rechercher les co-occurents fréquents des termes du test qui apparaissent dans les mêmes contextes et sont donc susceptibles de jouer le rôle de reformulants. L'avantage de cette méthode est qu'elle permet d'embrasser les différentes sortes d'anaphores nominales mentionnées plus haut. Cependant, son utilisation n'a pas permis d'identifier de reprises anaphoriques nominales pour les membres du groupe de test. Un cas, néanmoins, sera mentionné afin d'illustrer la méthode utilisée et de préciser la situation du terme « encours » dont l'usage dans le corpus ne correspond pas uniquement à celui mentionné dans les définitions données pour ce terme.

Le terme « encours » est ainsi défini dans le *LBMF* : « **Encours bancaires** : volume total des crédits accordés par les établissements de crédit à un moment donné. » Cette définition est différente de celle donnée par le *NPRE* : « Ensemble des effets remis par une entreprise à une banque et qui n'ont pas atteint la date de leur échéance. » On a alors décidé de ne pas considérer ce terme comme ressortissant à la banalisation terminologique mise en lumière dans le chapitre 6. En effet, on peut voir que, contrairement au mouvement observé habituellement, c'est ici le dictionnaire de langue qui donne une définition plus restrictive que le lexique spécialisé. D'autre part, on peut également constater que les deux définitions n'adoptent pas le même point de vue : dans la première, il s'agit de la banque, alors que dans la seconde, il s'agit de l'entreprise cliente². Enfin, si la première définition

¹ Liste du premier groupe d'expressions : « Ce ; ce ; Celle ci ; celle ci ; Celles ci ; celles ci ; Ces ; ces ; Cet ; cet ; Cette ; cette ; ceux ci ; Elle ; elle ; Elles ; elles ; Il ; il ; Ils ; ils ; L ; l ; La ; la ; Le ; le ; Les ; les ; Leur ; leur ; Leurs ; leurs ; Sa ; sa ; Ses ; ses ; Son ; son. »

² Néanmoins ce dernier point est mentionné dans le Grand Robert de la langue française pour les éditions de 1989 et de 2001 : « *Montant des effets escomptés par une banque, non arrivés encore à échéance. Encours de crédit ou encours : montant des crédits utilisés par un client auprès de sa banque ; (pour une banque) montant de l'ensemble des crédits utilisés par sa clientèle.* »

opte pour le pluriel, la seconde prend le parti du singulier. Ces constats suggèrent au moins que le statut du terme « encours » n'est pas clairement établi. Ce point gagne donc à être précisé à partir des données fournies par le corpus.

Dans le corpus des rapports d'activité, la forme « encours » est présente dans toutes les parties et possède 361 occurrences, dont 13 avec une majuscule initiale réservée aux contextes non syntactiques. On relève à partir des 13 occurrences de la forme avec majuscule initiale, les termes suivants : « Encours de crédit » (5 occurrences) et « Encours de collecte », (5 occurrences également). Cet extrait des occurrences permet d'effectuer deux précisions sur l'usage qui est fait d' « encours » dans le corpus : d'une part, il indique que le point de vue adopté dans les rapports d'activité est, de façon peu surprenante, celui de la banque ; d'autre part, l'utilisation du mot – peut-on encore parler de terme ? – ne semble pas être limité au domaine du crédit puisqu'il est également employé pour parler d'épargne (collecte). Cette diversité d'emplois est confirmée par l'examen des contextes d'apparition du mot.

Tout d'abord, il fonctionne très fréquemment avec une élision de l'expansion, comme c'est le cas dans les phrases suivantes : « *Confirmant sa première place auprès des épargnants, sa collecte totale, en hausse de 6,8 %, a atteint 121 milliards de francs, portant l'encours à 1 906 milliards de francs.* » (1995), ou « *Les Caisses régionales enregistrent une légère progression de leurs encours (+ 5,9%) sur un an, dans un marché en recul de 3, 2 %, à fin décembre.* » (2003). La répartition entre usage du pluriel ou du singulier n'est corrélée ni avec un emploi particulier ni avec un domaine spécifique (crédit, épargne ou assurance). L'analyse des segments répétés incluant la forme « encours » confirme la présence d'unités complexes, codées ou non, utilisées dans ces différents domaines :

*Tableau 44.
Segments répétés fréquents incluant la forme « encours »*

Segments	Long	Fréq
encours gérés ¹	2	78
encours de crédit	3	26
encours de collecte	3	7

¹ Vaut aussi bien pour les actifs financiers, les crédits, les contrats d'assurance.

encours d OPCVM	3	5
encours des financements	3	5
encours des prêts	3	5
encours des produits	3	4
encours gérés en assurance vie	5	3

Toujours dans le but de cerner le fonctionnement du mot « encours » dans le corpus, et en particulier pour rechercher des mots qui pourraient lui être substitués dans des contextes similaires, on a appliqué la méthode des co-occurents de co-occurents déjà utilisée dans le chapitre précédent à propos du syntagme lexicalisé « banque à distance »¹. On rappelle que cette méthode consiste à calculer d’abord les co-occurents les plus fréquents d’un mot – on utilise pour cela un calcul des spécificités sur une fenêtre correspondant à la phrase –, puis à sélectionner parmi eux ceux qui ont un coefficient de spécificité positif très élevé ; on réitère alors le calcul des co-occurents sur cette sélection. On vérifie ensuite l’hypothèse selon laquelle le fait de partager des contextes similaires (ces contextes sont en quelque sorte résumés par la sélection des principales formes co-occurentes obtenues lors de la première étape) constitue un indice de substituabilité et donc de proximité sémantique entre deux mots. On donne ci-dessous les formes utilisées pour le calcul et celles qui ont été collectées :

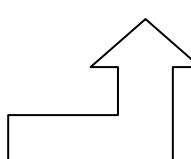
¹ W. Martinez (2002).

Tableaux 45 et 46.

Principaux co-occurents de la forme « encours » et sélection de co-occurents des co-occurents de cette forme

Forme	Frq. Tot.	Fréq.	Coeff.
gérés	175	96	***
euros	592	141	***
francs	625	198	***
milliards	828	312	***
progression	230	59	25
progressé	109	40	24
atteint	140	41	20
crédits	143	41	20
croissance	288	58	19
augmentation	109	29	14

Forme	Frq. Tot.	Fréq.	Coeff.
total	161	119	45
résultat	115	82	30
chiffre	66	55	27
montant	78	57	22



Dès lors, on peut s'interroger plus précisément sur la nature des relations établies entre ces unités dans les textes du corpus. Là encore, l'observation des contextes est facilitée par un premier repérage réalisé à partir de la carte des sections dans laquelle on a recherché les contextes lâches (rubriques), dans lesquels la forme pôle se trouve dans le voisinage de l'un des quatre co-occurents sélectionnés. On donne ici à titre d'exemples certains rapprochements mis en lumière :

E1 : « (...) les **encours** gérés par Sofinco ont augmenté de 21 % pour atteindre 21,6 milliards d'euros, les **encours internationaux** pesant plus du quart de **ce total**, en hausse de 26 %. » (2003)

E2 : « Les **encours** des deux principales filiales, CA-AM et CLAM ont atteint 273, 1 milliards d'euros, en hausse de 13 % par rapport à la fin 2002, sous le double effet des nouvelles souscriptions (18,3 milliards d'euros) et d'un effet marché positif (13,5 milliards d'euros). **Ce résultat** a pu être obtenu grâce à la mobilisation et au dynamisme commercial des différentes unités, qui ont toutes contribué à cette évolution. » (2003)

E3 : « L'activité de Predica a pleinement bénéficié de l'accentuation de la reprise du marché de l'assurance - vie. Son **chiffre d'affaires** a dépassé 10 milliards d'euros (66,4 milliards de francs), en hausse de 38 %, entraînant une progression de 13,6 % des **encours gérés** qui se sont établis à 70,3 milliards d'euros (460,8 milliards de francs). » (2000)

E4 : « *Banquier d'une entreprise sur quatre, le Crédit agricole a enregistré, sur ce marché, une augmentation de 11 % de l'encours des crédits, avec un montant de 161,5 milliards de francs.* » (1998).

Le fonctionnement d'« encours » dans le corpus suggère de façon générale que cette unité joue le rôle d'un mot passe-partout, dont la signification pourrait être résumée ainsi : montant total ou volume global des opérations de banque, des services d'investissements ou des activités parabancaires (assurance) pour une période donnée¹. On a, par exemple : encours des crédits utilisés par la clientèle ou consentis par la banque à un moment donné ; encours de collecte, d'OPCVM, de Sicav, encours de contrats d'assurance-vie, etc. Dans la mesure où l'on constate que le sens de ce mot tend à se diluer dans son domaine d'utilisation, on peut faire l'hypothèse qu'il correspond à une forme de jargon bancaire². Ce terme n'est probablement pas encore entré dans l'usage commun, car il exprime plutôt le point de vue du banquier. Dans le cas d'« encours », le lecteur néophyte est en quelque sorte livré à lui-même et il lui faut s'appuyer sur la langue commune pour échafauder un sens à partir des idées de montants ou de sommes « en cours », c'est-à-dire ayant un rapport avec l'activité de la banque durant une période donnée. On est donc très loin dans ce dernier cas des préoccupations dont on peut penser qu'elles devraient légitimement être associées à la vulgarisation d'une terminologie.

¹ On remarque dans le corpus que « volume » est réservé au domaine de la bourse – émissions, opérations, titres), ce qui peut expliquer son exclusion des contextes de « encours » alors qu'il pourrait jouer le rôle d'un quasi synonyme.

² De ce point de vue on peut le rapprocher d'« encaisse » (ce que l'on a encaissé, mis dans la caisse), « encours » se laissant paraphraser par « ce qui est en cours dans les comptes de l'année, les opérations qui ne sont pas closes durant cette période ».

Un vocabulaire peut-il être transparent, c'est-à-dire être compris quelles que soient les compétences linguistiques et communicationnelles des interlocuteurs ? Il semble que pour répondre à cette question il soit nécessaire de préciser au préalable le point de vue que l'on adopte. Le lecteur des rapports d'activité, obligé de se frayer un passage entre langue commune et terminologie banalisée sur fond de phraséologie économique un peu brumeuse, passe probablement à côté des phénomènes mis au jour par l'analyse. Tout est même fait pour lui dissimuler ces mouvements internes liés à la fabrication d'un discours qui doit paraître le plus lisse possible. On laissera donc le lecteur type des rapports d'activité de ce côté du miroir, pour voir ce qu'il en est de l'autre côté du discours, du point de vue de l'émetteur.

Le dispositif communicationnel de base des rapports d'activité repose sur l'utilisation d'un vocabulaire réputé accessible. Les changements qui interviennent dans la situation de communication en 2001 constituent une sorte de test grandeur nature. Devant l'élargissement de son public, le locuteur collectif des rapports d'activité ne recourt pas véritablement à de nouvelles armes. Il renforce plutôt l'utilisation de celles dont il dispose déjà. Dans les chapitres 2 et 3, on a vu apparaître un contexte mixte qui a tendance à mordre dans les contextes de texte suivi pour les trois dernières années du corpus. En ce qui concerne le vocabulaire, le même mouvement se traduit par une utilisation plus importante de la langue commune. La restitution de l'activité bancaire, déjà assez générale, tend alors vers une sorte de schématisation.

Ces mouvements ne peuvent être saisis dans leur complexité que si l'on prend soin de caractériser les variations de fréquence qui affectent les vocables du corpus. En effet, il convient de distinguer à chaque fois les causes de ces mouvements, faute de quoi aucune interprétation ne peut véritablement être proposée. Les effets de discrétion ou de valorisation doivent être séparés des variations enregistrées mécaniquement par les discours. Ce n'est que dans un deuxième temps que le rôle tenu par différentes variétés de langue peut être examiné. On voit alors que la langue commune est au service d'une restitution généraliste de l'activité bancaire, renforcée à partir de 2001 dans le dispositif de communication existant. Cette variation de fréquence n'est pas de même nature que

celle qui affecte la terminologie bancaire et la phraséologie économique des rapports d'activité. En effet, dans la mesure où ces variétés de langue sont plus étroitement en prise sur l'activité bancaire telle qu'elle est restituée, les fréquences des vocables qui leur sont rattachés ont tendance à fluctuer au gré de l'actualité. Ces variations de fréquence doivent néanmoins toujours être rapportées aux effets de mise en valeur ou de discrétion qui peuvent renforcer une tendance ou l'inhiber. La question de l'utilisation de variétés de langue différentes dans le corpus a également été abordée d'une autre manière. En effet, on s'est aussi demandé si l'utilisation de termes dont la banalisation n'est pas avérée entraîne une prise en charge énonciative particulière.

À une exception près (« organe central »), et encore ne s'agit-il que d'un cas indirect, les termes utilisés pour tester la prise en charge énonciative de la terminologie bancaire dans les rapports d'activité montrent que celle-ci ne fait pas l'objet de reformulations destinées à en expliciter le sens. Ni les marques typographiques, ni les noms ou les procédés métalinguistiques, ni le recours à l'anaphore ne sont utilisés pour construire un paradigme désignationnel autour d'un terme. On peut tirer plusieurs conclusions de ce constat. Tout d'abord, il est permis de se demander si la problématique de la reformulation a été abordée dans tous ses aspects. Comme on l'a indiqué plus haut, le but n'était pas de faire un tour d'horizon complet de cette problématique, mais de sonder les rapports d'activité. Il se peut donc que des cas aient échappé à nos investigations. Néanmoins, on conviendra que celles-ci ont été assez poussées pour ne rien laisser échapper au moins sur les points étudiés. Une autre question surgit alors : cette série de tests est-elle satisfaisante ? En effet, le résultat paraît assez maigre. On pense toutefois que cette exploration textométrique n'a pas été vaine.

En effet, il apparaît au terme de cette analyse que les rapports d'activité constituent une forme particulière de discours seconds, dans laquelle le parti pris communicationnel visant à produire un discours « lisse » a en quelque sorte pour contrepartie de bloquer les mécanismes classiques de vulgarisation, tels que la reformulation ou la définition des termes réputés difficiles. Seuls les noms propres semblent échapper à cette absence de prise en charge. C'est un point important qui ressort de ces analyses : en dehors de l'utilisation d'unités lexicales censément partagées par tous les destinataires (la langue commune, utilisée afin de promouvoir

une restitution commerciale de l'activité bancaire, les termes banalisés, ou encore, les mots que l'on a qualifiés d'économismes), ce genre de discours ne comporte pas de prise en charge explicite de certains termes réputés difficiles, que ce soit à l'aide de reformulations ou de définitions. On peut dire que l'usage qui est fait de la phraséologie des rapports d'activité est à l'image d'une restitution positive de l'activité bancaire : son existence tient guise d'explication. De ce point de vue, le tour de force des rapports d'activité consiste à laisser croire que des termes difficiles et une phraséologie imprécise relèvent du même degré d'évidence positive que les réalités qu'ils permettent d'évoquer.

En résumé, dans ce tableau du vocabulaire des rapports d'activité, les noms propres font figure d'exception. On trouve ici une heureuse confirmation de la capacité de ces unités à jouer le rôle de révélateurs, de vigies pourrait-on dire, lorsque l'on cherche à observer les caractéristiques d'une situation de déséquilibre cognitif entre interlocuteurs.

CONCLUSION DE LA DEUXIÈME PARTIE

Dans cette partie, on a cherché à caractériser le vocabulaire des rapports d'activité, parce qu'il constitue la matière première de l'extraction d'information nécessaire à l'élaboration du référentiel terminologique adaptable au contexte que l'on a projeté de mettre en place. Cette caractérisation ne va pas de soi : le vocabulaire des rapports d'activité relève-t-il de la langue commune, compte tenu du fait que ces textes ont un large public extérieur à l'entreprise, ou bien faut-il plutôt le rattacher à la terminologie propre aux domaines de la banque et de la finance ? Plutôt que d'apporter une réponse a priori, il a semblé plus fructueux d'interroger les rapports qu'entretient un genre de discours avec les variétés de langue qu'il mobilise. Plus précisément, il s'agit de rendre compte de la manière dont les variétés disponibles dans un parler d'entreprise à un moment donné sont réfractées par un genre de discours, ici le sous-genre de discours économique des rapports d'activité.

Dans cette approche bakhtinienne, on considère, en effet, que le genre joue le rôle d'un creuset, au sein duquel prennent forme les interactions entre une situation de communication et le vocabulaire utilisé pour évoquer une forme d'activité¹. Dans la mesure où il s'agit de collecter des unités lexicales ou terminologiques, on s'attache aussi à repérer, au sein d'une variété, les strates dans lesquelles les vocables sont plus particulièrement sélectionnés. Il n'est pas inutile, par exemple, d'identifier les termes utilisés préférentiellement pour restituer les aspects généraux de l'activité bancaire ; de même, il n'est pas indifférent de noter que la terminologie bancaire est utilisée dans ce genre de discours avec une acception élargie aux principaux aspects économiques de cette activité. En procédant ainsi, on se donne les moyens de préciser les propriétés et les limites de la collecte que l'on envisage pour un corpus donné, ce qui devrait permettre ultérieurement d'en combler les lacunes.

Ce vocabulaire a été étudié de manière certes qualitative, mais avec l'aide de données quantitatives, car celles-ci permettent une analyse méthodique et reproductible, qui semble mieux adaptée au but poursuivi. Pour être productive,

¹ M. Bakhtine ([1979] 1984 : 265-308). Quoique demeurées à l'état programmatique en ce qui concerne les genres de discours, les réflexions de M. Bakhtine demeurent éclairantes et fécondes. On renvoie pour la mise en perspective de ce texte à T. Todorov (1981 : 123-143).

cette approche doit aussi être contrôlée, ce qui passe, entre autres, par l'explicitation des options méthodologiques retenues à chaque étape de l'analyse. La caractérisation des variétés de langue suppose ainsi que l'on précise les propriétés de la norme retenue, pour décider du rattachement de tel vocable à la langue commune ou à la terminologie bancaire et financière. Dans la mesure où aucune norme n'est tout à fait satisfaisante, on a cherché à contrôler les résultats obtenus, que ce soit de façon manuelle ou automatique, en adoptant des points de vue complémentaires : corpus de référence, terminologie et dictionnaires externes. Le mot « banque », étudié de façon assez détaillée dans les chapitres de cette partie, a permis d'illustrer ce point.

Les graphies des vocables n'ayant pas été normalisées lors de la constitution du corpus, il a été nécessaire d'examiner séparément « banques », « banque » et « Banque ». L'étude du fonctionnement discursif de ces trois vocables a permis d'obtenir une caractérisation précise qui a montré, entre autres, que chaque graphie est utilisée majoritairement dans une strate différente. Si « banques » est bien employé dans le corpus avec le sens que l'on retrouve dans les dictionnaires de langue, y compris dans le *Dictionnaire fondamental de la langue française*, il n'en va pas de même pour « banque ». Cette graphie, du fait de son inclusion dans des syntagmes lexicalisés, relève de différentes strates terminologiques : terminologie bancaire du *Lexique de la banque et des marchés financiers*, terminologie bancaire généraliste – absente du *LBMF*, mais présente dans un corpus rassemblant des rapports d'activité d'autres établissements bancaires –, terminologie économique générale, voire phraséologie propre aux rapports d'activité du corpus. Enfin, la graphie « Banque », qui partage certaines strates terminologiques avec la graphie précédente, est en général incluse dans des dénominations propres assez variées, qu'il s'agisse de noms de banques ou de noms de pôles d'activité spécialisés à l'intérieur d'un même établissement.

Les vocables ainsi caractérisés ont également été répartis en fonction d'une grille distinguant des composants de base de l'activité bancaire : agents, objets, opérations économiques, ensembles complexes regroupant ces différents composants, propriétés des composants et divers. Cela a permis d'observer que chaque variété de langue ou chaque strate identifiée fournit des unités possédant des

propriétés spécifiques pour ces différentes catégories. La langue commune restitue l'activité bancaire en des termes assez généraux mais suffisants pour éclairer la tendance des résultats obtenus au cours d'une année. Elle propose une lecture simplifiée de l'activité bancaire, réduite à ses aspects commerciaux. La terminologie restitue la diversité de l'activité bancaire avec plus de finesse ; néanmoins, elle offre une présentation assez générique qui ne détaille pas les différents aspects de l'activité. Lorsque c'est le cas, cette présentation est ponctuelle et relativement peu développée. Enfin, la phraséologie des rapports d'activité recèle d'autres catégories qui s'échelonnent entre ces deux pôles formés par la langue commune et la terminologie bancaire. En effet, on a pu relever que certains mots sont utilisés de manière récurrente pour la seule facette économique de leur aire sémantique ; c'est le cas, par exemple, de « groupe » qui sert toujours à désigner un groupe financier ou industriel. L'étude des syntagmes lexicalisés a permis d'identifier de nombreuses variétés réparties sur différentes strates terminologiques : terminologie bancaire, mais aussi terminologie économique au sens large intégrant des aspects relatifs au marketing. Contrairement aux unités terminologiques, la plupart de ces syntagmes possèdent un sens général qui ne permet de désigner que des réalités économiques aux contours imprécis, ce qui contribue à renforcer l'impression d'un traitement à grosses mailles de l'activité bancaire. Il faut noter cependant la présence d'une terminologie spécifique au groupe bancaire Crédit Agricole, ainsi que de nombreuses dénominations propres.

L'organisation du corpus sous la forme d'une série textuelle chronologique a permis de mettre en évidence d'autres propriétés de la sélection lexicale propre au genre des rapports d'activité. En effet, qu'il s'agisse de phénomènes économiques généraux ayant un impact sur l'activité bancaire, ou bien de modifications intervenues dans la situation de communication des rapports d'activité en 2001 (changement de statut pour l'émetteur, élargissement du public des rapports d'activité, influence renforcée de la réglementation des marchés), l'approche diachronique a permis d'identifier la manière dont le dispositif communicationnel des rapports d'activité traduit les interactions entre phénomènes externes et sélection du vocabulaire. C'est ainsi que l'on a pu observer un renforcement du recours à la

langue commune à partir de 2001, qui peut être interprété comme la marque d'un souci de simplification du message des rapports d'activité destiné à un public élargi.

Les causes des variations de fréquence affectant les autres variétés de langue semblent plus diversifiées. Elles peuvent être d'ordre « mécanique » lorsque, par exemple, les contextes non syntactiques sont remaniés. En effet, les informations présentées à l'intérieur de ces contextes étant organisées au sein de dispositifs récurrents, la modification de ces derniers peut entraîner des variations de fréquence brutales, qui n'ont une portée thématique que très indirecte. Il en va de même pour un changement d'unité monétaire, qui modifie la fréquence de certains vocables, en quelque sorte par ricochet (les milliards de francs deviennent dans certains cas des millions d'euros). Surtout, il a été possible d'observer deux effets concurrents du dispositif communicationnel susceptibles de renforcer ou d'inhiber la fréquence des vocables. Il s'agit de ce que l'on a proposé d'appeler les effets de discrétion et de mise en valeur, qui agissent en fonction de la finalité assignée au genre de discours : restituer sous ses meilleurs jours l'activité du groupe bancaire, réduire autant que possible la place des sujets délicats, tant internes qu'externes. Cela permet de comprendre, entre autres, la variation structurelle des fréquences de certains types d'unités relevant de la terminologie prise au sens large. On voit ainsi que les termes relativement spécifiques, qui permettent de traiter des points de détail de l'activité bancaire, sont sollicités seulement en fonction d'une actualité qu'il s'agit de mettre en valeur ; de même, peut-on observer un parallèle entre le développement d'un secteur d'activité et l'augmentation de la fréquence du terme qui sert à le désigner. Cependant, en dehors de certains cas précis, il n'y a pas de relation mécanique observable entre ces deux ordres de phénomènes. Ce qui ressort de ces observations, c'est plutôt que la fréquence de chaque vocable est sous la dépendance directe du dispositif communicationnel qui l'utilise pour une variété de langue donnée.

On ne déduit pas de ces observations que le locuteur collectif des rapports d'activité contrôle parfaitement son discours au regard de l'objectif de communication grand public qu'il poursuit. Des lacunes sont en effet perceptibles lorsque l'analyse permet de rompre avec le sentiment d'évidence que peut laisser une lecture naïve. Lorsque l'on examine la ventilation des occurrences de différents syntagmes dans les neuf parties du corpus, il apparaît que la concurrence des

désignations (entre néologismes, ou entre unités appartenant à des strates terminologiques différentes), se double d'un renouvellement régulier du vocabulaire (voir chapitre 3). Cette instabilité des désignations, qui se chassent les unes les autres à l'occasion de bouleversements organisationnels, d'effets de mode ou lors de l'intégration de cultures d'entreprise différentes, est aggravée par l'instabilité sémantique qui affecte certains syntagmes lexicalisés. De même a-t-on constaté qu'un groupe de termes absents des dictionnaires de langue, contrairement à la majorité des termes récurrents qui sont sélectionnés dans le corpus, ne font l'objet d'aucune prise en charge énonciative destinée à en expliciter le sens. Il en va de même pour de nombreux syntagmes lexicalisés, dont le sens est ambigu ou imprécis. De cette absence de prise en charge énonciative, alors que c'est un large public qui est visé, on peut déduire que les rapports d'activité constituent une forme particulière de discours seconds. C'est pourquoi, on émet l'hypothèse selon laquelle l'absence de prise en charge explicite tient au caractère positif de ce genre de discours : le vocabulaire y est utilisé comme s'il était consubstantiel aux réalités désignées. Cette conception aurait pour effet de repousser toute prise en charge énonciative explicitement destinée à prévenir d'éventuelles difficultés de compréhension chez le lecteur.

Sous réserve de vérifications qui restent à mener, il semble que ces remarques puissent être étendues aux autres sortes de discours produits par les entreprises. Car, en donnant à voir l'acte d'énonciation et les positions réelles des interlocuteurs (qui parle, à qui, de quel sujet, dans quel but ?), une telle prise en charge énonciative explicite va à l'encontre d'un sentiment d'évidence, qui peut être cultivé à dessein par l'énonciateur¹. Dans tous les cas, les résultats obtenus lors de la caractérisation du vocabulaire des rapports d'activité montrent qu'un tel dévoilement paraît incompatible avec les finalités économiques poursuivies dans ce genre de discours.

¹ On a vu dans la première partie que ce sentiment a été identifié comme étant un obstacle à surmonter pour la réalisation des activités d'information. La démarche d'exposition des variations affectant un parler d'entreprise va donc non seulement contre ce sentiment, mais elle peut aussi aller contre la visée initiale de certains genres de discours, comme les discours de communication utilisés comme sources pour la constitution de référentiels terminologiques.

Seule l'opacité sémantique des noms propres est parvenue à toucher la conscience linguistique du locuteur collectif des rapports d'activité. Dans le chapitre 3 on a pu relever que les noms propres donnent indirectement accès aux modifications affectant le référentiel du groupe bancaire. Baisses et augmentations brutales de fréquences fonctionnent à la manière d'alertes qu'il s'agit ensuite de décoder. On a vu également que les procédés typographiques, métalinguistiques et anaphoriques refusés aux unités terminologiques susceptibles d'être reformulées sont fréquemment mis à contribution pour les noms propres. Or, il ne s'agit plus alors de reformulations : dans ce cas, ce sont des informations de nature encyclopédique qui ont été insérées dans le discours. Soit il s'agit de noms propres utilisés pour donner un cadre de référence, et dans ce cas ce sont plus naturellement les toponymes qui sont mis à contribution, soit il s'agit d'apporter une information sur une entité qui peut se limiter, comme pour les termes, à la mention d'un sigle ou de son développé. Ainsi, un toponyme mis entre parenthèses permet-il de préciser le pays de rattachement des banques dans lesquelles le Crédit agricole possède une participation. Il en est de même pour les sigles et leur développé. En outre, on a constaté à plusieurs reprises, dans les exemples donnés à propos des termes, que ces informations sont susceptibles de porter sur les entités elles-mêmes. Cependant, elles ne sont pas toutes de même nature ni ne bénéficient des mêmes développements. Il peut s'agir d'une catégorisation de l'entité. Cette catégorisation peut se transformer en description lorsqu'il s'agit de préciser le rôle d'une entité. Enfin, l'apport d'informations encyclopédiques relatives à une entité peut sortir du cadre de la catégorisation ou de la description pour prendre la forme du rattachement à un domaine d'activité. Ces premières indications seront reprises et amplifiées dans les chapitres suivants.

Ces investigations ont pu être menées grâce à l'utilisation de différentes techniques textométriques. Celles-ci se sont révélées capables d'aborder l'étude du vocabulaire à des niveaux d'analyse très différents. Une approche macroscopique a permis d'observer des mouvements d'ensemble présentés dans le chapitre 3, et de procéder à une comparaison globale avec une norme externe pour la langue commune (chapitre 4) ou la terminologie du domaine (chapitre 5). On a pu suivre l'évolution des fréquences de groupes de vocables représentatifs de variétés de

langue ou de certaines de leurs strates. Pour cela, l'approche textométrique a facilité l'identification de formes présentes dans toutes les parties du corpus (vocabulaire commun), mais aussi la constitution de groupes de formes représentatifs d'ensembles plus vastes. La présence relative de ces groupements dans les différentes parties du corpus a également permis de montrer des évolutions sensibles dans la répartition des variétés de langue utilisées sur une période d'une dizaine d'années seulement (chapitres 6 et 7). Par ailleurs, la méthode des spécificités a facilité le repérage des formes ou polyformes responsables des variations caractéristiques de telle partie ou de telle période du corpus. Bien plus, l'approche textométrique a grandement facilité une étude fine du fonctionnement discursif de certains groupes d'unités ou d'unités prises isolément.

Il a ainsi été possible d'observer des procédés discursifs tels que la reformulation, l'anaphore, mais aussi des signalements métalinguistiques, comme le recours aux parenthèses ou aux guillemets. Chaque forme ou polyforme a pu être diversement documentée selon les besoins : qu'il s'agisse de l'identification de variantes graphiques, de leur environnement immédiat sur l'axe syntagmatique (concordances et segments répétés), de co-occurents, ou encore, de co-occurents de co-occurents utilisés pour le repérage d'éventuels quasi-synonymes et de paradigmes désignationnels. La carte des sections de Lexico a facilité le repérage visuel de microphénomènes se produisant en grand nombre, tout en permettant de « zoomer » sur le contenu de certains paragraphes précis. On a ainsi pu constater que ces techniques d'exploration résistent à la difficulté et à la variété des problèmes soulevés. Elles répondent également au souhait, déjà évoqué, de disposer de procédures aisément reproductibles. Ces points doivent être pris en compte, dans la mesure où une approche méthodologique relative à la mise en place de référentiels terminologiques ne doit pas perdre de vue les aspects opérationnels inhérents à de tels projets.

Enfin, la caractérisation du vocabulaire des rapports d'activité a permis de préciser ce que l'on peut attendre d'un référentiel terminologique adaptable au contexte, élaboré à partir d'un corpus de rapports d'activité. Il doit combler des lacunes de plusieurs sortes. Pour des raisons déjà évoquées plus haut, les discours d'entreprise négligent fréquemment une dimension de vulgarisation, qui doit

nécessairement accompagner la diffusion d'informations particulières en dehors du cadre limité au sein duquel elles sont produites. Comme le but recherché n'est pas de réformer de tels discours, mais de faciliter l'accès à l'information qu'ils véhiculent, la solution peut résider dans le recours à un objet périphérique, le référentiel terminologique. Une première qualité pour un tel référentiel doit consister en ce qu'il répertorie les différentes manières de parler d'une même chose dans une situation de communication donnée, en veillant à ne négliger aucune strate des variétés de langue dans lesquelles les vocables sont sélectionnés : tout ce qui peut faire passerelle entre les différentes zones d'un parler d'entreprise a sa place dans un tel référentiel. Or, on a vu que ces zones peuvent être assez nombreuses dans le seul cadre d'un genre particulier et pour une période limitée à une dizaine d'années. Un autre aspect de la question se rattache à la sorte d'amnésie qui frappe ces discours lorsqu'on les étudie dans la durée : une phraséologie en chasse une autre, sans que les traces de ce renouvellement du vocabulaire soient conservées. Un référentiel devrait pouvoir être une sorte de journal de la manière dont un parler est utilisé dans une situation de communication type ; car rien ne dit que cette évolution du vocabulaire des discours soit coordonnée avec le référentiel individuel des salariés de l'entreprise ou, *a fortiori*, avec celui des lecteurs externes. L'expérience semble même indiquer le contraire. À l'évidence et à l'immédiateté des discours, le référentiel terminologique donnerait accès à la complexité du dire et aux traces laissées par son inscription dans la durée.

Ce point permet d'aborder une limitation importante que doit assumer un tel référentiel dès sa mise en place : il ne s'agit pas de tout inventorier ni de tout stocker, car si tant est que cela soit possible, il n'est pas certain que l'on parvienne ainsi à éclairer des processus de communication, tels que la recherche de l'information, son organisation ou sa diffusion. Le rôle du référentiel consiste donc à fournir à ses utilisateurs des indications utilisables pour elles-mêmes, et surtout, susceptibles de mettre en éveil leurs ressources linguistiques personnelles. En proposant des exemples de fonctionnements langagiers pour des situations de communication type, il doit favoriser toutes les extrapolations utiles à une meilleure circulation de l'information. On a suggéré précédemment que le fonctionnement des dénominations propres dans les discours pourrait constituer un point d'accès

possible pour la collecte de ces données : c'est la question qu'il s'agit d'examiner maintenant.

UNIVERSITÉ DE LA SORBONNE NOUVELLE – PARIS 3

École doctorale 268 – Langage et langues

DISCOURS D'ENTREPRISE ET ORGANISATION DE
L'INFORMATION

APPORTS DE LA TEXTOMÉTRIE DANS LA CONSTRUCTION DE
RÉFÉRENTIELS TERMINOLOGIQUES ADAPTABLES AU CONTEXTE

Tome 2

Thèse de doctorat de Sciences du langage

présentée et soutenue publiquement

par

Frédéric ERLOS

Sous la direction de :
André SALEM, Professeur

Jury :

Loïc DEPECKER
Brigitte GUYOT
Monique SLODZIAN

26 novembre 2008

UNIVERSITÉ DE LA SORBONNE NOUVELLE – PARIS 3

École doctorale 268 – Langage et langues

DISCOURS D'ENTREPRISE ET ORGANISATION DE
L'INFORMATION

APPORTS DE LA TEXTOMÉTRIE DANS LA CONSTRUCTION DE
RÉFÉRENTIELS TERMINOLOGIQUES ADAPTABLES AU CONTEXTE

Tome 2

Thèse de doctorat de Sciences du langage

présentée et soutenue publiquement

par

Frédéric ERLOS

Sous la direction de :
André SALEM, Professeur

Jury :

Loïc DEPECKER
Brigitte GUYOT
Monique SLODZIAN

26 novembre 2008

TROISIÈME PARTIE

DÉNOMINATIONS PROPRES ET RÉFÉRENTIEL TERMINOLOGIQUE

INTRODUCTION DE LA TROISIÈME PARTIE

La précédente partie a permis de dessiner en quelque sorte la « toile de fond » lexicale à partir de laquelle doit être réalisée la collecte de certaines unités appartenant au vocabulaire du corpus. Cette documentation de la source textuelle utilisée pour la constitution du référentiel terminologique a mis en évidence la manière particulière dont un parler d'entreprise est mis à contribution dans une situation de communication recourant à un genre de discours spécifique. Cette étude a également permis de préciser l'importance prise par les noms propres dans un tel contexte. On a pu voir dans la première partie que ces derniers fonctionnent comme des indicateurs pertinents des principaux mouvements ayant pu affecter un référentiel. Dans la seconde partie, l'importance des noms propres dans le vocabulaire du corpus a été précisée. Ainsi, ils représentent 1 / 5^e des unités les plus fréquentes appartenant au vocabulaire commun du corpus. Ce point sera approfondi plus loin, mais il semble déjà indiquer l'existence d'une correspondance entre la fréquence de certains noms propres dans les discours et la place qu'ils occupent dans le référentiel dont il est question. Par ailleurs, cette classe d'unités est la seule à faire l'objet d'une prise en charge énonciative, se faisant sous la forme d'un apport d'informations de nature encyclopédique.

On a présenté dans la première partie deux hypothèses relatives à l'utilisation des noms propres pour la constitution de référentiels terminologiques. La première consiste à considérer que les noms propres constituent un moyen utile afin de repérer certains des éléments privilégiés appartenant au référentiel d'une organisation. La pragmatique des ces unités particulières indique, en effet, que ce ne sont pas tous les particuliers présents dans un référentiel quelconque qui sont dotés d'une dénomination propre. Celle-ci, du seul fait de sa présence, désigne à l'attention certains êtres et objets possédant un statut spécial dans un contexte donné. Cette caractéristique repose sur trois propriétés au moins : le fait que le particulier fasse l'objet de fréquentes prédications, que le groupe de locuteurs éprouve le besoin de l'identifier ainsi, enfin, le fait que la langue ne propose pas de description définie plus économique. La fréquence des prédications suggère que le particulier en question revêt quelque importance dans un contexte donné ; l'adoption d'un nom particulier par le groupe de locuteurs au sein duquel il est évoqué,

correspond à la recherche d'un moyen simple de faciliter les échanges. De ce point de vue, le troisième critère est placé sur le même plan, en fonction du raisonnement implicite selon lequel c'est le moyen linguistique le plus accessible et le plus maniable qui doit alors être privilégié.

Le rappel de ces règles pragmatiques montre également que le lien de dénomination établi entre les noms propres et leurs référents ne constitue pas le moyen le plus économique d'un point de vue mémoriel pour référer à des objets. On utilise plutôt pour cela des mots désignant des concepts généraux. En effet, le « raccourci » que constituent les noms propres est contrebalancé par l'effort cognitif induit par leur apprentissage et leur stockage dans la mémoire longue des sujets. Cela explique pour une bonne part que leur usage, sauf exceptions qui, certes, peuvent paraître nombreuses à la lecture des dictionnaires de noms propres désignant des réalités dotées d'une importance culturelle certaine, n'a généralement de sens que dans le cadre d'un groupe plus ou moins étendu de locuteurs. C'est dans un tel cadre que la présence des noms propres semble constituer une piste pertinente donnant accès à un référentiel spécifique. Cette hypothèse comporte un volet ontologique qui n'est pas abordé dans ce travail ; la partie linguistique s'en tiendra à l'observation de l'usage qui est fait de ces unités dans le cadre d'un genre de discours. Ces observations devront donc être reconduites pour l'étude d'autres genres discursifs. En second lieu, on a formulé l'hypothèse selon laquelle les noms propres présents dans les discours rassemblés dans un corpus peuvent être utilisés afin de capter les éléments caractéristiques d'un vocabulaire. C'est à la vérification de ces deux hypothèses que l'on va s'attacher maintenant.

Afin d'évaluer la réponse que les textes du corpus peuvent apporter à la première hypothèse, on cherchera tout d'abord à établir précisément la place des noms propres dans leur vocabulaire, en termes de variété et de nombre d'occurrences. Cette étape constitue un préalable nécessaire. En effet, vérifier que les noms propres représentent bien en moyen d'accéder à certains pans d'un référentiel d'entreprise suppose, non seulement que leur présence dans les textes du corpus soit avérée, mais aussi qu'elle atteigne un niveau appréciable. Si les indications fournies par les deux précédentes parties permettent de penser que ces unités sont bien présentes dans les textes du corpus, leur recension doit reposer sur

une définition précise. On fera à ce sujet une distinction entre les dénominations propres et les unités dotées d'une majuscule initiale, à des fins de mise en valeur, « comme s'il s'agissait de noms propres ». Cette distinction opérée, le champ de la collecte visera à rendre compte du *continuum* s'échelonnant entre ces deux pôles.

Dans la mesure où cette collecte vise, entre autres, l'enrichissement d'un référentiel terminologique, elle devra permettre d'associer aux noms vedettes leurs variantes attestées dans le corpus. En effet, si les noms propres constituent une voie d'accès non négligeable à l'information, dès lors qu'ils appartiennent à un référentiel partagé entre émetteurs et destinataires¹, l'un des principaux obstacles à leur utilisation tient à la présence de variantes. Celles-ci seront néanmoins limitées aux unités attestées de façon récurrentes dans le corpus, car on sait, par ailleurs, que les variantes orthographiques accidentelles peuvent être fort nombreuses².

On exposera les questions soulevées par la réalisation de cette collecte et l'évaluation de la présence des noms propres dans les textes du corpus lorsque l'on recourt à des méthodes textométriques. En effet, on sait que le repérage des entités nommées est une tâche que l'on confie de plus en plus fréquemment à des programmes informatiques recourant majoritairement à un traitement automatique du langage naturel³. Ces programmes permettent d'obtenir des résultats remarquables à partir du moment où le logiciel a été « entraîné » sur un corpus de textes. En revanche, l'utilisation de ces techniques sur des corpus différents se révèle moins satisfaisante⁴. On notera seulement à ce stade que le recours à l'approche textométrique, plus robuste, ne vise pas seulement à garantir une meilleure portabilité dans un contexte complexe. On verra plus loin que l'enjeu concerne à

¹ M. Amar, dans sa thèse portant sur *les fondements théoriques de l'indexation*, écarte les noms propres comme unités susceptibles de représenter le contenu d'un document (descripteur). En revanche, l'auteur constate que les usages, représentés par plusieurs exemples de systèmes d'organisation de l'information, utilisent tous les noms propres comme moyen de réaliser l'indexation des contenus. Les classes privilégiées sont les noms de lieux et de personnes morales et physiques. M. Amar (2000).

² En septembre 2003, le moteur de recherche Google a communiqué les variantes du nom de la chanteuse Britney Spears saisies par ses utilisateurs. Ces variantes s'élevaient à six cents environ, allant de 40 134 occurrences pour « Brittany Spears » à 2 occurrences pour « Prittany Spears » ou « Briyany Spears ». La liste complète est donnée en annexe.

³ Un certain nombre d'expériences sont présentées, entre autres, dans B. Boguraev and J. Pustejovsky (1996).

⁴ Pour une présentation de l'état de l'art, on renvoie à T. Poibeau (2003 : 87-116).

titre principal l'adéquation de cette approche avec l'exploration des relations discursives établies entre les noms propres et les autres vocables du corpus.

Comme on vient de le voir, cette première étape pourrait constituer une fin en soi, dès lors qu'elle permet une utilisation améliorée des dénominations propres dans une perspective de recherche d'information. L'utilisation à bon escient des noms propres pour atteindre ce but est néanmoins dépendante de la connaissance que l'on en a. Ne serait-ce que pour cette raison, le recensement des noms propres gagne à être complété par celui d'autres unités qui leur sont rattachées dans les discours. C'est pourquoi, afin d'établir plus finement le rôle joué par les noms propres dans un référentiel spécifique, cette première collecte d'informations doit être précisée par une observation du rôle tenu par les noms propres dans les différents dispositifs du genre de discours étudié. Dès lors, la simple présence des noms propres, même si elle a déjà pu être documentée par les fréquences et la ventilation de ces dernières entre les différentes parties du corpus, doit être complétée par de nouvelles indications. En effet, si la classe des noms propres ne constitue pas un bloc monolithique, comment les textes du corpus mobilisent-ils telle ou telle sous-classe, comme les noms de lieux ou de personnes ? Par ailleurs, la répartition de cet ensemble d'unités entre composants de l'activité bancaire est-elle comparable à celle observée pour les unités ressortissant de la langue commune ou de la terminologie bancaire ? Ces interrogations conduisent à une étude du fonctionnement discursif des dénominations propres. C'est cette approche qui, semble-t-il, sera à même de valider ou d'invalider la seconde hypothèse formulée dans la première partie, concernant la capacité des noms propres à saisir des unités caractéristiques du vocabulaire d'un corpus.

Ce deuxième volet de la collecte, qui vise donc à utiliser les noms propres afin d'extraire des vocables caractéristiques du vocabulaire des rapports d'activité, doit permettre de constituer un référentiel terminologique adaptable au contexte. Cette approche nécessite, au préalable, que l'on puisse apporter des réponses aux nombreuses questions qu'elle soulève. Si toutes les sous-classes de dénominations propres ne sont pas amenées à jouer le même rôle dans les rapports d'activité, quelles incidences ont ces différences dans la propension de ces unités à réaliser une collecte productive ? Les relations établies dans les discours entre les noms propres

et les autres vocables sont-elles susceptibles d'une formalisation ; sur quelles bases faire reposer cette dernière, enfin, quel statut attribuer à ces relations ? Les réponses apportées à ces questions devront servir à guider la collecte, compte tenu des buts assignés à la mise en place d'un référentiel terminologique adaptable au contexte. En effet, on recherchera la ou les sous-classes susceptibles de fournir la collecte la plus productive et la plus représentative du vocabulaire du corpus. Car si la mise en place d'un référentiel terminologique ne vise pas une recension exhaustive du vocabulaire d'un corpus, il est d'autant plus nécessaire d'identifier des critères explicites destinés à orienter, et si besoin, à justifier, les étapes de la collecte.

Les réponses qui seront apportées à ces différentes questions convergent vers la construction d'un schéma du fonctionnement discursif des principales classes de noms propres présentes dans le corpus. De l'élaboration de cette construction découle la possibilité de la collecte, mais aussi sa valeur. De cette dernière dépend la qualité, mais aussi, la fiabilité des informations que l'on fera entrer dans le référentiel terminologique. On a indiqué plus haut, à propos du degré de figement de certains syntagmes, que les approches onomasiologique et sémasiologique sont concurrentes dans le travail de constitution de tels référentiels. L'introduction de noms propres dans un référentiel terminologique soulève de façon cruciale la question des données motivant l'établissement des relations sémantiques. Autrement dit, il sera nécessaire de discuter les bases d'un tel rattachement, sachant que celui-ci ne pourra pas avoir d'autre caution que discursive. La construction schématique du fonctionnement discursif de telle ou telle classe de dénominations propres devra donc être étayée par une analyse syntaxique à même de dégager les propriétés sémantiques justifiant l'établissement de relations entre unités. On étudiera alors plus particulièrement les manifestations linguistiques des opérations de catégorisation et de rattachement à des secteurs d'activité. En effet, ces deux sortes de relations semblent plus particulièrement désignées afin de guider la construction des arborescences des sites d'un intranet, ainsi que les regroupements thématiques de leurs contenus. Enfin, il conviendra de discuter le formalisme adopté pour la traduction de ces relations dans un référentiel terminologique.

Chacune des étapes de la collecte fera par ailleurs l'objet d'une évaluation. Celle-ci est non seulement nécessaire pour vérifier la fiabilité des informations

collectées, mais aussi pour permettre l'extension de cette approche à de nombreux corpus, ainsi que le partage des résultats obtenus pour des référentiels terminologiques différents. Les dénominations propres seront donc évaluées sous le rapport des relations qu'elles établissent avec les autres unités présentes dans le vocabulaire du corpus, mais aussi la productivité de la collecte qu'elles permettent d'obtenir. Par ailleurs, les résultats obtenus seront confrontés aux observations réalisées dans la précédente partie relativement aux variétés de langue présentes dans les textes du corpus. Ils seront également rapprochés de ce qu'il aurait été possible d'obtenir à l'aide d'un modèle construit a priori. Dans tous les cas, il s'agira de présenter les apports de l'approche préconisée.

On présentera donc tout d'abord les questions soulevées par l'identification et la collecte de dénominations propres à l'aide des techniques textométriques. La collecte des noms propres présents dans le corpus sera ensuite confrontée aux observations réalisées au sujet des autres composantes du vocabulaire (chapitre 8). Cette première étape sera approfondie par l'étude du rôle joué par une classe de dénominations propres dans le corpus. Les toponymes permettront ainsi de montrer que l'évaluation du rôle des dénominations propres doit être rapportée à l'usage qui est fait de ces unités dans le genre de discours étudié (chapitre 9). La notion de rôle devra être complétée par une construction reflétant le fonctionnement discursif des principales classes de noms propres présentes dans le corpus (chapitre 10). Cette construction schématique sera à même de guider les étapes suivantes de la collecte. Celles-ci consisteront à rechercher la manière dont certaines relations sont établies de façon récurrente entre ces unités et les autres éléments dont se compose le vocabulaire du corpus. Les résultats obtenus seront ensuite évalués au regard des exigences liées à la construction d'un référentiel terminologique adaptable au contexte (chapitre 11).

CHAPITRE 8
NOMS PROPRES EN DISCOURS

Les précédents chapitres ont permis de relever que les noms propres semblent occuper une place non négligeable dans le vocabulaire du corpus. Il s'agit maintenant de préciser ces observations, d'une part, en recensant les unités qui appartiennent à cette composante du vocabulaire et, d'autre part, en présentant les principales caractéristiques de cet ensemble. Cette recension des noms propres en discours¹ constitue la première étape de la démarche que l'on propose pour la construction d'un référentiel terminologique adaptable au contexte. On rappelle que ce dernier vise à documenter les activités d'information de deux manières au moins.

Il a pour but, tout d'abord, de faciliter l'usage des noms propres pour la recherche d'informations. On a rappelé plus haut que ces unités sont fréquemment utilisées dans ce cadre, aussi bien sur Internet² que dans les systèmes d'organisation de l'information. Cependant, dans une situation de déséquilibre cognitif, les noms propres possèdent un statut paradoxal. Ils combinent, en effet, deux propriétés contraires. D'une part, ils sont difficiles d'accès, parce qu'ils renvoient à une réalité encore mal connue par le néophyte et, d'autre part, ils constituent un moyen d'accès indirect, mais efficace, à ce qui est souvent difficile à verbaliser avec les expressions adéquates³, compte tenu de la raison qui vient d'être invoquée. De ce fait, cette collecte devra permettre de traiter deux points. Il s'agira tout d'abord de renforcer

¹ Le titre de ce chapitre est inspiré de celui de la journée d'étude organisée par le Céditec (2006), « Le nom propre en discours ».

² La lettre professionnelle *Recherche et référencement* du site Abondance (<http://www.abondance.com/>) a publié en février 2004 un bref article intitulé « L'année en mots clés selon Google » (disponible en juillet 2008 à l'adresse <http://actu.abondance.com/2004-02/google-zeitgeist.html>), qui présente les informations communiquées par le moteur de recherche Google à propos des recherches les plus fréquentes réalisées sur ce site pour l'année 2003. On remarquera la proportion écrasante de noms propres : « *Après Yahoo! la semaine dernière, c'est au tour de Google de nous fournir sa vision de l'année 2003 au travers des recherches effectuées tout au long de l'année sur son moteur. Le "2003 Year-End Google Zeitgeist" nous propose donc les mots clés les plus populaires de l'année (1. Britney Spears, 2. Harry Potter, 3. Matrix, 4. Shakira, 5. David Beckham), les nouvelles requêtes (1. Iraq, 2. Laci Peterson, 3. Kobe Bryant, 4. Bertrand Cantat, 5. RIAA), les marques (Ferrari, Sony, BMW), plus des infos sur la saisonnalité de termes comme Roland Garros, SARS, Iraq, carnaval, etc. Enfin, sachez qu'en France, le "Top 10" a été le suivant l'année dernière : 1. SNCF, 2. FNAC, 3. Star Academy, 4. TF1, 5. Matrix, 6. SFR, 7. Le Monde, 8. Britney Spears, 9. Air France, 10. Lorie.* » Les meilleurs scores de mots clés utilisés avec Google sont disponibles pour plusieurs pays et depuis l'année 2001 à l'adresse : <http://www.google.com/intl/en/press/intl-zeitgeist.html#fr>.

³ On pense, par exemple, à la littérature sur l'expression des besoins de l'utilisateur d'un système d'information. Pour un panorama de cette question, on renvoie à l'ouvrage de Y.-F. Le Coadic (2007).

l'utilité des noms propres comme critère de recherche, en recensant les dénominations propres attestées ainsi que leurs variantes. Ce type de ressource permet, en effet, de réaliser des expansions de requêtes, associant automatiquement à la dénomination propre soumise au moteur de recherche, ses différentes variantes. On vise ainsi l'amélioration du taux de rappel en permettant la sélection de certaines réponses où seules les variantes de la dénomination propre utilisée pour la recherche sont présentes. En second lieu, il conviendra de rendre ces dénominations plus accessibles aux néophytes, en les rattachant à des unités linguistiques autres, relevées dans leur voisinage textuel.

Une autre manière de documenter à l'aide des noms propres certaines activités d'information, comme la construction de l'arborescence de sites intranets, ou les regroupements thématiques de contenus, consistera à utiliser les informations collectées à partir de ces unités, afin de mettre en lumière les variations affectant l'usage d'un parler d'entreprise. Pour obtenir ce résultat, les noms propres présents dans les discours seront utilisés comme autant de points d'observation de leur cotexte. On disposera ainsi du moyen de collecter d'autres unités, terminologiques ou non terminologiques, en fonction de critères qui devront être explicités. L'organisation de ces données dans les référentiels terminologiques adaptables au contexte « exposera » alors des manières de dire et de classer propres à tel ou tel public d'un site. Cette image linguistique d'un public destinataire devra ensuite être injectée dans l'organisation des contenus, au moment de la réalisation de certaines activités d'information (voir le premier chapitre).

Ces deux utilisations complémentaires des noms propres rassemblés dans un référentiel terminologique ne nécessitent pas nécessairement le même type de collecte. Dans le premier cas, le recensement des noms propres et de leurs variantes peut suffire. On sera alors tenté de rechercher un recensement plutôt exhaustif. En revanche, pour satisfaire à la seconde utilisation, une sélection d'unités caractéristiques suffira peut-être à renseigner sur la manière d'utiliser telle ou telle strate d'un parler d'entreprise dans une situation de communication particulière. Par ailleurs, dans la première sorte de collecte les aspects morphosyntaxiques l'emportent (il s'agit de répondre à des questions telles que la délimitation exacte de ces unités dans leurs contextes syntactiques ou non syntactiques, ou encore,

l'association des sigles et de leurs développés), alors que pour le second type de collecte, il est nécessaire d'adopter une approche sémantique destinée à justifier l'établissement de relations formelles entre unités. Pour cela, cette deuxième collecte doit pouvoir s'appuyer sur une étude poussée du fonctionnement discursif des noms propres. Enfin, il faut aussi considérer que les deux collectes sont complémentaires par certains aspects : une collecte exhaustive est de nature à éclairer la place tenue par ces unités dans le vocabulaire d'un genre de discours ; or, et l'étude du fonctionnement discursif vient éclairer utilement certaines questions, comme celles soulevées par l'homographie ou encore, l'emprunt de signalétiques. On a donc opté dans ce chapitre et les suivants pour une approche ménageant les deux logiques à l'œuvre dans la collecte des noms propres du corpus. On abordera ici les questions soulevées par une collecte exhaustive, le fonctionnement discursif étant réservé pour les chapitres suivants.

Plusieurs techniques, plus ou moins outillées par des programmes informatiques, sont utilisables pour réaliser une telle collecte. On verra cependant qu'il y en a peu qui permettent d'embrasser tous les aspects qui viennent d'être évoqués. On proposera ici d'utiliser l'approche textométrique, qui a paru la plus portable et la plus robuste. En effet, en se plaçant en amont des traitements automatiques du langage naturel, elle ne nécessite pas de réglages particuliers de ses programmes en fonction des corpus et des variétés de langue devant être étudiés ou exploités. Par ailleurs, cette approche permet un contrôle des observations et des résultats obtenus, comme cela a déjà été montré à l'occasion du test des contextes à dominante syntactique et non syntactique. Ses applications seront présentées à chaque étape, pour la collecte initiale, mais aussi pour l'étude du fonctionnement discursif des noms propres.

Le recensement des noms propres présents dans un corpus soulève plusieurs questions auxquelles on tentera de répondre dans ce chapitre. En effet, la sélection de ces unités doit reposer sur une définition explicite, ce qui conduira à discuter l'approche du nom propre dans la grammaire classique. On verra, en effet, que la relation instaurée par le lien de dénomination propre peut reposer sur d'autres unités que le seul substantif. Il conviendra ensuite de situer cette première délimitation par rapport au champ que l'on souhaite voir couvert par la collecte. On présentera à ce

propos l'intérêt qu'il y a à prendre en compte des unités pour lesquelles l'usage de la majuscule fonctionne seulement comme une marque distinctive n'entraînant pas le rattachement à la catégorie des noms propres. Enfin, seront exposés les traitements appliqués aux formes graphiques en fonction de leurs fréquences, ainsi que l'évaluation de la collecte qu'ils ont permis de réaliser.

Il sera donc possible d'aborder les caractéristiques de la présence des noms propres par rapport au vocabulaire et à la topographie du corpus. On cherchera alors à évaluer de façon globale les possibilités offertes par la collecte d'autres unités à partir des noms propres. On proposera ensuite de catégoriser cet ensemble afin d'évaluer les types de composants de l'activité bancaire qui bénéficient plus particulièrement de ce type de dénominations dans le corpus. Ce sera l'occasion de présenter un certain nombre de repères facilitant la réalisation d'une catégorisation manuelle pour un nombre important d'unités. Différentes formes de constructions des dénominations propres (signalétiques) seront alors exposées, ainsi que l'existence d'un *continuum* homonymique particulier assez productif dans les textes du corpus.

I. RECENSEMENT DES NOMS PROPRES DU CORPUS

A. Les approches antérieures et leurs apports

1. La statistique lexicale : noms propres ou dénominations propres ?

Le recensement des noms propres dans un corpus de textes soulève des questions de deux ordres différents mais qui sont néanmoins interdépendantes. Comment caractériser les unités dont on veut faire la collecte et comment adapter les moyens informatiques sur lesquels on va s'appuyer pour réaliser cette tâche, tout en étant respectueux des conséquences de la définition que l'on aura retenue ? Pour aborder les différents points qu'appellent les réponses à ces questions on a commencé par présenter deux approches qui abordent les moyens de la collecte en s'appuyant sur des définitions des sortes d'unités collectées qui sont soit implicites, soit restrictives. La discussion de ces limitations permettra d'aborder dans un deuxième temps la caractérisation des noms propres que l'on a souhaité retenir dans le cadre de ce travail. Les moyens mis en œuvre ainsi que les résultats obtenus seront abordés dans la section suivante.

L'étude des dénominations propres en corpus ne va pas de soi, et l'usage dominant jusqu'à une époque récente semble s'être limité à l'établissement des *index nominum* figurant en fin d'ouvrage, ces listes de noms propres classés alphabétiquement et associés aux numéros de pages correspondant à leurs occurrences. Il est par conséquent nécessaire de commencer par situer la démarche que l'on souhaite adopter par rapport à deux orientations principales relativement proches apparues antérieurement.

L'approche textométrique qui a déjà été retenue précédemment pour l'étude de certains aspects du vocabulaire du corpus, semble également adaptée pour le recensement des noms propres. En effet, en prenant le parti de faire reposer ses traitements sur des formes graphiques brutes, elle rend l'exploitation du corpus indépendante, au moins dans un premier temps, des questions relatives à la lemmatisation et à la désambiguïsation. Elle offre ainsi une approche robuste, car portable d'un corpus à l'autre. Cela est d'autant plus appréciable en ce qui concerne

les noms propres, que les analyseurs morpho-syntaxiques généralistes ont tendance à assimiler dans des proportions relativement importantes à des mots inconnus¹. Cependant, l'étude des noms propres occupe une position assez marginale dans les travaux issus de la statistique lexicale et de la textométrie. Toutefois, la présentation de certains travaux permettra de préciser ce que recouvre la notion de nom propre dans un tel cadre et de vérifier son adéquation avec les buts poursuivis dans cette étude.

Au début des années 1950, les travaux pionniers de P. Guiraud en matière de statistique lexicale avaient pris le parti d'un recensement systématique des mots apparaissant dans les textes, et les noms propres n'en avaient donc pas été exclus a priori. Cependant, ces derniers n'avaient pas été retenus dès lors qu'il s'était agi d'identifier les mots-clés et les mots thèmes, mots jugés les plus caractéristiques d'une oeuvre². Cet ostracisme ne se justifiait pas pour des raisons statistiques, les noms propres étant présents dans des proportions parfois importantes, mais il résultait d'un implicite. En effet, ces travaux ayant pour cadre l'étude de la langue, tout ce qui sortait du champ du lexique en était tacitement exclu.

On retrouve une position similaire dans les travaux de C. Muller, pour qui les noms propres permettent de définir le lexique commun d'un corpus : « (...) *cette catégorie [les noms propres] devra être délimitée et, dans la plupart des cas, traitée à part du reste ; ce reste, nous l'appellerons alors, faute d'une meilleure dénomination, le « vocabulaire (ou lexique) commun »*³. » Cependant, C. Muller prend le soin d'argumenter sur ce point. Ses premiers travaux ayant porté sur certaines œuvres du théâtre classique français⁴, l'auteur insiste sur le fait que les noms propres de personnages ploient sous une charge thématique écrasante, et constituent donc des unités sans surprise dès lors qu'il s'agit d'opposer les caractéristiques d'œuvres ou de genres différents⁵. On comprend bien, en effet, que les noms des principaux personnages suffisent généralement à désigner une unique

¹ D. Maurel (2004).

² P. Guiraud (1953).

³ C. Muller (1967 : 17).

⁴ *Essai de statistique lexicale – L'illusion comique de Pierre Corneille* (1964), et *Etude de statistique lexicale – Le vocabulaire du théâtre de Pierre Corneille* (thèse publiée en 1967).

⁵ C. Muller (1967 : 90 et 155-157).

pièce, leur réapparition dans plusieurs œuvres d'un même auteur étant l'exception. En revanche, le phénomène est plus fréquent pour les seconds rôles. Or, C. Muller soulève une deuxième objection sur ce point précis : tel nom passe-partout sert-il à désigner le même personnage d'une pièce à l'autre ? L'auteur en doute et en déduit que, si le décompte à l'échelle d'une pièce ne pose pas de difficulté, le problème des « appellations répétées » est de nature à grever les décomptes réalisés à l'échelle de toute une œuvre¹. Ces deux caractéristiques l'ont conduit, par exemple, à ne pas retenir les noms propres dans l'étude de la répartition des vocables entre les différentes pièces du théâtre de Corneille².

Cet ostracisme qui frappe les noms propres dans les études statistiques n'a pas échappé à É. Brunet dont les travaux sur les textes littéraires des XIX^e et XX^e siècles ont progressivement accordé une place de plus en plus importante à ces unités. Entre un article inaugural sur *Les noms propres dans l'œuvre de Chateaubriand* que son auteur présente comme une gageure³, et les développements substantiels qui leur sont consacrés dans le *Vocabulaire de Victor Hugo*⁴, un souci de réhabilitation de l'étude des noms propres traverse les travaux d'É. Brunet : « A-t-on remarqué que la tradition exclut les noms propres des calculs, et que l'étendue d'un texte et l'étendue d'un vocabulaire sont mesurées comme si les noms propres n'existaient pas ? Les noms propres, même quand ils naissent sur le sol national, n'apparaissent jamais comme intégrés à la loi et à la langue commune. Individualistes, attachés aux particularismes thématiques, ils sont à l'écart de la société lexicale, et si on les catalogue dans les index et les encyclopédies, on les exclut des dictionnaires et des études quantitatives. Cet ostracisme est-il légitime ? Faut-il que les recensements de la population lexicale tiennent compte des mots étrangers, des immigrés, des apatrides, des asociaux⁵ ? » Si la réponse ne fait pas de doute, elle appelle en outre une autre question portant cette fois sur la morphologie des noms propres.

¹ *Ibid.* p. 31-32.

² *Ibid.* p. 163.

³ É. Brunet (1979: 94).

⁴ É. Brunet (1988: 307-341).

⁵ *Ibid.* (1988: 319, note 2).

En effet, les travaux quantitatifs qui viennent d'être cités reposent sur l'adoption de normes, indispensables au traitement des données textuelles par des programmes informatiques. Ces normes sont variables selon les idiomes et les buts assignés à la recherche, mais il est toujours nécessaire de définir précisément ce que l'on compte. Dans le cadre de ces travaux, les comptages statistiques sont réalisés sur les mots des textes, qu'il s'agisse des formes fléchies brutes ou de leurs lemmes. Le mot *y* est considéré comme « (...) *une unité graphique, séparée des unités voisines par un blanc ou une ponctuation (...)*¹. » Cela n'est pas très gênant lorsque les noms propres sont appréhendés d'un point de vue grammatical comme une classe de substantifs. En revanche, les types de noms propres déjà rencontrés dans les précédentes parties – noms d'organisations, de produits ou d'artefacts pour ne citer que trois exemples – se présentent majoritairement sous la forme d'unités complexes qu'il s'agit de prendre en bloc, et non de réduire à des formes simples, comme la statistique lexicale le réclame. C'est pour se démarquer d'une conception grammaticale du nom propre que B. Bosredon et I. Tamba en viennent à parler de dénominations propres au lieu de noms propres : « *Dès lors, les grammaires aborderont normalement la question du Npr au chapitre du **substantif** ou **nom**. Mais les dictionnaires de noms propres, qui, dans notre terminologie, deviennent des dictionnaires de **dénominations propres**, enregistreront régulièrement toutes les formes fixes de dénominations d'un référent unique, françaises ou étrangères, sans considération de statut grammatical, qu'il s'agisse de sigles (C.G.T.), de noms propres (la Corse), de constructions diverses allant du syntagme (la Cousine Bette, roman de Balzac) à la phrase (Le jour se lève, film de Marcel Carné)*². »

En accord avec ce point de vue, le recensement des noms propres du corpus doit en réalité correspondre à un recensement de dénominations propres. Partant, il

¹ C. Muller (1992 : 4-5). C'est la même notion que l'on trouve définie autrement dans L. Lebart et A. Salem : « **Occurrence** : suite de caractères non-délimiteurs bornée à ses extrémités par deux caractères délimiteurs de formes. **Caractères délimiteurs / non-délimiteurs** : distinction opérée sur l'ensemble des caractères, qui entrent dans la composition du texte permettant aux procédures informatisées de segmenter le texte en occurrences. On distingue parmi les caractères délimiteurs : les caractères délimiteurs d'occurrence (encore appelés « **délimiteurs de formes** ») qui sont en général : le blanc, les signes de ponctuation usuels, les signes de préanalyse éventuellement contenus dans le texte (...). »

² B. Bosredon et I. Tamba (1994 : 123).

est nécessaire de trouver une solution au sein de l'approche textométrique qui permette la manipulation de telles dénominations. En effet, il ne s'agit pas de perdre les avantages de cette approche dès lors que l'on souhaite la mettre à contribution pour collecter des unités complexes. Certes, cela rend la tâche plus ardue, au point que C. Muller a pu voir dans le traitement des unités complexes une voie sans issue¹. Cette conception du « nom propre grammatical suffisant » est partagée par É. Brunet.

On empruntera à cet auteur un exemple, afin d'illustrer les conséquences de cette conception que l'on juge appauvrissante pour la collecte des dénominations propres. Lorsque cet auteur remarque au sujet des occurrences « (...) de *Victor* (1561) et de *Hugo* (858) ² » que ces deux mots réunissent 2419 occurrences, il comptabilise des noms propres grammaticaux. Lorsque l'on aborde la question en termes de dénominations, il est a priori nécessaire d'envisager les utilisations d'au moins trois et non de deux unités seulement. Il faut en effet prendre en compte non seulement les cas où ces mots sont bien utilisés pour désigner Victor Hugo, soit « Victor » et « Hugo » employés seuls, ainsi que le nom complet « Victor Hugo », mais aussi ceux où ils sont employés pour désigner d'autres individus par leur prénom « Victor » ou leur patronyme « Hugo »³. Dans les développements qui

¹ Concernant les unités qui nous intéressent, cet auteur est très clair : « *On rappellera pour mémoire le cas embarrassant des noms de nombres (cent dix, mais dix-neuf, etc.). Quant aux sigles (O.N.U., P.D.G., etc.), ils représentent généralement des mots dont la suite est fortement liée. Mais il est évident que si l'on décide de compter O.N.U. pour une seule unité, il faudra faire de même pour Organisation des Nations unies, qui deviendra ainsi distincte de Nations unies, et où le mot nations ne sera pas compté avec les occurrences du vocable NATION s.f. (...) Cela risque d'entraîner assez loin ; car, quand on nomme le comte Alfred de Vigny, on peut, suivant les situations, employer toute cette séquence, ou une partie (Alfred de Vigny), ou même l'un ou l'autre de ses éléments seulement : le titre, le prénom, ou le patronyme. S'agit-il de variantes d'une même unité de lexique, ou d'unités à mettre à la même mesure ?* » C. Muller (1992 : 18). On voit à partir de ce qui a été dit plus haut qu'il s'agit d'une impasse surtout lorsque c'est le Lexique de la Langue qui est visé et que la lemmatisation des vocables peut rendre la tâche exorbitante.

² É. Brunet (1988 : 320).

³ Cette question se pose évidemment dans des termes différents pour un mot donné en fonction du corpus. S'agissant des *Mémoires d'Outre-Tombe* É. Brunet note ainsi : « *Il est symptomatique qu'en face de Bonaparte ce soit le nom même de Chateaubriand qui vienne en second lieu avec 510 emplois, ce qui est surprenant de la part d'un auteur qui a choisi de parler à la troisième personne. Cette insistance montre non seulement le prix que Chateaubriand donne à sa famille, père, mère, frère et cousins, mais aussi l'attachement nobiliaire à un nom qui remonte à Saint Louis.* » É. Brunet (1979 : 92). Par ailleurs, dans le *Vocabulaire de Zola* É. Brunet mentionne explicitement ce qui constitue à ses yeux une double difficulté : « Eugène Rougon » n'est pas vu comme un ensemble du fait de la norme de segmentation qui ne tient compte que des chaînes de caractères délimitées par

suivent, les appellations « noms propres » et « dénominations propres » seront considérées comme synonymes. L'exploitation des données du corpus nécessitera donc, d'une part, d'identifier les dénominations et toutes leurs variantes, et d'autre part, de manipuler les dénominations complexes de la même manière que les dénominations simples, afin de pouvoir additionner les occurrences ou les distinguer en cas de besoin.

Enfin, les travaux réalisés dans le cadre de la statistique lexicale permettent d'éclairer un dernier point concernant la présence des dénominations propres dans un corpus électronique. La relation entre le recensement des noms propres et la saisie d'un référentiel spécifique, comme celui d'un auteur, a déjà été exprimée par É. Brunet en ces termes : « *Il nous a semblé que les relevés statistiques pouvaient rendre service dans le cas-limite des noms propres. Il ne nous échappe pas qu'il s'agit là d'une gageure, où les contraintes thématiques s'imposent bien lourdement. Il n'en reste pas moins que l'étude des noms propres – du moins des noms réels – permet une approche – et presque une mesure – de la culture d'un écrivain et de la place qu'il se donne dans le système référentiel du temps et de l'espace¹.* » Cependant, le même auteur attire l'attention sur le fait que cet accès est variable selon les auteurs, les genres et les époques auxquels se rattachent les textes d'un corpus.

Dans *Le vocabulaire de Victor Hugo* É. Brunet compare le nombre de noms propres différents au nombre d'unités lexicales du vocabulaire commun² dans les œuvres de quatre auteurs, Victor Hugo, François-René de Chateaubriand, Émile Zola et Marcel Proust. Il constate : « *alors que la proportion est d'environ 15% chez Proust et Zola, elle dépasse le tiers chez Chateaubriand et s'approche de la moitié chez Hugo³.* ». Par ailleurs, il apparaît que, pour chaque auteur, cette proportion

deux séparateurs (ici les espaces) ; par conséquent il n'est pas possible de dire à partir de la liste des noms propres si « Rougon » s'applique dans les textes à Pierre, Eugène, Sidonie ou Pascal. Il en va de même pour les prénoms. L'auteur reconnaît donc que la population des noms propres recensés n'a pas pu être désambiguïsée. É. Brunet (1985 : 275).

¹ É. Brunet (1979 : 94).

² C. Muller (1967 :17) « *cette catégorie [les noms propres] devra être délimitée et, dans la plupart des cas, traitée à part du reste ; ce reste, nous l'appellerons alors, faute d'une meilleure dénomination, le « vocabulaire (ou lexique) commun ».* »

³ É. Brunet (1988 : 311).

varie selon les genres et les périodes. Ainsi, pour Victor Hugo, les noms propres sont plus fréquents dans le théâtre et la correspondance, qui s'opposent sur ce terrain au roman et à la poésie, même si cette opposition doit être nuancée en tenant compte des spécificités des œuvres et des périodes. En dehors de la littérature, les textes de vulgarisation scientifique seraient d'une manière générale moins accueillants pour les noms propres que, par exemple, les textes journalistiques. Autrement dit, il n'est pas possible de postuler a priori une présence égale des noms propres dans tous les textes. Cela indique la nécessité d'un recensement relativement complet, si ce n'est exhaustif. Dans tous les cas, il faudra que cette présence puisse être évaluée en fonction des besoins de la collecte.

2. Le repérage automatique des entités nommées : catégorisation et majuscule

De nombreux programmes destinés au repérage des noms propres dans des corpus et reposant sur un traitement automatique du langage naturel sont apparus depuis le début des années 1990. Les tâches confiées à ces automates peuvent être résumées ainsi : « *l'identification de noms propres connus ou la découverte des nouveaux noms propres, et leur catégorisation*¹. » L'extraction complète consiste alors à stocker le résultat des opérations de reconnaissance et de catégorisation dans une base de données. Ces techniques ont été particulièrement développées dans le cadre des conférences *MUC (Message Understanding Conference)* aux Etats-Unis de 1987 à 1998² pour le repérage des entités nommées (*named entities*). Celles-ci correspondent, dans la tâche « *ENAMEX* » des *MUC*, aux noms de personnes, d'entreprises et de lieux présents dans un texte. On retiendra que ces techniques qui s'appuient sur des méthodes statistiques, linguistiques ou d'apprentissage, ou sur une combinaison de techniques, ont atteint une maturité certaine, même si leur diffusion hors des laboratoires reste encore timide : « *La reconnaissance des entités nommées à partir de textes écrits est actuellement la tâche d'extraction d'information qui obtient les meilleures performances. Les taux combinés de*

¹ B. Daille et E. Morin (2001 : 602).

² T. Poibeau (2003 : 87-102 et 187-189).

précision et de rappel sont comparables à ceux des humains, avec un taux de l'ordre de 0,90 P&R¹ sur des dépêches journalistiques². »

Cette approche, pour aboutie qu'elle soit, pêche néanmoins par plusieurs aspects. Une difficulté importante tient à la faible portabilité des solutions d'un corpus à l'autre : dictionnaires et grammaires locales utilisées pour le repérage valent pour un corpus donné mais sont difficilement transposables sans modifications sur un corpus aux propriétés différentes, que celles-ci portent sur le registre de langue (des dépêches d'agence très homogènes opposées à des mails relativement informels) ou le domaine (le général opposé au spécialisé). Or, dans le contexte qui a été retenu ici pour l'adaptation d'un référentiel terminologique, il faut envisager que pour un site donné plusieurs corpus soient nécessaires, certains partagés et d'autres spécifiques, le nombre de sites étant de plusieurs dizaines.

Un second point se révèle moins adapté encore et concerne la catégorisation a priori des entités nommées. Outre que le champ des entités peut se trouver de cette manière artificiellement restreint (les trois classes *ENAMEX* des conférences *MUC* – personnes, lieux ou organisations – peuvent ainsi être opposées à la petite centaine de classes recensées par J.-L. Vaxelaire³), la catégorisation présente l'inconvénient d'être établie à l'avance, sans tenir compte de la catégorisation effectivement constatée dans le corpus étudié, celle-ci étant ramenée au rang se scorie. Or, c'est précisément l'une des informations que l'on cherchera à obtenir : comment telle situation de communication modifie-t-elle la manière de catégoriser un même objet et quels mots sont alors utilisés ? Il reste néanmoins possible d'utiliser des catégorisations a priori dès lors qu'elles ont une fin heuristique destinée principalement au regroupement de centaines d'unités qui, sans cela, formeraient une masse de données difficilement manipulable. Enfin, mais il s'agit d'une

¹ La précision correspond au nombre de réponses correctes par rapport au nombre de réponses fournies par le repérage (correctes et incorrectes). Le rappel correspond au nombre de réponses correctes trouvées par rapport au nombre de réponses attendues. L'ensemble des réponses correctes ayant été déterminé manuellement au préalable. La formule P&R permet de faire une synthèse de ces

deux mesures :
$$P\&R = \frac{2 \times \textit{précision} \times \textit{rappel}}{\textit{précision} + \textit{rappel}}$$

² T. Poibeau (2003 : 88).

³ J.-L. Vaxelaire (2005 : 312-398).

limitation qui tient plus aux finalités de ces extractions (veilles industrielle, économique, policière ou militaire¹) qu'à un défaut interne, le repérage automatisé utilise plutôt le contexte² comme moyen pour détecter les noms propres et les classer, alors que l'on cherchera dans cette étude à collecter certaines unités appartenant au voisinage du nom propre. Il reste que cette approche demeure éclairante sur plusieurs points.

En effet, cette approche impose tout d'abord une démarche d'évaluation de la collecte de noms propres, quel que soit le moyen utilisé, automatisé, semi-automatisé ou manuel. Dans l'approche manuelle ou semi-automatisée que l'on a retenue, elle permet de définir des critères partageables autorisant les comparaisons et les vérifications. Par des sondages, on dispose ainsi du moyen de décider de façon relativement objective de la complétude de la collecte effectuée en fonction des buts à atteindre et des contraintes du moment.

En second lieu, elle permet d'avoir un retour d'expérience documenté sur l'utilisation de la majuscule comme critère pour le repérage des dénominations propres. D'un point de vue théorique, on ne peut que souscrire à la remarque J.-L. Vaxelaire : « *Nous avons déjà noté que la majuscule ne pouvait être un critère fiable, on ne peut donc avoir de résultats réellement satisfaisants en se fondant uniquement sur celui-ci*³. » Dans une perspective terminographique ou lexicographique, il est cependant nécessaire d'aller plus loin. En dehors du reproche d'eurocentrisme linguistique, qui n'est pas rédhibitoire lorsque les langues des corpus étudiés sont le français et l'anglais, les cas de noms propres sans majuscule initiale se révèlent plutôt rares et relèvent en tout cas de l'exception⁴. C'est

¹ Les MUC ont été sponsorisées par la DARPA (Defense Advanced Research Projects Agency) ; voir aussi T. Poibeau (1999).

² Certains auteurs utilisent le contexte pour préciser le rattachement d'un nom propre à une catégorie générale, mais le contexte est restreint à la phrase. Par exemple, dans la phrase « *An industry analyst, Robert B. Morris III in Goldman, Sachs & Co.'s San Francisco office, said...* », « San Francisco » est identifié dans un premier temps comme une combinaison de deux mots inconnus, dans un deuxième temps comme un bloc, ou « *polyword* », de deux unités, étiqueté comme nom propre, et dans un troisième temps, la présence de « *office* » permet à l'automate de préciser qu'il s'agit d'un nom propre de lieu. D.D. McDonald (1994 : 33-37).

³ J.-L. Vaxelaire (2005: 690).

⁴ J.-L. Vaxelaire (2005: 73-75).

probablement pourquoi la majuscule est utilisée de manière systématique comme premier élément déclencheur pour le repérage des noms propres, dans ce cadre, mais aussi dans celui de la textométrie¹. On l'utilisera de la même manière, ce qui nécessitera, par exemple, de départager les noms propres des débuts de phrase. En revanche, on pourra se montrer plus accueillants pour l'emploi de la majuscule initiale lorsqu'elle est utilisée pour valoriser des unités terminologiques ne correspondant pas a priori à des dénominations propres. L'importance que revêt le cadre pragmatique pour la création de ces unités invite en effet à ne pas négliger ces marques graphiques. Comme certains auteurs y invitent, il est en effet possible de considérer que l'on affaire à un *continuum* entre taxon et noms propres au sein d'une même terminologie, ou mieux, au sein d'un même parler d'entreprise². Il n'en reste pas moins qu'une frontière existe entre termes et dénominations propres³.

Enfin, les traitements automatiques du langage, qui doivent affronter la diversité des discours et le caractère protéiforme des unités linguistiques, ont permis d'identifier des difficultés plus réelles que celles liées directement à l'emploi de la majuscule. C'est le cas pour la délimitation gauche du nom propre, qui pose la question de l'inclusion ou non du nom commun classificateur, comme pour les odonymes (« rue de la Paix », « comte Alfred de Vigny » pour reprendre l'exemple donné précédemment). La délimitation droite peut également se révéler difficile à identifier, comme pour ces exemples de noms d'entreprises ou de directions au sein d'une entreprise : « Relations avec les Caisses régionales » ou « Caisse régionale de Crédit agricole Anjou Mayenne ». D'une manière générale, il s'agit de prendre en compte les questions relatives aux différentes sortes de variantes d'une même dénomination, comme celles qui ont été recensées par B. Daille et E. Morin :

- « - les variations graphiques : *Parti Socialiste, Parti socialiste, parti socialiste ; École Normale Supérieure, Ecole normale supérieure ;*
- les sigles ou les abréviations : *École Normale Supérieure → ENS ;*

¹ É. Brunet (2001) *Hyperbase Logiciel documentaire et statistique pour la création et l'exploitation de bases hypertextuelles, Manuel de référence, version 5.2.*

² J. Humbley (2006 : 107-124).

³ Voir *infra*.

- *les variations morphosyntaxiques : les habitants de Nantes → les Nantais, la politique de Jospin → la politique jospinienne ;*
- *les coordinations : le couple Montand-Signoret, le Grand et le Petit Palais ;*
- *les ellipses : École Normale Supérieure → Normale, Normale sup ;*
- *les métaphores : l'Everest → le toit du monde¹. »*

Il convient donc d'adopter des procédures de collecte qui limiteront autant que faire se peut les oublis et les pertes dus à ces phénomènes de variation assez étendus. Par ailleurs, on voit déjà en quoi la possibilité de ces variations justifie la nécessité de collecter et de relier entre elles les variantes d'une même dénomination.

B. Statut des dénominations propres et périmètre de la collecte

1. Questions soulevées par la détermination du périmètre de la collecte

a Fonctionnement sémantique des noms propres prototypiques

Comme on vient de le constater, la prise en compte des noms propres tels qu'ils se présentent dans les textes rend nécessaire de les envisager comme des dénominations propres et non comme une catégorie restreinte aux unités correspondant aux traditionnels noms propres grammaticaux. Il est cependant nécessaire d'aller plus loin et de tenter de préciser ce que l'on entend par dénominations propres afin de préciser le périmètre de la collecte et de cerner le statut de ces dénominations en terminologie.

L'abondante littérature linguistique relative aux noms propres qui a paru depuis quelques dizaines d'années, au moins pour ce qui concerne le domaine français, peut être analysée comme un symptôme de la difficulté qu'il a à définir linguistiquement une catégorie d'unités hétérogènes². Pourtant, cette difficulté paraît presque inévitable, si l'on admet que toute unité linguistique peut faire office de

¹ B. Daille et E. Morin (2001 : 608).

² Pour un état complet de la question, on renvoie à la thèse de G. Kleiber publiée en 1981 et à la synthèse la plus récente des travaux de G. Kleiber (2004), ainsi qu'à la thèse de J.-L. Vaxelaire publiée en 2005. On trouve dans ces travaux une bibliographie très complète sur la question des noms propres.

nom propre dès lors qu'on lui assigne pour fonction d'être la forme fixe de la dénomination d'un référent unique. De ce fait, on a affaire à un ensemble ouvert, culturellement normé et empruntant sa morphologie et sa syntaxe à la langue dans laquelle ses éléments sont apparus. De même, si l'on admet que tout référent préalablement catégorisé est susceptible de porter un nom propre, on a à nouveau affaire à un ensemble ouvert, où la norme sociale fournit les représentants typiques en même temps qu'elle est en permanence remise en question par les exceptions. Il convient néanmoins de dépasser ce constat général, ne serait-ce que pour délimiter le champ de la collecte et examiner les conditions d'un rattachement de ces unités à un référentiel terminologique.

Cette dernière tâche nécessite que l'on précise la position adoptée dans ce travail au sujet du fonctionnement sémantique des noms propres. On essaiera de limiter l'examen de ce point au strict nécessaire pour le cadre de cette étude, sans aborder les réflexions encore en cours sur un sujet amplement discuté en France, au moins depuis la thèse de Georges Kleiber¹. Dans ce travail, on admet que la catégorie des dénominations propres possède des éléments prototypiques, les noms propres de personnes et de lieux, mais aussi d'autres éléments moins prototypiques dont le statut est parfois mal assuré. Pour le premier ensemble, on reprendra la position défendue par M.-N. Gary Prieur, qui s'appuie elle-même en partie sur les travaux de G. Kleiber, tout en proposant une notion utile, dès lors que l'on vise une étude des dénominations propres en contexte. Pour M.-N. Gary Prieur, il convient de distinguer le sens du nom propre, qui correspond à ses propriétés lexicales, et son contenu, qui ajoute à ce sens premier certaines propriétés de son référent.

Selon G. Kleiber, les noms propres possèdent un sens dénominatif qualifié d'« instructionnel », par opposition à « descriptif » ou « conceptuel » : *« Ce sens dénominatif n'est alors plus perçu comme une propriété ou description du référent, mais comme l'instruction de chercher et de trouver dans la mémoire stable le référent qui porte le nom en question. Le point essentiel dans l'affaire est que la paraphrase descriptive de ce sens ne peut plus être considérée comme une expression synonyme : de même que l'individu qui prononce (cette occurrence de)*

¹ G. Kleiber (1981) pour la version publiée.

« je » n'est pas sémantiquement équivalent à je, de même si je a pour instruction de prendre en compte celui qui a prononcé l'occurrence de je, l'individu appelé Charles n'est pas sémantiquement égal à Charles, même si Charles a pour sens instructionnel de prendre en compte le porteur de ce nom. Autrement dit, tout comme les symboles indexicaux et d'autres expressions du même acabit procédural, les noms propres sont irréductibles : ils ne se laissent pas réduire à une description correspondant à la donation dénomminative utilisée¹. » M.-N. Gary Prieur met en perspective cette conception du sens du nom propre relativement à une situation d'énonciation, en précisant qu' « (...) il consiste à donner l'instruction d'associer à la forme du nom propre un individu dont on sait, dans une situation d'énonciation donnée, qu'il est l'unique porteur du nom à prendre en considération. Par sa seule forme, donc, un nom propre établit une relation directe, dans une situation donnée, avec son référent². »

Ce sens instructionnel est complété par un contenu sur lequel on reviendra lorsque seront abordées les questions relatives à la collecte d'informations à partir des dénominations propres. On se contente de préciser à ce stade que ce contenu est défini à partir du fonctionnement discursif du nom propre et qu'il correspond à « (...) un ensemble de propriétés attribuées au référent initial de ce nom propre dans un univers de croyance³. » La notion de contenu semble a priori opératoire pour les noms propres prototypiques et ceux qui le sont moins. En revanche, il semble nécessaire de préciser le statut sémantique d'unités moins prototypiques rencontrées dans le corpus, comme cela a été indiqué plus haut.

b Fonctionnement sémantique des noms propres non prototypiques

Les noms de produits font partie des cas de noms propres « non-prototypiques » rencontrés dans le corpus. L'ampleur des débats relatifs à leur statut, à l'intérieur de la question plus générale des noms propres, semble signaler une

¹ G. Kleiber (1996 : 573).

² M.-N. Gary Prieur (2001 : 10).

³ M.-N. Gary-Prieur (1994 : 51). La notion d'univers de croyance est empruntée à R. Martin : « on appellera « univers de croyance » ou « univers » l'ensemble indéfini des propositions que le locuteur, au moment où il s'exprime, tient pour vraies ou qu'il peut accréditer comme telles. » (1992 : 38).

difficulté particulière. Comme l'indique J.-L. Vaxelaire¹, la majorité des auteurs ayant abordé la question propose de distinguer les noms de marques des noms de produits, les premiers ayant un statut de nom propre qui est refusé aux seconds. Cependant, à la fin de la section qu'il consacre au sujet, J.-L. Vaxelaire finit par mettre en doute la portée de cette distinction générale. Dans le cadre de ce travail, on admettra que les noms de produits possèdent un double statut. Le premier permet de les appréhender comme des termes se rapportant à un domaine spécifique et bénéficiant d'une définition précise. Conçus généralement au sein d'une collectivité organisée, les produits sont désignés alternativement par des descriptions définies intégrant le terme approprié, comme « le nouveau contrat d'assurance-vie multisupport », ou par des noms de code, valant aussi pour le projet dans le cadre duquel ils sont élaborés. Ce n'est que lorsque l'étape préalable à la commercialisation est atteinte que le nom du produit est définitivement fixé. Ce dernier est alors soumis à une double contrainte : il doit permettre de distinguer le produit de ses concurrents, sans pour autant le décrire.

Le nom du produit ne doit pas être descriptif au sens où il reprendrait une expression ordinaire pour caractériser celui-ci. C'est en quelque sorte le prix dénommatif à payer pour que la distinction puisse être réalisée. Les règles régissant le dépôt d'une marque sont à cet égard très claires :

« Ne choisissez pas certains signes, comme :

- un signe, un mot ou une expression qui sert à désigner votre produit ou service.

Exemple : le mot "baba" ne peut être déposé seul pour désigner de la pâtisserie ;

- un signe, un mot ou une expression qui décrit votre produit ou service. Exemple : l'expression "pure laine" seule ne peut être choisie pour un tapis en laine². »

Cette contrainte juridique bloque par conséquent la possibilité d'utiliser une expression spécialisée au non, proposant une description « claire » du produit.

En revanche, la dénomination propre constitue l'unité idéale permettant de respecter cette contrainte, tout en rendant possible une communication médiatique massive en direction d'un public de consommateurs non initiés. L'opacité de la

¹ J.-L. Vaxelaire (2005 : 348-351).

² In *La marque – Tout ce qu'il faut savoir avant de déposer une marque*, INPI, Paris, 2007, 18 p.

dénomination propre offre alors un moyen sûr de distinguer le produit parmi les éléments appartenant à la même classe. En effet, une dénomination propre est à même d'opérer une telle distinction du seul fait de son utilisation car, avant même d'orienter vers l'identification de son référent, elle est interprétée comme la « marque » d'une opération de dénomination ayant eu lieu préalablement à l'échange communicationnel. Le destinataire est informé par son truchement qu'il doit rechercher en mémoire stable le référent sur lequel l'énonciateur a souhaité attirer son attention. Il va sans dire que cet objectif est atteint au détriment de la description du produit, qui aurait pu être réalisée à l'aide de la motivation qui soutend ordinairement les unités terminologiques¹. C'est donc la contrainte de l'usage commercial qui est responsable de l'occultation du terme, désignation du concept, au profit de la dénomination propre correspondant à la marque du produit.

Dans les usages courants, le nom du produit demeure l'héritier de ce double statut et fonctionne, soit comme un terme plus ou moins banalisé, lorsque des instances du produit sont visées, soit comme un nom propre, lorsque c'est le produit portant telle marque précise qui est pointé. Il reste que les deux aspects sont liés et forment un tout du point de vue juridique. En effet, la dénomination propre est protégée contre les usurpations par un enregistrement auprès de l'Institut national de la propriété industrielle² – elle jouit alors du statut de marque déposée pour une période et un périmètre géographique donnés –, et la définition du concept doit respecter un cadre législatif précis destiné, entre autres, à protéger les acheteurs.

Afin d'illustrer ce double statut du nom de produit, on présente le fonctionnement de la dénomination du produit « Floriane » dans des extraits tirés du corpus et de deux autres sources, l'une technique et l'autre commerciale. Dans le

¹ On sait que les recherches de noms visent, entre autres, à compenser cette lacune en réintroduisant une forme de description du produit par le biais des connotations que les locuteurs associent à son nom. On renvoie pour ce point à P. Siblot (1995 : 147-160).

² « *Au sens de la propriété industrielle, la marque est un « signe » servant à distinguer précisément vos produits ou services de ceux de vos concurrents.* » Le formulaire de dépôt de marque précise par ailleurs qu'un tel « *dépôt de marque n'a pas vocation à protéger un nom de manière générale, mais un nom identifiant des produits et/ou des services.* » Ce document oppose sur la même page deux zones distinctes, le nom ou « *modèle de la marque* » et la catégorie à laquelle le produit se rattache. Pour caractériser celle-ci, les termes employés « *doivent permettre à toute personne d'en délimiter le contenu de façon immédiate, certaine et constante.* », in *Le formulaire marque – Comment remplir votre formulaire de dépôt de marque ?*, INPI, Paris, 2007, 22 p. Consulté le 13 mars 2008 à l'adresse : <http://www.inpi.fr/>.

rapport d'activité de 2001, la dénomination apparaît avec une extension attributive prenant la forme d'une apposition et pouvant faire office de définition : « *Floriane, contrat d'assurance-vie conçu par Predica et donnant accès à un fonds composé de 70 OPCVM gérés par Crédit Agricole Asset Management.* » Cette présentation se trouve en quelque sorte à mi-chemin entre une définition technique, comme on en trouve dans les fiches signalétiques des produits de la banque et de l'assurance, et une présentation commerciale destinée à la clientèle : « *FLORIANE est un contrat d'assurance sur la vie, à capital différé, à capital variable investi selon le choix du client :*

- *sur des supports financiers (obligations, actions, fonds de fonds)*
- *sur un fonds en euros¹.* » Sur un site commercial, les mêmes ingrédients bénéficient d'un traitement censé être plus accessible pour les destinataires :

« *Vous voulez diversifier votre patrimoine en combinant opportunités des marchés financiers et avantages fiscaux de l'assurance vie ? (...) Floriane est un contrat d'assurance vie en unités de compte, qui met à votre disposition une large gamme de supports. Ce contrat, particulièrement souple, vous permet de piloter vous-même la répartition de votre capital en fonction des évolutions de la Bourse et de vos objectifs patrimoniaux. (...)*

Avec Floriane, vous disposez d'une large ouverture aux marchés financiers notamment internationaux, grâce aux nombreux supports proposés : supports monétaires, obligataires, actions, fonds à formule, supports immobiliers et fonds de fonds.² »

Dans ces différentes citations on voit fonctionner côté à côté dénomination propre et périphrase définitionnelle, à la manière de ce que l'on peut rencontrer dans un ouvrage de vulgarisation scientifique pour un terme non encore banalisé. Si les périphrases définitionnelles possèdent des différences, il n'en reste pas moins qu'elles présentent toutes les principaux traits caractéristiques du produit et de son concept : un contrat d'assurance-vie offrant le choix des supports financiers servant

¹ Extrait de la fiche produit de Floriane (document interne).

² Extrait de l'argumentaire présenté sur le site Internet du Crédit Agricole de l'Anjou et du Maine (http://www.ca-anjou-maine.fr/Vitrine/CatalogueProduit/FichProd/P_FichProd_HdG_Floriane.jsp), consulté le 12 mars 2008.

de placements. On observe également que la reprise dans le dernier extrait se fait à l'aide du générique « contrat ». Mais il faut également convenir que le nom du produit est utilisé comme une dénomination propre, lorsqu'il occupe une position initiale dans les trois extraits.

Enfin, ces exemples rappellent aussi que la faible maniabilité¹ d'un terme complexe tel que « *contrat d'assurance sur la vie, à capital différé, à capital variable investi selon le choix du client* » ou encore « *contrat d'assurance vie en unités de compte* » est également susceptible d'orienter les utilisateurs initiés, comme les commerciaux chargés de vendre le produit, vers l'adoption de dénominations courtes, comme « contrat Floriane », dont le caractère hybride manifeste sous une autre forme le double statut du nom de produit.² Ces remarques incitent à adopter pour la collecte une conception englobante, capable d'embrasser un *continuum* allant des noms propres prototypiques aux termes, en passant par les dénominations propres d'artéfacts, tels que les produits bancaires dont il vient d'être question. Même si cette question fait toujours débat au sein de la terminologie, il semble néanmoins que cette discipline constitue un cadre accueillant pour une telle approche.

c Une approche terminologique globale

Les réflexions menées par certains auteurs, au sujet de l'intégration des noms propres dans le cadre de la terminologie, vont permettre de préciser l'option d'une approche englobante que l'on souhaite retenir pour la collecte.

Les normes qui encadrent le travail terminologique ne reconnaissent que deux sortes d'unités, les termes, désignant des concepts généraux, et les appellations, désignant des concepts uniques. Une partie des dénominations propres est donc exclue a priori : « *Une appellation désigne un concept unique, c'est-à-dire, un concept dont l'extension ne correspond qu'à un seul objet. En terminologie,*

¹ Sur cette notion, voir R. Kocourek (1991 : 140).

² G. Kleiber (1981 : 314-315) dans une analyse pragmatique des noms propres, note que ces derniers se substituent avantageusement aux descriptions définies pour une raison similaire : « *Lorsque le particulier abstrait est l'objet de fréquentes prédications, on peut éprouver le besoin, comme nous le verrons ci-dessous avec les particuliers spatio-temporels, de se dispenser de l'emploi, peu économique dans le cas de nombreuses répétitions, de la description définie qui délimite le particulier en question. Dans ces conditions l'emploi du nom propre s'impose.* »

l'accent est mis sur les noms des concepts uniques représentant des objets appartenant à un ensemble mais qui présentent une certaine individualité, par opposition aux noms qui désignent des individus en tant que tels (par exemple Marie, Pierre)¹. » Par ailleurs, on voit à travers les exemples² qui sont donnés par la norme, que la catégorie des appellations ne prend pas en compte les noms de produits.

*Tableau 47.
Exemples d'appellations de concepts uniques dans la norme ISO 704 :2000*

L'Assemblée nationale, le Ministère de l'agriculture	— institutions ou entités politiques spécifiques
Les Nations Unies, La Fondation de France	— organisations spécifiques
La Légion d'honneur, le Prix Nobel de la paix	— récompenses spécifiques
La comète de Halley, Saturne	— phénomènes scientifiques spécifiques
Walkman, Airbus, Kleenex	— noms de marque spécifiques

Néanmoins, l'introduction de la notion d'appellation ainsi que ces derniers exemples constituent une évolution importante par rapport à la précédente version de la norme datant de 1987, où le sujet était ignoré. En effet, le statut des noms propres en terminologie, jusqu'à une époque très récente, pouvait être résumé ainsi par R. Kocourek « *Ayant inclus les unités lexicales brachygraphiques, nous pouvons examiner un autre type d'expression habituellement éliminé des études linguistiques et terminologiques : celui des noms propres³.* » Or, cet auteur, qui plaide pour l'intégration des noms propres dans les terminologies, réclame également le développement d'une terminographie ayant pour socle les textes technoscientifiques. En effet, les deux semblent liés, car une approche qui prend appui sur les textes produits dans le cadre d'un domaine a nécessairement affaire avec ce type d'unités : « *Les noms propres des chercheurs, des inventeurs, des rédacteurs, des organismes, des ouvrages, des localités, des usines, des animaux, etc. apparaissent dans les textes, sans parler de l'emploi des noms propres dans les bibliographies, qui sont l'un des lieux de concentration des Propria. D'autres unités lexicales sont apparentées aux noms propres. Les éponymes, c.-à-d. les unités lexicales dont la*

¹ NF ISO 704 :2000, 7.4. p. 29.

² *Ibidem.*

³ R. Kocourek (1991 : 94).

source est un nom propre, sont assez nombreux (...). On trouve également un très grand nombre de noms de marque et d'autres appellations de produits, de modèles, etc.¹. »

À travers les exemples donnés ici, on perçoit que R. Kocourek adopte implicitement l'idée d'une sorte de *continuum* pouvant aller du nom d'un chercheur, nom propre de personne appartenant au premier groupe prototypique, au nom de produit ou de modèle, dont l'appartenance à cette classe d'unités ne fait pas l'objet du même consensus. Confirmant sa pensée, R. Kocourek admet « *les noms propres comme faisant partie du lexique technoscientifique soit en qualité de noms propres terminologiques, unidisciplinaires ou interdisciplinaires, soit en qualité de noms propres non terminologiques, plurivalents mais précis². »*

Dans un article plus récent³, J. Humbley brosse un bilan du statut des noms propres dans la théorie, la littérature scientifique et la pratique terminologiques, pour parvenir à une vision assez voisine de celle de R. Kocourek, en proposant une conception scalaire de la terminologie, qui serait capable d'intégrer certaines catégories de noms propres parmi les unités terminologiques. L'auteur constate à l'aide de plusieurs exemples, noms d'objets célestes, d'institutions et de toponymes, que les frontières théoriques séparant les *nomens* ou taxons, relevant de la terminologie, et les noms propres s'estompent dès lors que ces unités entrent dans le champ des pratiques terminologiques. C'est du reste ce que l'on peut constater à travers les nomenclatures de dictionnaires spécialisés ou les bases de données terminologiques, qui mêlent termes et noms propres. J. Humbley suggère que ces réflexions pourraient être prises en compte par la théorie en élargissant « *la définition du terme pour englober toutes les expressions qui véhiculent une connaissance spécialisée, qui nécessitent un apprentissage particulier et qui ne relèvent pas des compétences linguistiques du locuteur mais des capacités cognitives. Ce sont les mots qui nécessitent le cas échéant des ajustements pour*

¹ *Ibid.* (1991 : 96). Ce constat est également fait par Dardo de Vecchi dans sa thèse (1999).

² R. Kocourek (1991 : 97).

³ J. Humbley (2006 : 107-124).

ménager les relations entre ces connaissances et leurs expressions linguistiques ou autres¹. »

La présentation de ces points de vue confirme donc, s'il en était encore besoin, que ni la catégorie des noms propres, ni celle des termes ne sont homogènes. Il faut donc convenir que chacune fonctionne sous la forme d'un *continuum* avec des éléments plus ou moins prototypiques. En dehors de la position extensive de R. Kocourek, l'approche terminologique admet que certaines dénominations propres puissent entrer dans son champ, mais elle répugne néanmoins à admettre les noms propres prototypiques. Cette position est éclairée par ce qui a été dit plus haut : le sens instructionnel n'étant pas un sens de nature conceptuelle, les noms propres prototypiques ne peuvent pas désigner un concept unique. Il reste que ce clivage, justifié sur un plan théorique, se heurte à l'intuition que l'on a néanmoins affaire à un même ensemble, dès lors que l'on adopte un point de vue communicationnel. C'est parce qu'il part des textes, c'est-à-dire de productions s'insérant dans un dispositif communicationnel donné, que R. Kocourek adopte une conception extensive de l'intégration des noms propres dans le champ terminologique. C'est la même raison, mais vue sous un autre angle, qui pousse J. Humbley à considérer que font partie du même ensemble « *les expressions qui véhiculent une connaissance spécialisée, qui nécessitent un apprentissage particulier²* ». Il restait à traduire cette approche globale dans la procédure de collecte.

Pour que la collecte soit réalisée de manière fiable et contrôlable, il est nécessaire de partir d'un critère simple, exploitable par les programmes informatiques restant dans le cadre textométrique. On a vu dans la section précédente que les approches automatisées d'identification et de collecte des dénominations propres prennent généralement appui sur la présence d'une majuscule initiale pour faire le départ entre les unités qui vont potentiellement entrer dans cette catégorie et celles qui vont en être exclues. On reprendra donc ce critère graphique comme point de départ. Si cette approche empirique permet une avancée

¹ *Ibid.* (2006 : 122).

² Référence donnée plus haut.

concrète, elle présente néanmoins l'inconvénient de faire dépendre la collecte de l'utilisation de la majuscule dans les textes du corpus. On a vu que l'approche est acceptable pour des langues comme le français et l'anglais qui sont présentes dans le corpus utilisé. Cependant, la majuscule n'est pas seulement destinée à marquer les noms propres : les débuts de phrase ou de vers, les énumérations ordonnées, les citations complètes, les titres, les intitulés de colonnes de tableaux, les légendes de graphiques, etc., produisent également des unités possédant une majuscule initiale dans les textes. D'autre part, il faut aussi envisager que les règles typographiques ne sont pas toujours respectées de manière exemplaire : certains signes de ponctuation, lorsqu'ils ne terminent pas une phrase (deux-points, points d'interrogation, points virgule, points de suspension), ne sont normalement pas suivis d'une majuscule, mais cette règle, parmi d'autres, est souvent ignorée¹. Enfin, on sait que l'usage « *tend à employer les majuscules pour mettre en valeur des « mots importants », souvent estimés tels par la subjectivité du scripteur (la presse abuse parfois de ce procédé)*² » ; on ajoutera que les textes relevant de la communication institutionnelle reprennent facilement à leur compte cette forme de mise en valeur très fréquente dans le domaine de la publicité. L'identification des dénominations en contexte devra donc faire l'objet d'un contrôle strict.

On fera néanmoins une distinction entre les deux sortes de phénomènes qui viennent d'être évoqués. Autant les nécessités typographiques ou l'ignorance de certaines règles sont susceptibles de produire des unités qui ne seront pas retenues pour la collecte, autant la valorisation à l'aide de la majuscule de certains mots jugés importants par l'émetteur pourra entraîner leur sélection, tout au moins dans un premier temps. En effet, lorsque ces unités sortent du cadre des dénominations propres, elles peuvent néanmoins entrer dans le champ de la collecte qui doit être réalisée à partir des dénominations propres. Par ailleurs, on verra plus loin que leur étude introduit à un phénomène d'homonymie qui semble délibérément entretenu dans le cadre du parler d'entreprise que l'on étudie. La liste complète des

¹ On renvoie, entre autres, à L. Guéry (2000) et au *Lexique des règles typographiques en usage à l'Imprimerie nationale* (2002).

² M. Riegel *et al.* (2004 : 75).

dénominations recensées étant donnée en annexe, on donne ici des exemples destinés à illustrer le périmètre couvert par la collecte :

- noms de personnes attestés, et non des noms propres grammaticaux : « Jean Laurent » a été retenu, mais « Jean » et « Laurent », non attestés seuls, ont été éliminés.
- toponymes, (par exemple, « France »)
- noms d'institutions, comme celles qui correspondent plus haut à des appellations,
- noms de personnes morales, (par exemple, « Crédit agricole »),
- noms de marques, (par exemple, « Mozaïc »),
- noms de produits, (par exemple, « Livret Jeunes Mozaïc »),
- noms d'artéfacts très divers allant d'un type d'avion, A 330, au texte de loi.
- un ensemble de termes faisant l'objet d'une valorisation spéciale dont la majuscule est la marque écrite : titres divers (Directeur Général), des statuts valorisés (Primary Dealer), des notations (Fitch AA+), des noms de directions appartenant à des organigrammes et qui sont également des termes (Moyens de paiement) ; enfin, des valorisations institutionnalisées entrant ou non dans un contexte de reprise anaphorique, comme « Groupe » reprenant le syntagme « le groupe Crédit agricole » ou désignant cette banque comme « le Groupe ».

2. Relation établie dans la terminographie entre une dénomination propre et ses variantes

Les normes terminologiques utilisées pour l'établissement de terminologies (en particulier les normes NF ISO 704 : 2001 et 1087-1 : 2001) ne traitent pas à proprement parler de la question des variantes d'une même dénomination. Ces normes reconnaissent, par exemple, la présence de sigles et d'acronymes¹ pour une dénomination complète, mais ne précisent pas le type de relation qui peut être établie entre ces unités. Il en va de même pour les notions de terme « privilégié », « toléré » et « déconseillé »² qui semblent pourtant indiquer que des variations de dénominations pourraient être prises en compte. En fait, seules les dénominations

¹ NF ISO 1087-1 : 2001, 3.4.10, 3.4.11 et NF ISO 704 :2001, A.2.3.

² NF ISO 1087-1 : 2001, 3.4.15 à 3.4.17 et NF ISO 704 : 2001, 8.1.

tout à fait différentes sont mentionnées, et l'application de la cote terminologique semble essentiellement destinée à faire le tri parmi des synonymes : « *Une des fonctions essentielles d'une terminologie normalisée est d'indiquer si un terme est privilégié, toléré ou déconseillé. Un terme recommandé par un comité technique doit être considéré comme un terme privilégié, alors qu'un terme toléré doit tenir lieu de synonyme acceptable pour un terme privilégié. Les termes déconseillés sont des termes qui ont été rejetés. Un certain nombre de raisons justifient qu'un terme soit rejeté ou déconseillé. Un terme qui peut être un synonyme possible d'un terme privilégié est rejeté pour favoriser la monosémie. Un terme peut être erroné ou impropre¹.* » Par ailleurs, la relation d'équivalence est réservée dans la norme NF ISO 1087-1 :2001 à la « *relation entre désignations représentant le même concept dans des langues différentes².* ». Il reste donc la relation de synonymie, même si elle ne semble pas avoir été conçue pour être employée dans le cas des variantes, puisqu'elle établit une « *relation entre termes représentant le même concept dans une langue donnée³* », autrement dit, elle établit une relation entre des désignations différentes à un degré plus élevé que celui des simples variantes.

C'est cependant cette relation que l'on retiendra, afin de relier entre elles les variantes d'une même dénomination propre. Du reste, il s'agit d'une extension de sens qui est pratiquée dans d'autres normes dédiées à une forme de terminologie, les thésaurus. En effet, les normes ISO 2788 : 1986 et NF Z 47-101 : 1990 relatives à l'établissement de thésaurus monolingues et multilingues, proposent une acception extensive de la relation d'équivalence. Celle-ci s'applique comme dans la norme NF ISO 1087-1 : 2001 aux désignations d'un même concept dans des langues différentes⁴, mais elle s'étend également aux synonymes compris de manière extensive comme pouvant aussi correspondre, entre autres, à des désignations appartenant à des registres de langue différents, aux variantes pour des notions

¹ NF ISO 704 : 2001, 8.1.

² NF ISO 1087-1 : 2001, 3.4.21.

³ NF ISO 1087-1 : 2001, 3.19.

⁴ NF Z 47-101 : 1990, 5.1.

nouvelles et « *aux variantes orthographiques, y compris les variantes de racines et les pluriels irréguliers*¹. »

II MÉTHODE DE COLLECTE DES DÉNOMINATIONS PROPRES ET RÉSULTATS OBTENUS

A Identification des dénominations propres et de leurs variantes

1. Établissement d'une liste brute de formes avec majuscule initiale

La méthode utilisée pour la collecte des dénominations propres du corpus constitue une tentative de mise en pratique des principes et des indications qui viennent d'être exposés, dans le cadre de l'approche textométrique. Le fait de devoir tenir compte des dénominations propres telles qu'elles sont attestées, et non des seuls noms propres grammaticaux, induit un certain nombre de contraintes qu'il est nécessaire de prendre en compte. Ainsi, on veillera aux bonnes délimitations gauche et droite des dénominations collectées ; par ailleurs, l'utilisation de la majuscule ne constituant qu'un point de départ, il conviendra également de pousser les investigations au-delà de ce seul critère. Enfin, le repérage des variantes devra pouvoir être fait en contexte et permettre de désambiguïser les formes recensées. Dans les sections suivantes, la collecte sera évaluée en termes de rappel et de précision, puis les propriétés générales diachroniques et topographiques des dénominations propres du corpus seront présentées.

L'identification des dénominations propres et de leurs variantes se fait en plusieurs étapes. La première d'entre elles consiste à établir la liste des formes simples avec majuscules ou en capitales présentes dans le corpus. Le logiciel Hyperbase conçu par É. Brunet propose un traitement automatique d'identification des noms propres. On a vu plus haut que son concepteur vise implicitement les noms propres grammaticaux, formes simples, et non les dénominations propres. Cependant, comme il ne s'agit à ce stade que d'identifier une liste de formes simples, une sortie automatique permettrait un gain de temps avantageux en isolant les véritables « noms propres » des simples formes dotées d'une majuscule initiale

¹ ISO 2788 : 1986, 8.2.

du fait de leur position dans la phrase. Le programme comporte à cet effet des algorithmes qui visent à séparer le bon grain de l'ivraie. Par exemple, si un mot apparaît en début de phrase avec une majuscule et qu'en dehors de cette position il est toujours écrit avec une minuscule initiale, il sera éliminé automatiquement de la liste des noms propres.

Cependant, on a pu constater que ce traitement produit deux sortes d'artefacts qui perturbent la composition de la liste¹. Le premier a pour conséquence d'éliminer des formes qui existent à la fois comme occurrences d'un mot simple – et qui par conséquent sont susceptibles d'entrer dans le cas de figure évoqué plus haut –, et comme parties d'une dénomination propre. Par exemple, « Banque » est éliminé de la liste des noms propres, car il correspond à une forme présente en minuscules dans les phrases (hors position initiale) des textes du corpus ; cependant, la forme avec majuscule correspond aussi au premier membre d'une trentaine de dénominations propres complexes construites sur un patron qu'illustre l'exemple « Banque Commerciale de Grèce ». Le deuxième artefact transforme une partie des mots en capitales en des mots avec majuscule initiale absents du corpus. Cela provient du fait que les lettres accentuées ont été transposées sans accent lorsque certains mots ont été écrits en capitales dans les textes du corpus – ce phénomène est du reste assez fréquent². Lorsque ces mots ont été transformés par le programme en formes simples avec majuscule initiale, ils n'ont pas retrouvé leurs accents pour autant. Cela a eu pour effet de produire des mots qui n'existent pas à proprement parler dans les textes, comme « TRESORERIE » qui est devenu « Tresorerie ».

Afin d'éviter les biais introduits par des traitements automatiques du type de celui qui vient d'être présenté, on a opté pour une source très proche de l'étape primitive de segmentation du texte en formes simples. On est donc parti du tableau lexical entier produit par Lexico ; celui-ci donne pour chaque forme la ventilation de

¹ Pour ce test, on a comparé à l'aide d'Excel la liste des noms propres du corpus extraite automatiquement par Hyperbase avec la liste des parties de dénominations propres dotées d'une majuscule initiale ou en capitales. Plus de 500 formes sur 2800 possédant une majuscule initiale et entrant dans la composition de dénominations propres sont absentes de la liste des noms propres produite par Hyperbase. Le résultat complet de la comparaison est donné en annexe. Ce test permettra peut-être, on le souhaite du moins, d'améliorer les traitements dédiés aux noms propres dans Hyperbase.

² N. Catach (1996 : 83).

ses occurrences dans les différentes parties du corpus. Toutes les formes présentes dans ce tableau avec une majuscule initiale ont donc été sélectionnées. On obtient ainsi une liste brute de 3945¹ formes dont on ne peut pas dire à ce stade s'il s'agit de noms propres ou, *a fortiori*, de dénominations propres.

*Tableau 48.
Vingt premières formes du dictionnaire du corpus avec majuscule initiale*

Forme	Fréquence
Crédit	2992
Agricole	1713
Caisses	1047
Indosuez	827
Directeur	724
Président	724
Total	688
Le	666
MF	663
Groupe	603
France	542
Les	519
A	500
En	489
Général	485
La	467
Tél	449
S	437
Banque	435
L	412

La lecture de ce tableau permet rapidement d'identifier les principales catégories de formes que l'on obtient ainsi. On distingue d'abord les formes *a priori* indécidables et pour lesquelles il sera nécessaire de vérifier si elles constituent une dénomination propre à elles seules, si elles entrent dans la composition d'une dénomination complexe, ou bien si elles tiennent les deux positions à la fois : « Crédit », « Agricole », « Caisses », « Indosuez », « Directeur », « Président », « Total », « MF », « Groupe », « France », « Général » et « Banque ». Le deuxième

¹ Ces 3945 formes représentent 35% des formes présentes dans le corpus qui en compte 11 231. La liste se décompose en plusieurs paliers de fréquence. On dénombre ainsi 21 unités dont la fréquence va de 2992 à 399 ; un deuxième palier comporte 539 unités dont les fréquences s'échelonnent de 265 à 10 ; un troisième palier regroupe 1818 unités dont les fréquences s'échelonnent de 9 à 2 ; enfin, un quatrième palier regroupe 1567 hapax.

groupe se compose d'articles et d'une préposition, tous susceptibles d'occuper la tête d'une phrase : « Le », « Les », « En », « La », « L ». Un troisième et dernier groupe rassemble des éléments plus énigmatiques : « A », « Tél », « S ». Les deux derniers groupes ne posent pas de grandes difficultés car leurs formes sont relativement peu nombreuses et une vérification sommaire permet d'identifier le fonctionnement des grammaticaux et de lever certaines énigmes. Les difficultés se concentrent sur le groupe le plus important pour les éléments duquel des vérifications approfondies du fonctionnement en discours devront être appliquées.

La première difficulté tient au caractère indécidable a priori du statut de ces formes par rapport aux dénominations propres. Elle oblige à procéder à des contrôles systématiques, car on a vu avec l'exemple donné plus haut que certains raccourcis aboutissent à d'inévitables oublis. La seconde difficulté tient à la délimitation des dénominations qui sont susceptibles de comporter de façon banale quatre ou cinq parties, nombre qui peut parfois s'élever jusqu'à la dizaine comme pour le « Forum de l'Agriculture Raisonnée Respectueuse de l'Environnement ». Cette difficulté est en quelque sorte redoublée par un phénomène d'inclusion des dénominations les unes dans les autres, comme dans l'exemple ci-dessous :

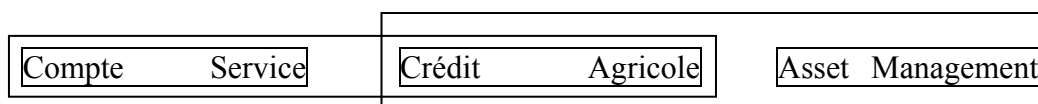


Figure 54.

Exemples de dénominations propres emboîtées les unes dans les autres

Il est donc nécessaire de pouvoir disposer de moyens textométriques permettant d'opérer un repérage fiable sur un nombre de formes et des fréquences relativement importants.

2. Segments répétés, concordances et identification des dénominations propres

Les segments répétés ont déjà été utilisés dans la deuxième partie, lorsqu'il s'est agi de déterminer l'usage de certains mots-clés dans le corpus. Le fait qu'un mot-clé soit inclus dans un ou plusieurs termes complexes ou dénominations propres à hauteur de plus des deux tiers de ces occurrences a été considéré comme un critère pour caractériser un usage majoritaire sortant de celui de la langue commune. Ce

faisant, on avait abordé indirectement les relations pouvant être établies entre dénominations complexes et segments répétés.

Les segments répétés, dont on rappelle qu'ils correspondent à des suites de formes dont les fréquences sont égales ou supérieures à 2 dans le corpus¹, n'ont pas eu pour destination première de permettre l'identification de dénominations propres, celles-ci ne constituant qu'un sous-ensemble marginal des segments d'un texte². Le calcul des segments a été conçu au sein du laboratoire de l'ENS Saint-Cloud lorsqu'il a semblé nécessaire d'identifier et de quantifier par des moyens textométriques la présence récurrente dans les textes d'expressions plus ou moins longues : *« Peut-on objectiver par des comptages cette sensation que le discours politique est un genre où les répétitions sont très abondantes, que certaines séquences reviennent souvent, martelées au détriment du sens de chacune des formes qui les composent, allant même jusqu'à produire cette impression, souvent décrite de manière polémique, que le discours est produit par une « langue de bois » ? On perçoit bien, à la simple lecture de ces textes, des répétitions qui mettent en jeu des unités plus longues que les simples formes du vocabulaire. Mais en l'absence de comptages précis on en est réduit à se demander si ces répétitions formelles sont bien propres au genre étudié³. »*

L'utilisation des segments répétés pour l'identification des dénominations propres présentes dans le corpus offre néanmoins plusieurs avantages, à condition de les utiliser pour des unités relativement fréquentes, la fréquence 2 étant le minimum requis, les fréquences égales ou supérieures à 10 constituant l'ensemble qu'il est intéressant de traiter de cette manière. Cette restriction faite, le premier avantage de cette approche est de donner accès à des suites de formes dans un contexte textométrique, ce qui permettra ensuite de bénéficier de tous les calculs qui peuvent être appliqués aux formes simples : dictionnaire, recensement des occurrences, ventilation de ces dernières dans les parties du corpus, calcul des spécificités. Le second avantage, qui découle du précédent, est de permettre l'ordonnement des

¹ L. Lebart et A. Salem (1994).

² En anticipant sur les indications qui sont fournies plus bas, 1151 segments ont été identifiés comme correspondant à des dénominations complexes, sur un effectif total de 42 771 segments, soit un peu moins de 3%.

³ A. Salem (1987 : 22).

segments par fréquences, des plus élevées aux plus basses. Ce classement permet de filtrer avantageusement un volume conséquent de scories, les dénominations complètes ayant dans tous les cas une fréquence plus élevée que les syntagmes dans lesquels elles peuvent être incluses (il s'agit, par exemple, des cas où la dénomination propre est précédée par un mot grammatical, ou lorsqu'elle est suivie d'une particule verbale). Néanmoins, et c'est le troisième avantage, les formes captées par le segment parce qu'elles appartiennent à la phrase, mais qui n'appartiennent pas à la dénomination, permettent fréquemment d'aider à délimiter la dénomination sans ambiguïté. On trouve, par exemple, parmi la dizaine de segments les plus fréquents incluant la forme « **UI** » (Frq. 82), les informations suivantes :

*Tableau 49.
Segments incluant une dénomination propres et ses délimiteurs*

Type de délimitation	segment attesté	Occurrences
délimitation gauche	<i>le groupe UI</i>	5
délimitation droite	<i>UI a</i>	16
sigle et son développé	<i><u>Union d Etudes et d Investissements</u> UI</i>	11

Le calcul des segments prend appui sur la segmentation des textes du corpus en formes simples. Cette segmentation repose sur la distinction entre deux sortes de caractères entrant dans la composition des textes : les caractères délimiteurs de formes et les caractères non délimiteurs. Les premiers correspondent en général au blanc, aux signes de ponctuation usuels et aux caractères utilisés pour formater le texte, comme par exemple « < » et « > » qui encadrent la balise servant à déclarer une partie du corpus pour le logiciel Lexico. Les caractères non délimiteurs correspondent au reste, c'est-à-dire aux lettres de l'alphabet et aux chiffres. En général, pour déterminer les suites de formes récurrentes du corpus, le calcul des segments répétés tient compte des délimiteurs de séquence. Ceux-ci sont un sous-ensemble des délimiteurs de formes, correspondant aux ponctuations faibles et fortes (les différentes sortes de points et de virgules, les guillemets, les tirets et les parenthèses). Lexico utilise ces délimiteurs pour calculer les segments répétés, mais il offre aussi la possibilité de ne pas les prendre en compte. On vient de voir plus

haut l'avantage que l'on peut tirer de cette option. En effet, on peut ainsi capter le plus grand nombre possible d'indices de délimitation des dénominations, comme par exemple le sigle « UI » qui est séparé dans le texte de la dénomination complète par une paire de parenthèses. La liste des segments répétés du corpus que l'on a utilisée pour l'identification des dénominations propres comporte les segments calculés sans délimiteurs de séquence, complétés par les segments calculés avec les délimiteurs mais absents du premier ensemble¹. On obtient ainsi une liste de 42 771 segments répétés.

L'identification des dénominations repose ensuite sur la projection des formes appartenant aux deux premiers paliers de la liste des 3945 formes simples possédant une majuscule initiale ou étant écrites en capitales, sur la liste des 42 771 segments répétés. Pour réaliser cette opération, on a utilisé le tableur Excel dans lequel les deux listes de formes et de segments avec leurs fréquences ont été copiées. Chaque forme simple est ensuite recherchée dans la liste des segments triés par ordre décroissant de fréquence. Cette recherche se fait à l'aide de la fonction « contient » et permet la sélection de tous les segments contenant la forme, sans que la casse soit prise en compte. Ce dernier point permet de contrôler d'éventuelles dénominations utilisées sans majuscule initiale, comme les adresses Internet. L'utilisation du tableur Excel permet en outre de sélectionner les formes ou les segments correspondant à des dénominations propres au fur et à mesure que les requêtes sont jouées. On conserve ainsi une trace qui permettra ensuite de constituer facilement le sous-ensemble de formes et de segments correspondant effectivement à des dénominations propres. On donne ci-dessous un les premiers segments répétés incluant la forme « Indosuez » ; les segments correspondant à une dénomination comportent le chiffre 1 dans la colonne dédiée à cette marque.

¹ Cette différence s'explique par le mode de calcul des segments. En effet, le programme élimine les segments répétés inclus dans des segments plus larges et ayant la même fréquence. Certains segments calculés avec les séparateurs n'apparaissent pas dans la liste des segments calculés sans séparateurs quand il existe un segment répété enjambant un délimiteur et qui possède la même fréquence que le segment inclus calculé avec les délimiteurs de séquence. Les deux listes de segments ont été comparées à l'aide d'Excel, ce qui a permis de repérer facilement les segments manquants.

*Tableau 50.
Dix premiers segments incluant la forme « Indosuez » (827 occurrences)*

Liste des segments calculés avec et sans délimiteurs	Fréquence	Dénomination ?
Agricole Indosuez	594	
Crédit Agricole Indosuez	593	1
de Crédit Agricole Indosuez	100	
Indosuez Cheuvreux	86	
Crédit Agricole Indosuez Cheuvreux	82	1
Indosuez W I Carr	67	
Indosuez W ¹	67	
Banque Indosuez	64	1
Indosuez W I Carr Securities	62	1
Indosuez a	58	

Les fréquences associées aux segments facilitent un repérage rapide des dénominations propres. En effet, les segments correspondant aux dénominations sont fréquemment encadrés par d'autres segments correspondant, soit à une partie de la dénomination, et dans ce cas ils ont une fréquence plus élevée, soit à un morceau de phrase dans lequel la dénomination est incluse, et dans ce cas ils ont une fréquence plus faible. On dispose alors d'un premier indice ; il est ensuite aisé de vérifier à l'aide des concordances si les différences de fréquences correspondent à une variante ou à un quelconque incident typographique. Si l'on reprend l'exemple qui a été retenu ici, les concordances des 594 occurrences de « Agricole Indosuez » triées sur leurs contextes gauches permettent immédiatement d'identifier la seule occurrence dans laquelle « Crédit » est écrit sans accent.

¹ Ce segment qui possède la même fréquence que celui qui précède et dans lequel il se trouve inclus n'a pas été éliminé par le logiciel car il provient d'un deuxième calcul des segments dans lequel les délimiteurs ont été maintenus (voir plus haut). Les lettres W et I étant suivies de points dans la dénomination propre, le segment calculé avec les délimiteurs de séquence s'est logiquement arrêté avant le point.

t hors - bilan) , Credit Agricole Indosuez Fixed Income est
ces Fastnet (30 % Crédit **Agricole** Indosuez) est spécialisée
le S . A . et 35 % Crédit **Agricole** Indosuez) est spécialisée
Management et 45 % Crédit **Agricole** Indosuez) qui coiffe déso
le S . A . et 48 % Crédit **Agricole** Indosuez) est dédiée aux

Figure 55.

Extrait des concordances du segment répété « Agricole Indosuez »

Pour les formes appartenant aux autres paliers, comme dans le contrôle réalisé dans l'exemple précédent, les classiques concordances se sont révélées efficaces, d'autant que le logiciel Lexico permet de traiter de cette manière des séries de formes. Par exemple, pour des formes de fréquence 4, il est possible de travailler par lots de cinq formes ou plus ; pour les formes de fréquence 1, on travaille par lots de trente formes, ce qui permet un repérage très rapide des dénominations. On donne ci-dessous l'exemple d'un lot de cinq formes de fréquence 4 (Autonomie, Automates, Auvray, AUBRY, Atlantico) projetées sur les concordances.

```

non agricoles . # 1967 Autonomie financière de la
e universelle . # 1967 Autonomie financière de la
e universelle . # 1967 Autonomie financière de la
e universelle . # 1967 Autonomie financière de la
8 910 000 9 442 440 . Automates bancaires 5 968 6
943 910 10 495 724 . # Automates bancaires 6 737 7
9 442 440 9 943 910 . Automates bancaires 6 737 7
8 223 000 8 910000 . # Automates bancaires 3 560 3
cteur Général : Gilles Auvray 43 Caisses locales 1
cteur Général : Gilles Auvray 43 Caisses locales 9
cteur Général : Gilles Auvray . # Nord . 10 , aven
cteur général : Gilles Auvray . # FINISTÈRE 7 , ro
. 35 ; Président : Guy AUBRY ; Directeur Général :
5 35 . Président : Guy AUBRY . Directeur général :
35 35 Président : Guy AUBRY Directeur général : J
35 35 Président : Guy AUBRY Directeur Général : J
e Janeiro Banco Inter - Atlantico SA A : Rio Branc
5 % dans Banco Inter - Atlantico qui détient 85 %
pital du Banco Inter - Atlantico ; spécialisé dans
Banco Boavista Inter - Atlantico . # Chili : Banco

```

Figure 56.

Concordances simultanées des cinq formes « Autonomie », « Automates »,
« Auvray », « AUBRY » et « Atlantico »

A partir des concordances des cinq formes de fréquence 4 il est possible d'identifier de façon sûre les cinq dénominations propres suivantes : « Gilles Auvray », « Guy AUBRY », « Banco Inter - Atlantico SA », « Banco Inter – Atlantico », « Banco Boavista Inter – Atlantico ». Deux de ces formes étant une variante de la même dénomination (« Banco Inter - Atlantico SA » et « Banco Inter – Atlantico »). Les deux premières formes correspondent sans ambiguïté, pour la première (« Autonomie »), à un début de phrase, pour la seconde (« Automates »), à un intitulé de ligne dans un tableau contenant des données chiffrées.

Ces deux procédures ont permis d'établir une première liste de 1151 dénominations propres complexes et 544 dénominations propres simples de fréquence supérieure ou égale à 2, ainsi que 1137 dénominations propres simples ou complexes de fréquence 1, soit un total de 2832 dénominations propres.

3. *Évaluation de la collecte*

Comme cela a été indiqué plus haut, la procédure d'identification et de collecte des dénominations propres doit pouvoir être contrôlée en sorte de pouvoir se faire une idée précise de la qualité du résultat obtenu. Pour y parvenir, on a procédé à des sondages à partir de la carte des sections de Lexico. Quatre lots de vingt-cinq paragraphes dont les coordonnées ont été tirées au hasard ont fait l'objet d'un dépouillement systématique afin de relever les dénominations propres. Les résultats obtenus ont permis de mesurer le taux de rappel, le taux de précision et de calculer la mesure P&R¹. Le nombre de dénominations propres collectées rapporté au nombre de celles qui ont été relevées dans le cadre des tests donne un rapport de 0,97 (rappel) ; le nombre de dénominations propres collectées correspondant bien à des dénominations propres identifiées lors du test donne un taux de précision de 0,99. Dans le premier cas, le test révèle que quelques dénominations propres n'ont pas été identifiées lors de la phase de collecte ; dans le second, le test a montré un cas d'homonymie qui aurait dû être distinguée lors de la phase de collecte et qui ne l'a pas été. La mesure P&R donne un résultat de 0,98 qui est tout à fait comparable à celui que des annotateurs humains ont obtenu dans un autre cadre pour les

¹ Pour les formules, voir *supra*.

conférences MUC évoquées précédemment¹. On voit de cette manière que la collecte pourrait être encore affinée, mais que les résultats obtenus dès la première phase d'identification sont presque exhaustifs. Par conséquent, ce sont les dénominations propres obtenues lors de la phase d'identification qui seront situées par rapport aux autres composantes du vocabulaire des rapports d'activité.

4. Le repérage des variantes et des sigles

Dans un premier temps, les variantes et les sigles sont identifiés et collectés comme les autres dénominations propres. Mais la méthode adoptée est redondante pour les dénominations complexes qui sont susceptibles d'être retrouvées autant de fois qu'elles possèdent de composants. Pour illustrer ce point, on peut partir de l'exemple fourni plus haut par la dénomination « Banco Inter-Atlantico ». Cette dénomination fait partie d'un ensemble de variantes qui ont pu être rencontrées à cinq reprises au moins et avec les deux méthodes présentées plus haut. On a rassemblé dans le tableau ci-dessous les différents cas de figure rencontrés.

¹ T. Poibeau (2003 : 101).

*Tableau 51.
Identification des variantes d'une dénomination propre*

Forme simple	Frq.	Segments répétés	Concordances
Banco	167	Banco Boavista InterAtlantico (3 occ.) Banco Boavista InterAtlântico (3 occ.) Banco Inter-Atlantico (3 occ.) le Banco Boavista (3 occ.) Banco InterAtlantico (2 occ.) le Banco Boavista InterAtlantico (2 occ.)	
Boavista	12	Banco Boavista (10 occ.) Banco Boavista InterAtlantico (3 occ.) Banco Boavista InterAtlântico (3 occ.)	Banque Boavista (1 occ.) banque Boavista (1 occ.)
Inter	9		Banco Inter-Atlantico (3 occ.) Banco Boavista Inter-Atlantico (1 occ.)
Atlantico	4		idem (voir concordance donnée plus haut)
InterAtlântico	3		Banco Boavista InterAtlântico (3 occ.)

Les recoupements qui peuvent être faits au moment de la collecte ne suffisent pas toujours comme dans cet exemple pour démêler l'écheveau des dénominations et des transformations qu'elles peuvent subir du fait des variations graphiques, mais aussi en raison des modifications qui affectent la réalité qu'elles dénomment. En effet, ce n'est pas une seule dénomination qui a été ici recensée, mais trois. Pour établir ce point, un retour au contexte est indispensable. Il permet de comprendre que, dans un premier temps, le Banco InterAtlantico a fait l'acquisition d'une autre banque brésilienne, le Banco Boavista. Cette acquisition a eu, entre autres conséquences, pour effet de modifier la dénomination de la banque acheteuse qui est alors devenue le Banco Boavista InterAtlantico. On a donc trois ensembles dénommatifs avec leurs variantes respectives :

- Banco InterAtlantico = Banco Inter-Atlantico
- Banco Boavista = Banque Boavista = banque Boavista
- Banco Boavista InterAtlantico = Banco Boavista InterAtlântico = Banco Boavista Inter-Atlantico

On peut remarquer à l'occasion de cet exemple qu'un contrôle par les concordances apporte parfois des compléments dans les basses fréquences. Il est également possible de procéder à un contrôle de deuxième niveau en repartant de la liste des dénominations propres pour la transformer en unités simples. On contrôle ensuite la manière dont les unités simples de fréquence supérieure à 1 (hormis les grammaticaux comme le déterminant « de ») sont présentes dans la liste des dénominations complètes. Il faut noter, néanmoins, que quelques variantes ne sont rattachées à une dénomination principale qu'au cours de manipulations ultérieures de la liste complète. Ces repérages ont permis d'identifier 479 variantes dans la liste complète des dénominations, soit 17%.

Enfin, le repérage des sigles et des acronymes, que l'on a comptabilisés dans les variantes, a été réalisé selon les mêmes procédures, à une différence près. En effet, on a vu dans la deuxième partie que les dénominations propres étaient à peu près les seules unités à faire l'objet d'une prise en charge énonciative particulière se traduisant, entre autres, par les mentions consécutives d'une dénomination complète et de son sigle. On en a donné un exemple plus haut avec le sigle UI, ce cas de figure se répétant un nombre de fois relativement important : sur les 671 paires de parenthèses que compte le corpus¹, 18% sont utilisées pour donner un sigle ou un développé de dénominations propres dans pratiquement tous les cas. Sauf dans les cas de faibles fréquences, ces développés ont été identifiés à l'aide des segments répétés.

B. Les dénominations propres dans le vocabulaire du corpus

1. Aspects statistiques relatifs aux dénominations propres

La présentation d'un certain nombre d'aspects statistiques globaux relatifs aux dénominations propres vont permettre de préciser la part qu'elles prennent dans des phénomènes qui ont été observés, principalement dans la première partie, à propos du vocabulaire du corpus pris dans son ensemble. Afin de pouvoir mener ces observations, il a été nécessaire de réduire la liste des dénominations propres à la

¹ Cette liste est donnée en annexe.

liste des formes simples dotées d'une majuscule initiale ou écrites en capitales. Les formes sans majuscules qui faisaient partie de ces dénominations ont été retirées de la liste parce qu'elles fonctionnaient également dans le corpus en dehors des dénominations propres. On a cependant fait une exception pour les classificateurs entrant dans les odonymes (rue, avenue, etc.). La graphie de ces unités varie d'une partie du corpus à l'autre, tantôt avec majuscule initiale, tantôt sans, mais comme leur utilisation est toujours la même, à savoir former une dénomination d'odonyme, ils ont été conservés dans la liste¹. Cette liste peut comporter un certain nombre d'occurrences de formes avec majuscule initiale, entrant dans la composition des dénominations propres, mais pouvant également être utilisées en début de phrase en dehors de telles dénominations. Le contrôle de ce point, qui a porté sur les fréquences les plus élevées, a montré que le risque de distorsion était faible, sans commune mesure avec celui qu'aurait provoqué la conservation des parties de dénominations ne comportant pas de majuscule initiale.

Cette liste a servi dans un premier temps à construire un sous-corpus dans lequel les découpages en parties ont été conservés, mais où seules les formes appartenant à la liste qui vient d'être décrite ont été conservées. Le but de cette opération est de cerner la part du groupe constitué par les dénominations propres dans les évolutions du vocabulaire du corpus. Deux mesures principalement ont été retenues : l'accroissement du vocabulaire et l'analyse factorielle des correspondances. Avant de présenter le résultat de ces deux mesures, on donne ci-dessous les principales caractéristiques textométriques du sous-corpus.

¹ Cette liste est donnée en annexe.

Tableau 52.

Principales caractéristiques textométriques du corpus des rapports d'activité réduit aux formes graphiques entrant dans la composition de ses dénominations propres

Num	Partie	Occurences	Formes	Hapax	Fmax	Forme
1	1995	2254	896	639	198	Crédit
2	1996	2306	827	592	207	Crédit
3	1997	3159	978	665	319	Crédit
4	1998	2993	902	593	328	Crédit
5	1999	3548	995	619	367	Crédit
6	2000	4186	1072	668	427	Crédit
7	2001	4672	1095	632	426	Crédit
8	2002	3162	800	462	259	Crédit
9	2003	4717	1301	829	461	Crédit

On note que, contrairement à la plupart des corpus textuels en français, la forme « de » n'est pas la plus fréquente. Bien qu'elle entre dans la composition des dénominations propres, elle a été retirée de la liste présentée plus haut, car les emplois de ce déterminant dans le corpus dépassent très largement leur seule présence dans les dénominations propres. Une fois constitué, le sous-corpus a été traité de manière classique par le programme Lexico. Le premier calcul présenté ici est celui de l'accroissement du vocabulaire.

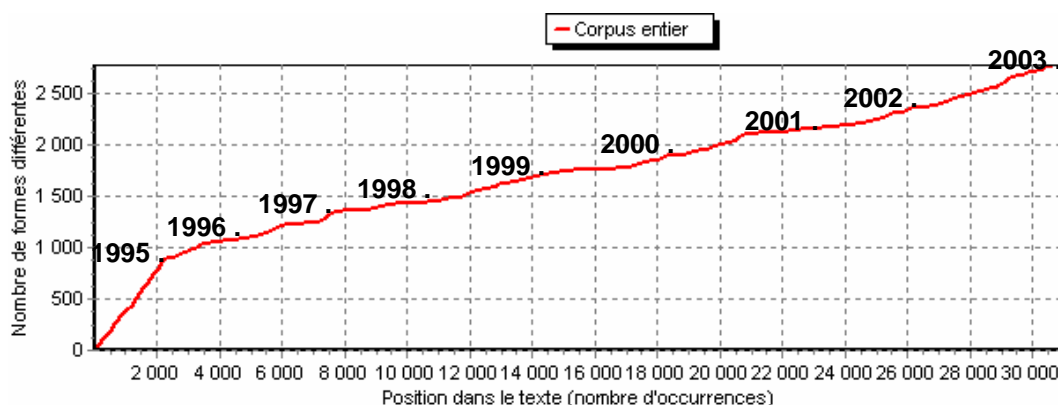


Figure 57.

Accroissement du vocabulaire des rapports d'activité réduit à ses formes simples avec majuscules ou en capitales incluses dans des dénominations propres

Lecture de la figure 57 : les parties du corpus correspondant aux années de rapports d'activité sont réparties sur une courbe ayant pour ordonnées la progression du nombre de formes et pour abscisses la progression du nombre d'occurrences au fur et à mesure que l'on parcourt les textes du corpus, et cela jusqu'au dernier mot. Une année correspond à la portion de la courbe comprise

entre deux repères. Les formes et leurs occurrences pour l'année 1995 correspondent au tronçon allant du point zéro au repère « 1995 », les formes et leurs occurrences pour l'année 1996 correspondent au tronçon allant du repère « 1995 » au repère « 1996 », et ainsi de suite.

On remarque tout d'abord que la courbe de ce graphique est globalement alignée sur celle du vocabulaire pris dans son ensemble. On constate la même progression du nombre d'unités nouvelles d'année en année, même si cet accroissement ne se fait pas toujours dans les mêmes proportions. La position des années correspondant aux parties du corpus sur la courbe signale deux écarts plus importants que les autres : celui qui sépare 2000 de 2001 et l'autre séparant 2002 de 2003. Comme cela a été indiqué au-dessus, cette « séparation » est d'une nature particulière, car elle indique que l'année 2001 est particulièrement riche en occurrences et que l'année 2003 cumule cette propriété avec, en outre, un accroissement important de formes nouvelles. On peut donc voir que les dénominations propres entrent pour une part notable dans l'accroissement du vocabulaire du corpus. Pour l'année 2003, sur les 1500 formes nouvelles qui apparaissent, un peu moins de 400 appartiennent à des dénominations propres, soit un quart environ.

Lorsqu'elle est appliquée à ce corpus, l'analyse factorielle des correspondances permet de faire ressortir les caractéristiques propres à l'année 2003 par rapport à ce qui a pu être constaté au sujet du vocabulaire pris dans son ensemble. En effet, dans la première partie, on a pu voir que l'AFC partage le corpus en deux ensembles : les années 1995-2000 et les années 2001-2003. Cette coupure a été interprétée comme une recombinaison du vocabulaire des années précédentes, entraînant des modifications importantes de fréquences pour certaines formes. Ce phénomène a été compris comme étant principalement une répercussion du changement de statut de l'émetteur des rapports d'activité. Cela avait été l'occasion de relever aussi que les dénominations propres étaient un bon indicateur de ce type de bouleversements. Ce rôle est confirmé par la lecture que l'on peut faire des résultats de l'AFC appliquée au sous-corpus des rapports d'activité.

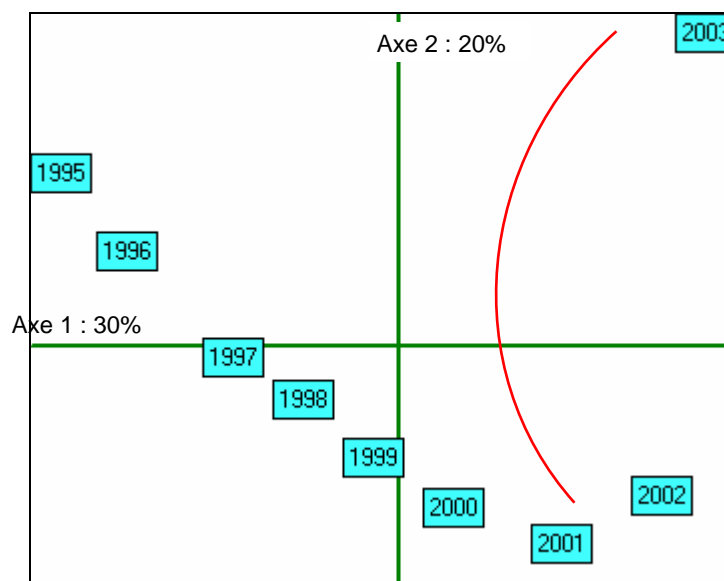


Figure 58.

AFC sur le corpus RA9503 réduit aux seules formes simples ayant une majuscule initiale ou en capitales, incluses dans des dénominations propres

De façon peu surprenante, compte tenu de ce qui a été vu dans le chapitre 3, l'analyse factorielle des correspondances propose une disposition typique des parties du corpus lorsque celui-ci est organisé en série textuelle chronologique. Or, contrairement au résultat obtenu dans la première partie lorsque ce calcul a été appliqué au vocabulaire pris dans son ensemble, la réduction du corpus aux dénominations propres oppose un *continuum* allant de 1995 à 2002 et la seule partie 2003. On a ici une confirmation de ce qui a été avancé plus haut au sujet de cette année : elle cumule un nombre important d'occurrences avec l'apparition de formes nouvelles. On a vu dans le chapitre 4 que l'année 2003 était caractérisée par la présence de nombreuses dénominations propres nouvelles parmi ses formes spécifiques positives : Crédit Lyonnais, Crédit Lyonnais Asset Management, Calyon, etc. Il n'est donc pas surprenant que ces dénominations ou parties de dénominations propres soient responsables de l'isolement de l'année 2003 par rapport aux autres parties du corpus :

Tableau 53.

Formes dotées du plus fort coefficient de spécificité positif pour la partie 2003 dans le corpus RA9503 réduit aux formes correspondant à des dénominations ou parties de dénominations propres

Forme	Frq. Tot.	Fréquence	Coeff.
Lyonnais	162	120	***
Calyon	54	54	45
CLAM	29	29	25
AM	53	38	20
Finaref	30	21	12
CA	129	47	10
Comité	31	20	10
France	542	137	10
Eurofactor	15	13	10
UAF	11	11	10
CL	10	10	9
Lixxbail	9	9	8

En effet, la majorité de ces dénominations propres ou parties de dénominations propres apparaissent en 2003 parce qu'elles sont liées au rachat du Crédit Lyonnais (Lyonnais, Calyon, CLAM, Eurofactor, UAF, CL, Lixxbail). Encore une fois, les dénominations propres se montrent pertinentes pour capter les bouleversements qui sont susceptibles d'affecter en profondeur le référentiel du groupe bancaire que l'on a pris comme exemple.

2. Les dénominations propres et le vocabulaire du corpus

Ces deux calculs ont permis de préciser, en l'isolant, le rôle tenu pas les dénominations propres dans les phénomènes qui affectent les caractéristiques du vocabulaire du corpus. Aussi, on peut chercher à savoir si ce groupe actif du vocabulaire est toujours présent dans les mêmes proportions dans toutes les parties du corpus, pris cette fois dans son ensemble. Pour avoir une idée de la manière dont le groupe des dénominations se répartit sur les différentes parties du corpus des rapports d'activité, on a projeté la liste qui les regroupe sur les parties du corpus RA9503. Cette projection est rendue possible par la fonctionnalité de gestion de groupes de formes proposée par Lexico. En effet, la liste complète de parties de dénominations propres ayant une majuscule initiale a pu être chargée dans un groupe

de formes, et l'ensemble du groupe a ensuite été projeté sur un graphique de ventilation des formes.

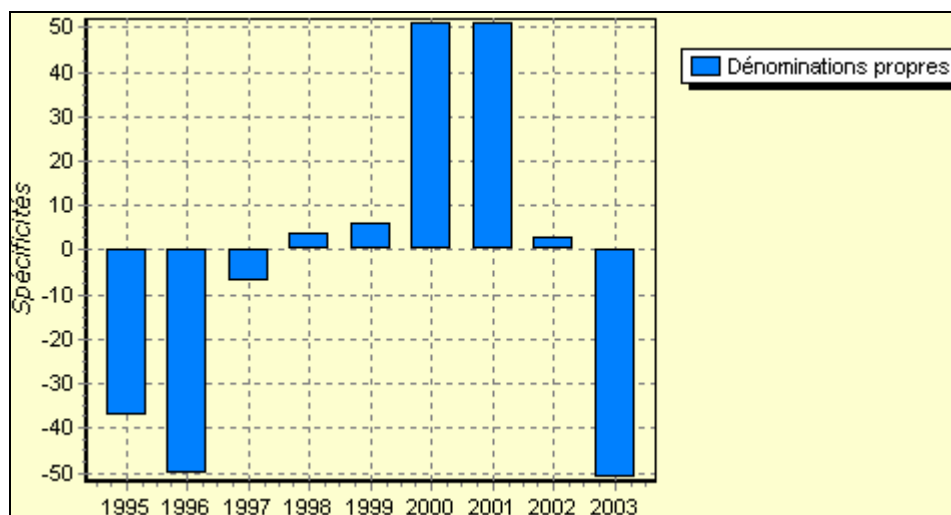


Figure 59.
Spécificités caractérisant la répartition des occurrences de la liste des dénominations propres sur l'ensemble du corpus

Le calcul des spécificités met en lumière la manière dont le groupe des dénominations propre est présent dans chaque partie relativement au reste du vocabulaire. Par rapport à la norme observée pour le corpus, le calcul indique si la présence des dénominations propres est banale, et dans ce cas les spécificités sont nulles. En revanche, si les éléments dont se compose ce groupe sont sous ou surreprésentés, les spécificités seront négatives ou positives. Le calcul oppose dans le graphique trois ensembles : les années 1995-1997 sont caractérisées par une sous-représentation des dénominations propres, la dernière année indiquant une inversion de tendance ; les années 1998-2002 sont, en revanche, caractérisées par une surreprésentation culminant en 2000 et 2001 pour diminuer en 2002. Enfin, l'année 2003, comme 1996, est caractérisée par une sous-représentation nette de la catégorie.

On peut d'ores et déjà tirer deux remarques des résultats présentés dans ce graphique. Tout d'abord, la ventilation du groupe de dénominations propres sur des textes homogènes organisés en série chronologique montre des variations de répartition importantes d'un texte à l'autre. L'étude des dénominations propres dans

un seul texte ou dans un corpus regroupant des textes hétérogènes risque, par conséquent, d'enregistrer des phénomènes ne pouvant pas être expliqués à partir de l'observation de ces seules unités. On a déjà eu l'occasion de relever ce point dans la deuxième partie, notamment pour la terminologie bancaire véhiculée par les rapports d'activité. La deuxième remarque tient au fait que la présence des dénominations propres semble devoir être expliquée par plusieurs raisons appartenant à des plans différents.

Ce point est sensible dès que l'on essaie d'expliquer les principales variations révélées par le graphique. Il en est ainsi des deux années pour lesquelles le déficit en dénominations propres semble le plus accentué : 1996 et 2003. De façon macroscopique, on peut considérer que ces déficits interviennent à deux reprises, précisément lorsque le groupe bancaire connaît des événements majeurs : achat d'une banque d'investissement en 1996, et achat de l'une des principales banques françaises en 2003. Cependant, si l'on regarde la courbe des accroissements des dénominations propres que l'on a donnée plus haut, on s'aperçoit que les deux années fournissent chacune de nouvelles formes et de nouvelles occurrences, mais dans des proportions deux fois moindres pour 1996. Les deux observations peuvent apparaître à ce stade comme étant contradictoires. Il est donc nécessaire de pousser plus loin les investigations. On s'aperçoit alors que le résultat équivalent présenté par les spécificités repose en fait sur deux mouvements inverses. Pour l'année 1996, l'achat de la banque Indosuez modifiant la dimension géographique de l'activité du groupe bancaire, un nombre important de coordonnées géographiques disparaissent¹, en attendant que le référentiel soit mis à jour, ce qui sera fait l'année suivante. Par ailleurs, les nouvelles dénominations apportées par l'achat de la banque Indosuez et de ses filiales ne sont pas encore définitivement fixées. Elles font donc une apparition timide dans le rapport : de nouvelles formes apparaissent mais elles sont peu représentées. Ces nouvelles formes compensent donc mal les pertes d'occurrences entraînées par le délai de latence observé pour la mise à jour du référentiel géographique. Un déficit est donc enregistré pour 1996, plus fort que pour l'année 1995. Pour l'année 2003, on observe un phénomène inverse: non

¹ Voir chapitre suivant.

seulement de nouvelles formes apparaissent, mais elles drainent avec elles un nombre d'occurrences élevées. Dans ce cas, le déficit indiqué par le calcul des spécificités doit s'expliquer par la dilution des occurrences des dénominations propres dans le reste du vocabulaire de cette partie, dont le nombre d'occurrences est presque le double de celui enregistré pour les autres années (on compte en moyenne 20 000 occurrences pour les autres parties et 39 000 pour 2003). On note encore à cette occasion que la proportion des dénominations propres ne varie pas mécaniquement en fonction de la taille du vocabulaire d'un corpus ou de l'une de ses parties.

Ces mouvements, très visibles sur le graphique, ne doivent néanmoins pas masquer un autre mouvement que les variations de fréquence des dénominations propres enregistrent de façon indirecte. Ce mouvement général a pour toile de fond les transformations du groupe bancaire. Comme on l'a vu précédemment haut que le groupe procède par acquisitions pour se développer dans des secteurs nouveaux de l'activité bancaire, si bien que les équilibres se recomposent en permanence à l'intérieur du groupe bancaire. Ainsi, tel secteur au départ peu développé, occupe une place importante une dizaine d'années plus tard. Le poids économique de ce secteur s'étant accru dans le groupe, une place plus large lui est donc accordée dans la restitution des activités. Comme ce sont plusieurs secteurs qui connaissent un essor comparable durant la même période, et que ces nouvelles activités s'ajoutent à celles qui existaient déjà, il résulte de ce mouvement une complexité de plus en plus grande du groupe bancaire. Le dispositif des rapports d'activité est alors modifié pour tenter de maintenir une certaine lisibilité à l'organisation de cet ensemble de plus en plus multiforme. Ce mouvement est sensible dès 1997, avec l'apparition de nouvelles rubriques tendant pour la plupart à présenter une vision schématique du groupe bancaire : les organigrammes se multiplient, de même que les restitutions géographiques, ainsi que les listes de filiales. Parallèlement, apparaissent des tentatives de recomposition du découpage de l'activité. Des pôles d'activité sont créés afin de regrouper ce qui autrefois faisait l'objet de présentations séparées. On voit ainsi la « Banque de proximité » venir encadrer les noms des principaux segments de clientèle (Marché des particuliers, des professionnels, etc.). Dans les deux cas, la recomposition de l'activité du groupe bancaire produit de nouveaux

espaces discursifs pour lesquels les dénominations propres sont fréquemment sollicitées. C'est ce mouvement général de recomposition qui traverse presque tout le corpus. L'année 2002 marque une sorte de coup d'arrêt à cette prolifération qui reprend cependant son cours en 2003, mais avec cette fois un doublement de la taille du vocabulaire¹. On peut chercher à savoir comment ces mouvements sont exprimés dans les textes.

3. Topographie des dénominations propres

Comme dans les deux précédentes sections, on examinera ici le comportement du groupe des dénominations propres pris dans son ensemble. Cet examen portera sur les deux des principaux contextes qui ont été identifiés dans la première partie et que l'on aura l'occasion d'utiliser dans les chapitres suivants.

On peut tout d'abord s'attacher à étudier la répartition des dénominations propres en fonction des contextes syntactiques et non syntactiques. Cette étude permettra d'éclairer ce qui vient d'être dit au sujet du développement des espaces de présentations des dénominations propres. Afin d'étudier ce point, on a projeté la liste des dénominations propres sur un graphique de répartition pour lequel le vocabulaire du corpus a été réparti entre les deux principales sortes de contextes présents dans les rapports d'activité.

¹ Ce mouvement est sensible dans le graphique d'accroissement du vocabulaire du corpus réduit aux dénominations propres : entre la première partie et la dernière, le nombre de dénominations propres par parties s'accroît régulièrement, en dehors de l'année 2002 assez atypique, et fait plus que doubler en 2001 et 2003.

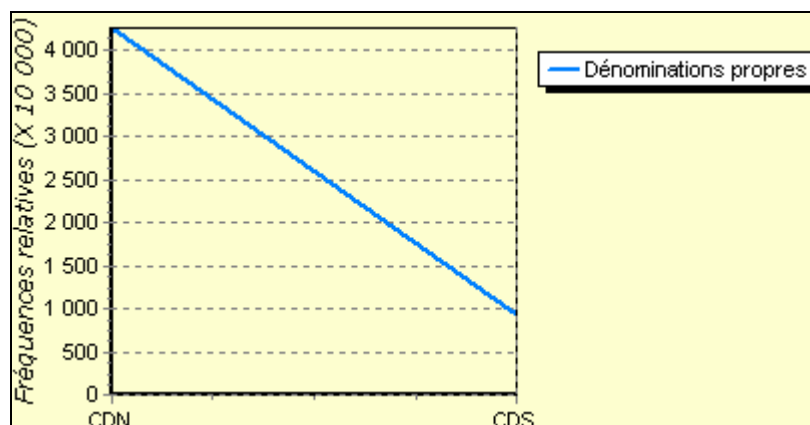


Figure 60.

Ventilations des dénominations propres en fonction des contextes syntactiques et non syntactiques du corpus rapports d'activité

Ce graphique permet de mettre en évidence la surreprésentation des dénominations propres dans les contextes de type non syntactique. On sait que les substantifs peuvent fonctionner isolément, en dehors du contexte habituellement fourni par la phrase, sous la forme de listes et, en l'occurrence, d'organigrammes. De ce point de vue, le développement de dispositifs destinés à synthétiser l'organisation du groupe bancaire dans les rapports d'activité a exploité une propriété « naturelle » des dénominations propres. Il conviendra de tenir compte de ce point lorsque l'on envisagera d'utiliser les dénominations propres pour capter le vocabulaire présent dans leur voisinage. En effet, il est probable que les informations collectées varient sensiblement selon qu'elles proviennent de l'un ou de l'autre contexte. Par ailleurs, il faudra prendre garde au fait que toutes les sortes de dénominations ne sont peut-être pas également réparties entre les deux contextes. Ces points seront illustrés dans les chapitres suivants. Il convient au préalable d'examiner un autre aspect de la répartition des dénominations propres à la surface du texte.

On observe maintenant la manière dont les dénominations propres sont réparties entre les différents paragraphes du corpus. Ce point doit être vérifié, afin d'estimer la représentativité de la collecte par rapport aux informations contenues dans les rapports d'activité. En effet, on doit chercher à capter un certain nombre d'unités présentes dans le voisinage textuel des dénominations propres. Autrement

dit, on utilisera les dénominations propres comme pivots dont on calculera les co-occurents. La portée théorique de ce point par rapport à la collecte d'informations sera examinée dans la section suivante. Il reste que l'espace textuel dans lequel s'effectue le calcul des co-occurents doit être défini en accord avec l'objet de la recherche.

Dans la deuxième partie, on a déjà eu recours à un calcul de ce type¹. Il s'agissait alors de préciser l'usage discursif de certaines unités terminologiques. Le paragraphe avait été retenu comme fenêtre de calcul des co-occurents, car il fournissait un bon compromis entre le volume de données à exploiter et l'exploration thématique que l'on en attendait. Qu'en est-il pour les dénominations propres ? Dans un premier temps, on tentera de répondre à cette question d'un point de vue théorique, en examinant les relations qui peuvent être tissées dans le discours entre ces unités et le découpage du texte en paragraphes. On examinera ensuite la manière dont les dénominations propres « couvrent » la surface textuelle du corpus, lorsqu'elle est découpée en paragraphes.

Les résultats d'un calcul des co-occurents dépendent, en plus de la méthode utilisée², de la fenêtre retenue. Celle-ci doit être à son tour déterminée par la nature des informations recherchées, et donc provenir d'une hypothèse sur la manière dont les données encyclopédiques attachées aux dénominations propres peuvent être réparties dans les textes. S'inscrivant dans la lignée des travaux de Chastain, Corblin, Kleiber, Charolles et d'autres, le travail de C. Schnedecker apporte des précisions relatives à la description des chaînes de référence initiées par des noms propres. Une chaîne de référence y est définie comme « *la suite des expressions référentielles (homogènes) encadrée par deux Np ou deux SN pleins identiques coréférents*³. » Dans un contexte où un seul Np est présent, ou si la saillance d'un nom propre ne fait pas de doute, sa reprise, appelée « redénomination », manifesterait la volonté du locuteur d'aborder son sujet sous un éclairage thématique

¹ Voir, entre autres, le chapitre 6.

² La méthode utilisée ici est celle qui est mise en oeuvre par Lexico et qui a été proposée par P. Lafon sous le nom de cooccurrences spécifiques (P. Lafon, 1984). Pour la présentation d'autres approches : W. Martinez (2003).

³ C. Schnedecker (1997 : 190).

différent : « *Métaphoriquement, nous dirions que le Np instruit du fait qu'on peut fermer un fichier référentiel pour en ouvrir un autre du même dossier ou à propos du même référent. Linguistiquement parlant, le Np signalerait que le locuteur initie une nouvelle chaîne pour saisir le référent dans un contexte tout différent ou sans rapport nécessaire avec celui qui précédait*¹. » Dans cette perspective, la relation existant entre une chaîne de référence initiée par un nom propre et le découpage d'un texte en paragraphes ne relève ni d'une hypothétique régularité formelle, ni de la simple coïncidence. Elle s'explique par la rencontre de deux fonctionnements parallèles : d'une part, la constitution de blocs référentiels, imputable à la capacité que possède le nom propre d'empaqueter des données relatives à son référent, et d'autre part, le marquage textuel de blocs thématiques à l'aide des paragraphes. Par conséquent, il est nécessaire que la collecte de certaines des unités co-occurentes d'une dénomination propre soit réalisée dans l'espace textuel au sein duquel elles ont le plus de chances d'avoir été distribuées. D'autre part, on verra plus loin que la collecte d'unités telle qu'elle est conçue dans ce travail n'a pas besoin de considérer a priori que la dénomination propre doive jouer un rôle premier dans la construction thématique, comme cela semble être le cas pour les chaînes de référence. Le paragraphe constitue donc le palier textuel nécessaire et suffisant qui doit être adopté ici comme fenêtre pour la recherche des co-occurents des dénominations propres. À ce stade, il est encore nécessaire de pouvoir évaluer la portée de la collecte par rapport à l'ensemble des paragraphes du corpus parmi lesquels les dénominations propres se trouvent réparties. En effet, autant une bonne répartition pourra laisser escompter une collecte représentative du contenu du corpus, autant la concentration des dénominations propres dans certaines zones obligera à des explorations complémentaires, destinées à évaluer la nature des informations qui seront restées dans l'ombre.

Afin de savoir dans quelle mesure les dénominations propres offrent une couverture représentative de la surface textuelle du corpus, on a projeté la liste de celles-ci sur la carte des sections de Lexico. Dans cette carte, que l'on a donnée ci-dessous, les carrés représentent des paragraphes. Ceux qui sont colorés en rouge

¹ C. Schnedecker (1997 : 150-151).

contiennent au moins une partie de dénomination propre (forme simple avec majuscule initiale ou forme en capitales). On voit à l'aide de ce graphique que la couverture des paragraphes des textes du corpus est assez bonne, d'autant qu'une proportion importante de carrés blancs correspondent à une balise servant à identifier certains découpages textuels (cela correspond à 798 paragraphes sur 6048, soit 13 % du nombre total de paragraphes). Le nombre de paragraphes non sélectionnés, après avoir fait le décompte de ceux qui ne contiennent que des balises, s'élève à 417, soit un peu moins de 7 % de l'ensemble. Ce seuil est jugé tolérable, d'autant que la carte des sections permet d'identifier facilement les zones de texte où les dénominations propres sont absentes. Des observations complémentaires sont donc toujours possibles. Ainsi, on a pu préciser ce que l'on pouvait escompter de la couverture d'une collecte d'informations réalisées à partir des dénominations propres dans un corpus de rapports d'activité tels que celui que l'on utilise dans ce travail.

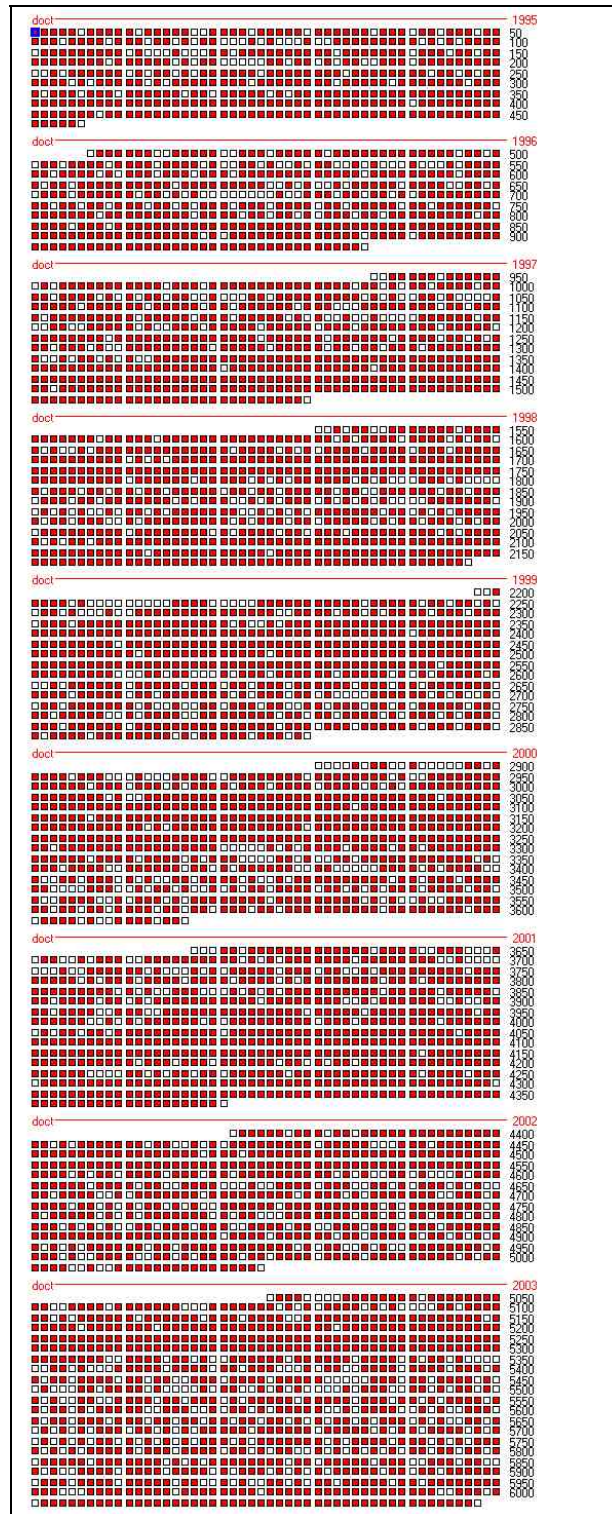


Figure 61.

Cartes des paragraphes du corpus RA sur laquelle ont été projetées les formes simples appartenant à une dénomination propre et ayant une majuscule initiale ou figurant en capitales

III. DÉNOMINATIONS PROPRES ET COMPOSANTS DE L'ACTIVITÉ BANCAIRE

A. Une catégorie atypique : l'agent « personne morale »

La place des dénominations propres dans la terminographie et la terminologie a été abordée dans la deuxième partie ainsi que dans la première section de ce chapitre. On a pu ainsi constater, d'une part, qu'elles sont bien intégrées dans la production terminographique, avec l'exemple du *Lexique de la banque et des marchés financiers* et, d'autre part, que la théorie éprouve une certaine réticence à leur intégration dans le champ de la terminologie. On a pu voir également que cette position connaît un infléchissement relativement récent, surtout lorsque le champ d'étude est étendu aux unités impliquant les capacités cognitives du locuteur. De ce point de vue, le fait de soumettre les dénominations propres à la grille des composants bancaires permettra d'approfondir la connaissance de cet ensemble d'unités dans le cadre du parler d'entreprise que l'on étudie.

Or, on peut également chercher aussi à caractériser plus avant la place tenue par ces dénominations dans l'utilisation qui est faite d'un parler dans le cadre du genre discursif auquel se rattachent les rapports d'activité. Dans la première partie, on a identifié d'un point de vue théorique la place centrale que devaient tenir les dénominations propres dans un référentiel : les rapports d'activité font-ils la même place à tous les objets dotés d'une dénomination propre ? Toujours dans cette première partie, on a constaté que les dénominations propres de personnes morales (entreprises) jouent un rôle important dans les bouleversements affectant le vocabulaire du corpus lorsque l'émetteur des rapports d'activité change de statut. On peut alors se demander si l'on n'a pas alors enregistré indirectement un phénomène plus général lié au rôle tenu par les dénominations de personnes morales dans les textes du corpus. De ce point de vue, l'étude du vocabulaire commun, menée dans la deuxième partie, a permis de mettre au jour des unités particulièrement présentes dans le corpus et appartenant à l'ensemble des dénominations propres. En effet, on a alors recensé des noms propres de personnes physiques et morales dans la catégorie des agents, ainsi que des toponymes, regroupés dans une nouvelle catégorie ouverte baptisée « divers », et réservée aux éléments ne relevant pas a priori des composants bancaires à proprement parler.

Afin de préciser ces points, on commencera par présenter et commenter la répartition des dénominations propres entre les différentes catégories de composants, pour formuler dans un deuxième temps des remarques relatives aux relations s'établissant entre les différentes catégories et les dénominations propres ; on abordera enfin les questions afférentes aux ensembles homonymiques auxquels appartiennent certaines dénominations propres.

Dans la deuxième partie, on a proposé de répartir certaines unités appartenant au vocabulaire du corpus en catégories générales caractéristiques des principaux composants de l'activité bancaire. Il s'agit des agents, des objets, des opérations, des ensembles complexes (regroupant ces trois catégories dans des cadres particuliers), et des propriétés. Comme cela a été rappelé plus haut, une sixième catégorie a été créée afin de faire entrer certaines unités du vocabulaire commun thématique ne relevant pas a priori à l'activité bancaire. La liste des dénominations propres a été soumise dans son ensemble à cette catégorisation¹. On donne ci-dessous le résultat de ce classement, avec pour rappel les classements précédemment effectués dans la deuxième partie.

¹ La liste correspond aux 2528 lignes du tableur utilisé pour répertorier les dénominations. Dans ce tableau, une ligne peut correspondre à deux dénominations lorsqu'une variante a été identifiée. Lorsque deux variantes ont été identifiées, une deuxième ligne a été créée, dans laquelle la dénomination propre de référence est reprise dans une première colonne, et la variante ajoutée dans une deuxième colonne, et ainsi de suite pour les autres variantes. Le cas le plus fréquent est celui dans lequel aucune variante n'est identifiée, puis une seule variante, etc. La catégorisation a donc été appliquée deux fois dans les cas où deux variantes ont été identifiées pour la même dénomination propre, mais il s'agit de proportions négligeables. En revanche, comme avec ce système les cas où une seule variante a été identifiée n'ont été comptabilisés que comme une seule ligne (ou dénominations propres), le nombre de lignes est inférieur au nombre total de dénominations propres, variantes comprises, qui s'élève à 2868.

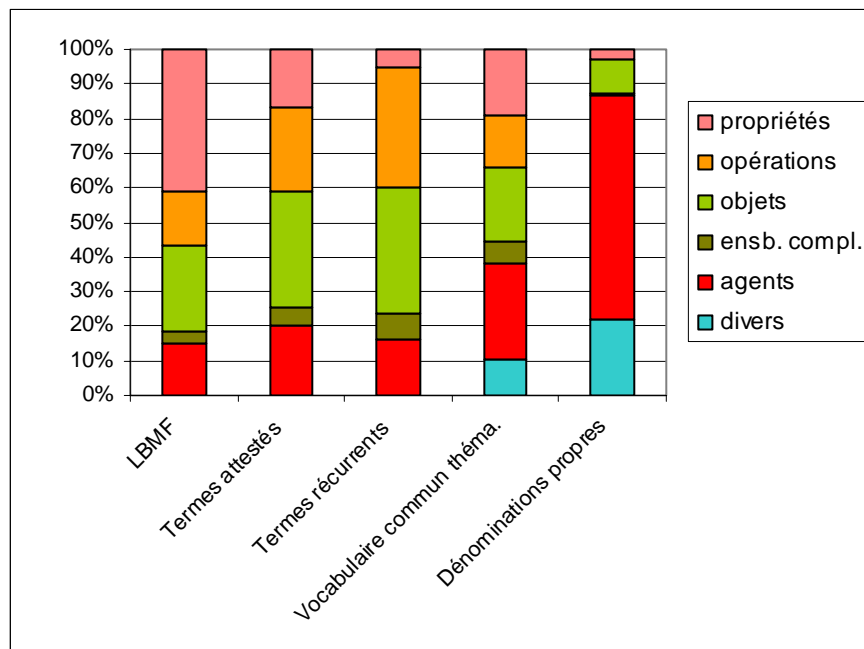


Figure 62.

Répartition des dénominations propres selon les principaux composants de l'activité bancaire et précédents classements

Le graphique présenté ci-dessus met en lumière la situation atypique des dénominations propres lorsque ce groupe d'unités est classé en fonction des principaux composants de l'activité bancaire. En effet, les unités relevant de la catégorie des agents représentent à elles seules les deux tiers de la population. Les dénominations propres d'opérations sont quasiment inexistantes (0,1 %). Les dénominations de propriétés, d'objets et d'ensembles complexes représentent à elles trois moins de 15 % de l'ensemble. Seule la catégorie « divers », apparue dans le cadre de l'étude du vocabulaire commun thématique, double sa taille avec un effectif d'un peu plus de 20 % de l'ensemble. Cette catégorie se décompose en « lieux », « secteurs¹ », « événements », « récompenses », « artefacts » et « infrastructures ».

¹ Les secteurs correspondent à des secteurs économiques valorisés par une majuscule et entrant par ailleurs dans un *continuum* d'homonymies qui rend difficile leur rattachement univoque à une catégorie ou à une autre. Les deux catégories concernées sont « agent » et « secteur » ; à ce stade un classement à grosses mailles a attribué une seule catégorie pour des unités homonymes. Le principe de ce classement repose sur l'usage majoritaire constaté. La première catégorie est retenue lorsque l'usage attesté est majoritairement celui du nom d'une direction ou d'un département pour une organisation. L'autre catégorie est retenue lorsque l'usage attesté est majoritairement celui d'un nom de secteur d'activité. Lorsque chaque unité est étudiée isolément, une distinction doit être faite entre les homonymes qui n'appartiennent pas à la même classe de composants.

Les toponymes représentent plus des deux tiers de la catégorie ; on revient plus bas et dans le chapitre suivant sur cette catégorie particulière. La sous-catégorie « secteurs » a une taille intermédiaire, proche de la centaine d'unités, alors que les autres sous-catégories ont des effectifs plus modestes d'une vingtaine d'unités chacune. On dispose donc à ce stade d'une indication générale sur la répartition des classes de dénominations propres dans le corpus. Pour préciser ce point, il est nécessaire d'examiner de façon plus détaillée la manière dont les principales catégories sont constituées.

Que ce soit d'après ses effectifs (1627) ou la proportion de fréquences élevées qu'elle comporte, la catégorie des agents arrive nettement devant les autres. On donne ci-dessous ses dix premiers items.

*Tableau 54.
Principales dénominations propres appartenant à la catégorie « agents » avec leurs fréquences et leurs variantes*

Dénominations	Variantes	Frq¹
Crédit Agricole	Crédit agricole	1699
Crédit Agricole	CREDIT AGRICOLE	1062
Caisses régionales	Caisses régionales de Crédit Agricole	659
Caisses régionales	Caisses régionales de Crédit agricole	659
Caisses régionales	Caisses Régionales de Crédit Agricole	659
Caisses régionales	CAISSES REGIONALES DE CREDIT AGRICOLE	659
Caisses régionales	Caisses Régionales	659
Caisses régionales	CAISSES REGIONALES	659
Groupe	GROUPE	603
Crédit Agricole Indosuez	CREDIT AGRICOLE INDOSUEZ	593
Crédit Agricole Indosuez	CAI	593
Crédit Agricole S A		410
Caisses locales	Caisses Locales	372
Crédit Agricole Asset Management	CREDIT AGRICOLE ASSET MANAGEMENT	182
Crédit Agricole Asset Management	CA Asset Management	182
Crédit Agricole Asset Management	CA AM	182
CNCA	Caisse nationale de Crédit agricole	175
CNCA	Caisse Nationale de Crédit Agricole	175
CNCA	CAISSE NATIONALE DE CREDIT AGRICOLE	175
CNCA	Caisse nationale	175
Sofinco	SOFINCO	165
Crédit Lyonnais		162

Les principales fréquences sont attribuées à des dénominations propres de personnes morales ou à des regroupements de personnes morales toujours présentés sous la forme d'un terme au pluriel avec majuscule initiale sur le premier membre de la dénomination ou sur tous ses membres (« Caisses régionales » et « Caisses locales »). Quant à « Groupe », il sert à désigner dans les reprises le syntagme « le groupe Crédit agricole ». La liste des agents se compose également de dénominations propres de personnes physiques à hauteur de plus de 20 % (380 unités sur 1627 que comporte la catégorie « agents »). Inférieure en termes

¹ La fréquence d'une dénomination propre correspond au cumul des fréquences de la dénomination principale et de ses variantes. Cette fréquence a été reportée pour chacune des lignes recensant des variantes.

d'effectifs à la sous-catégorie des personnes morales (1247 unités), la liste des dénominations propres de personnes physiques présente également des fréquences beaucoup moins élevées. On donne ci-dessous les principales d'entre elles, classées par ordre décroissant de fréquence.

*Tableau 55.
Principales dénominations propres de personnes physiques appartenant à la catégorie « agents »*

Dénominations	Variantes	Frq
Jean Laurent		32
Marc Bué		23
Bernard Michel		21
Thierry Coste		21
René Carron		18
Pierre Bastide	Pierre BASTIDE	15
Yves Barsalou	Yves BARSALOU	15
Gilles Gramat		14
Marc Antoine Autheman		14
Michel Villatte		14

La deuxième catégorie la mieux représentée parmi les dénominations propres du corpus correspond à un ensemble hétérogène dans lequel les toponymes occupent une place prépondérante. En effet, les 558 dénominations propres de cet ensemble se répartissent en 398 toponymes (71 %), en noms de secteurs (16 %), noms d'événements, de normes, d'artefacts et d'infrastructures, représentant chacune de 3 à 4 % de l'ensemble avec à chaque fois une vingtaine d'items. Les toponymes ne constituent pas une catégorie tout à fait homogène, et le chapitre suivant sera consacré aux différents usages qui en sont faits dans le corpus. Les noms de secteurs forment un ensemble aux contours difficiles à cerner, du fait, entre autres, de la présence de nombreux homonymes. Ils feront l'objet d'une étude plus détaillée à la fin de cette section.

Les quatre autres sous-ensembles de la catégorie « divers » regroupent des éléments qui sont a priori extérieurs à l'activité bancaire, comme les toponymes. Les noms d'événements concernent des événements extérieurs à l'activité du groupe bancaire mais auxquels celui-ci s'associe généralement sous la forme du mécénat (Assises de l'eau, France 98, Festival des régions, etc.). Il peut s'agir également d'événements aux implications économiques directes ou indirectes : an 2000,

Conseil de Madrid, etc. L'ensemble des normes regroupe pour sa part la législation et les chartes nationales et supranationales ayant une incidence sur l'activité économique de l'organisation ; il s'agit aussi bien de normes comptables publiques (M 14), que de normes liées au développement durable (Carbone Disclosure Project), de réglementation européenne (PAC) ou encore de lois nationales (loi Madelin). Le groupe des dénominations propres d'artefacts rassemble des objets particuliers ayant un rapport avec l'activité du groupe bancaire, soit de façon directe (participation au financement de la restauration de la frégate de La Fayette), soit indirecte, comme « Internet ». Enfin, le sous-ensemble relatif aux infrastructures rassemble essentiellement des bâtiments cités à l'occasion de la présentation, une nouvelle fois, des activités de mécénat.

Il ressort des remarques qui précèdent que les fréquences permettent d'isoler des ensembles, dénominations propres de personnes morales et toponymes, qui semblent plus particulièrement utilisés dans le cadre des rapports d'activité¹. Ces deux ensembles sont ainsi désignés comme devant faire l'objet d'une attention particulière dans la perspective d'une collecte du vocabulaire réalisée à partir des dénominations propres². En second lieu, le regroupement des dénominations propres a permis de soulever certaines questions relatives à leur catégorisation et susceptibles d'éclairer leur fonctionnement dans le corpus.

B. La catégorisation manuelle des dénominations propres

1. Les indices de type « internal evidence »

Le classement des dénominations propres en composants de l'activité bancaire, réalisé dans un premier temps à partir de la liste des seules dénominations,

¹ Les personnes morales possèdent avec les toponymes le plus grand nombre de formes simples dotées d'une majuscule et de fréquence supérieure ou égale à 10 (respectivement 183 et 53 formes). De même, parmi les 153 formes simples dotées d'une majuscule initiale (et entrant dans la composition des dénominations propres attestées) et présentes dans les neuf parties du corpus, 45 appartiennent à des dénominations propres, 21 à des toponymes, 20 à des noms de personnes physiques, 2 à des objets et 7 à diverses catégories ; 95 de ces formes sont utilisées dans des dénominations propres, et 58 appartiennent à des dénominations dont la fréquence est inférieure à 10.

² Il sera question des toponymes dans le chapitre suivant et des dénominations propres de personnes morales dans les deux derniers chapitres.

s'est révélé plus ou moins aisé à établir selon les cas. Ces variations deviennent mesurables lorsque l'on observe le temps consommé pour chaque arbitrage au moment de rattacher une dénomination à une catégorie. Le laps de temps nécessaire à cette opération s'étend d'une fraction de seconde à plusieurs minutes dans certains cas, lorsque le retour au contexte s'avère nécessaire. Afin de préciser ce qui distingue de ce point de vue les dénominations qui ont été recensées, on peut tenter de dégager les propriétés linguistiques sur lesquelles s'appuie le raisonnement qui préside au classement des dénominations propres.

Une première série d'indices permet de regrouper certains patrons très répandus, utilisés pour la formation des dénominations propres dans des contextes plus larges que celui qui est offert par les textes du corpus. Ces patrons permettent également des ajustements en fonction des contextes d'utilisation. Le corpus offre au moins deux exemples de ces constructions typiques. Il s'agit des noms de voies, ou odonymes, et des dénominations propres de personnes. Dans le premier cas, la construction typique est du type « classificateur pour le type de voie + Nom propre », comme dans « Avenue Victor Hugo ». Le second exemple se construit selon le patron « Prénom(s) + Patronyme(s) », comme dans « Jean Dupont ». Ce fonctionnement ordinaire du mode de construction de certaines dénominations propres a été utilisé dans les travaux visant à l'identification automatique des entités nommées. Du point de vue des traitements informatiques, cette propriété correspond à ce qu'après Mc Donald on a nommé *internal evidence*¹ et que l'on peut traduire par « indice interne à la dénomination ». Pour reprendre l'exemple des noms de personnes, l'indice est alors constitué par le ou les prénoms apparaissant dans les textes devant la forme ou la suite de formes dotées d'une majuscule initiale². En effet, une liste fermée de prénoms associée aux algorithmes de repérage des entités nommées permet alors de caractériser l'ensemble « prénom reconnu + forme

¹ « *Internal evidence is taken from within the sequence of words that comprise the name. This can be definitive criteria, such as the presence of known "incorporation terms" that indicate companies ("Ltd.", "G.m.b.H."), or it can be heuristic criteria such as abbreviations or known first names, which often indicate people.* » David D. Mac Donald (1996: 22).

² Lorsqu'une analyse morphosyntaxique a été utilisée au préalable pour étiqueter le corpus, les formes avec majuscule initiale sont fréquemment étiquetées lors d'une première passe comme appartenant à une catégorie inconnue.

inconnue avec majuscule initiale » comme correspondant à une dénomination propre de personne. Il reste que le recours au contexte fourni par la phrase (*external evidence*) est souvent nécessaire afin de procurer aux automates les indications nécessaires permettant, par exemple, de distinguer les noms de personnes de leur réutilisation sous la forme de marques commerciales.

2. *Fréquence ordinaire et patron type caractéristiques*

Le repérage manuel des dénominations est en général très rapide pour ce type de constructions. Il fournit, en outre, deux sortes d'indications utiles susceptibles de faciliter et donc d'alléger la tâche de collecte lorsqu'elle est réalisée dans un cadre textométrique. La première de ces indications est fournie par le décompte du nombre d'occurrences. On a évoqué précédemment le fait qu'une dénomination propre de personne pouvait être également utilisée dans le même contexte comme nom de marque commerciale. Ce phénomène de circulation des dénominations entre différentes classes de référents est courant. Cependant, lorsqu'une hésitation est possible au moment d'attribuer une catégorie à une dénomination propre, la fréquence ordinairement observée pour les membres de la classe pressentie est à même de confirmer ou d'infirmer le jugement. Pour illustrer ce point, on partira d'un exemple présent dans le corpus : « rue La Boétie » et « Rue La Boétie ». Si l'on est tenté de considérer à partir de la seule liste qu'il s'agit d'une dénomination et de sa variante, la fréquence de l'ensemble est de 17 occurrences, soit deux fois plus que la fréquence maximale observée pour les noms de voies. Cette indication chiffrée fait office d'alerte et incite à pousser le contrôle plus loin. On identifie alors la particule « SAS » placée devant l'odonyme et qui avait été assimilée à une scorie lors du repérage initial à partir de la liste des segments répétés. En réalité, cette particule fait bloc avec l'odonyme pour former une nouvelle dénomination propre dont le référent est une société anonyme¹ et non une rue. Contrairement à l'interprétation initiale, la différence de fréquences a permis de déclencher le processus de vérification à l'issue duquel deux dénominations propres distinctes

¹ « SAS » est l'acronyme de « société par actions simplifiée ».

cohabitant dans le corpus ont été identifiées : l'une correspondant à un odyonyme et l'autre à un nom de société.

Une autre indication est fournie assez rapidement lors du repérage manuel. Elle concerne l'utilisation préférentielle d'un patron de dénomination propre banal. Lorsque, par exemple, les dénominations propres de personnes apparaissent systématiquement, au fur et à mesure de la collecte, formées sur le patron « Prénom + Nom », il est permis d'extrapoler en considérant qu'il s'agit du patron utilisé de préférence à d'autres possibles. C'est que la présence répétée d'un patron type renseigne indirectement sur la propension des textes du corpus à faire se succéder des registres de langue ou des dispositifs textuels susceptibles de produire des séries de variantes ou de synonymes du type « François Mitterrand », « Mitterrand », « Tonton » ou « François¹ ». Autrement dit, lorsque l'on peut constater le recours récurrent au même patron dénominatif, on dispose d'une indication sur l'homogénéité de l'utilisation qui en est faite dans un corpus donné. Là encore, le cadre textométrique adopté pour la collecte permet d'acquérir une vision d'ensemble sur certains usages linguistiques et, partant, de limiter les tâches de contrôle en contexte qui seraient très fastidieuses. Cependant, en dehors de ces exemples, la collecte fait également apparaître des patrons qui semblent plus spécifiques du corpus étudié. Ceux-ci, loin d'être anecdotiques, sont parfois très productifs. Le cadre de réflexion fourni par l'étude des signalétiques peut aider à rendre compte de leur fonctionnement linguistique.

3. Signalétiques et classes de dénominations propres

C'est dans le cadre d'une réflexion ayant pris pour objet les titres de tableaux que B. Bosredon a été conduit à avancer la notion de signalétique pour rendre compte du fonctionnement de ce type de dénominations propres. Cependant, la portée de cette conception s'étend bien au-delà de ce seul objet d'étude, puisqu'elle propose d'articuler cadre pragmatique et contraintes linguistiques afin d'expliquer la formation de familles de dénominations propres correspondant à des usages sociaux très diversifiés : « *Contrainte en quelque sorte par la spécificité du domaine*

¹ Exemple emprunté à R. Eluerd (1985 : 142).

référentiel auquel elle s'applique, cette fonction [de nomination] génère des lignées de formes appellatives possédant un air de famille qui correspondent à des domaines d'appellation distincts et opposables entre eux. Si cette hypothèse est vérifiée, il doit exister par conséquent un mécanisme discursif contrôlant la production de formes appellatives construites puis réutilisées par les usagers du discours. On appellera signalétique l'emprise linguistique de ce dispositif de contrôle : elle se manifeste par un ensemble d'indices convergents, capables d'évoquer de façon plus ou moins directe le champ d'appartenance d'une dénomination construite¹. »

La prise en compte de la présence de signalétiques présente un double inévitement. La portée de cette information, au même titre que celles qui ont été mentionnées plus haut, permet de faciliter la collecte mais aussi de la fiabiliser : la reconnaissance d'un patron de dénomination rend également attentif aux exceptions. D'autre part, la mise en évidence de patrons implique la compréhension de leurs variantes, responsables à leur tour de variantes de dénominations souvent difficiles à identifier. On peut illustrer ce point à l'aide des variantes rencontrées pour les dénominations de certaines caisses régionales de Crédit agricole. Ainsi, pour la caisse régionale « Centre France », on rencontre dans le corpus les variantes suivantes :

- Centre France
- CENTRE FRANCE
- CRCAM Centre France
- CRCAM de Centre France
- Caisse régionale de Crédit agricole Centre France
- Caisse régionale de Crédit Agricole de Centre France

Les rôles d'alerte et de guide pour le repérage des variantes que peuvent jouer l'identification et la connaissance de signalétiques particulières sont d'autant plus nécessaires lorsque le genre textuel utilisé pour la collecte néglige d'autres

¹ B. Bosredon (1997 : 231). Le glossaire en début d'ouvrage propose cette autre définition : « *Signalétique. Ensemble de traces linguistiques (constituées de marqueurs récurrents et caractérisées par des sélections sémantiques privilégiées) entrant dans la fabrication de dénominations relatives à un domaine référentiel spécifique.* » *Ibid.* p. 9.

patrons plus répandus. En effet, la collecte des dénominations propres d'entreprises montre que la mention du type de société (extensions du type « SA » pour société anonyme) est relativement peu utilisée dans le corpus des rapports d'activité. Contrairement à ce qui a pu être observé par ailleurs dans des corpus de dépêches d'agences utilisés pour l'identification automatique des dénominations¹, la mention du type de société semble réservée ici aux dénominations n'apparaissant que dans des contextes non syntactiques, en dehors de l'exception constituée par « Crédit Agricole SA ». Dans le corpus, ces extensions ne sont utilisées que pour 2 % des occurrences de dénominations propres d'entreprises². En dehors de cet aspect, d'autres régularités ont néanmoins pu être mises en évidence. Celles-ci paraissent relever de signalétiques particulières adoptées pour les dénominations propres d'entreprises appartenant au Crédit agricole.

On illustrera ce point à l'aide de l'exemple fourni par la dénomination propre « Crédit agricole », qui fonctionne à la fois comme dénomination propre du groupe bancaire Crédit agricole et comme marque commerciale utilisée de manière exclusive par la banque éponyme. On peut avoir une idée de la productivité de la réutilisation de la dénomination comme marque, en interrogeant la base de données recensant les marques commerciales déposées pour la France et, plus récemment, pour l'Europe³ : ce sont environ trois cents dénominations qui font actuellement l'objet d'un dépôt. Cependant, on s'intéressera ici aux cas de réutilisation de la dénomination pour former d'autres dénominations propres d'entreprises. Cet usage semble traduire le souci « ontologique » de ceux qui établissent les dénominations, de marquer par ce moyen l'appartenance au groupe bancaire des entités nouvellement acquises. Cependant, la législation ne les oblige pas à suivre précisément les patrons que l'on va présenter plus loin. D'ailleurs, les exceptions

¹ On en trouve des exemples dans B. Boguraev et J. Pustejovsky (1996).

² Les formes « AB, AG, AS, DTVM, GmbH, Inc, Ltd, PTE, PVT, SA, SAL, SNC, SVB », correspondant à des types de sociétés totalisent, en dehors du cas atypique de « Crédit Agricole S.A. », 290 occurrences sur un ensemble de 13 000 occurrences de dénominations propres de personnes morales recensées dans le corpus.

³ Ce service de l'Institut national de la propriété industrielle (INPI) est accessible à l'adresse <http://www.icimarkes.com/> consultée en août 2007. La liste des 277 noms de marques déposés est donnée en annexe.

sont nombreuses pour lesquelles l'entreprise achetée conserve sa raison sociale d'origine. En effet, la notoriété acquise au préalable par la marque constitue parfois un obstacle à l'application de la signalétique habituellement utilisée.

La dénomination propre « Crédit agricole » a été utilisée pour former 159 des dénominations propres d'entreprises qui ont été recensées dans le corpus. Pour en faciliter la présentation, il convient au préalable de mentionner certaines variantes qui ont été mises au jour durant la collecte. La dénomination propre « Crédit Agricole » s'abrège dans le sigle « CA », mais il s'agit d'une version relativement récente de la dénomination qui a pour origine « Crédit Agricole Mutuel » avec le sigle correspondant « CAM ». Le remplacement d'une version par l'autre s'est effectué, surtout dans le domaine du marketing et de la publicité, à partir du milieu des années 1970, mais les deux dénominations coexistent encore dans les secteurs où prévalent les aspects institutionnels et juridiques¹. On donne ci-dessous les différents patrons réutilisant la dénomination.

Le premier patron générique est du type « Crédit Agricole + X » la dénomination propre initiale étant placée en tête de celle qu'elle permet de construire, à la manière d'un régissant. Les autres formants appartiennent à une liste fermée, mais peuvent se combiner entre eux. Il s'agit du nom d'une autre entreprise (NE), d'une caractérisation du type de société ou d'organisation (TS), du rattachement à un domaine d'activité (DA) et, enfin, de la localisation géographique de l'activité (LA). On trouve une douzaine de combinaisons utilisées pour soixante-huit dénominations propres d'agents correspondant à des personnes morales (entreprises du secteur financier). On donne ci-dessous des exemples de combinaisons :

1. Crédit Agricole + NE : Crédit Agricole Indosuez
2. Crédit Agricole + TS : Crédit Agricole S.A.
3. Crédit Agricole + DA : Crédit Agricole Asset Management.
4. Crédit Agricole + LA : Crédit Agricole Suisse

¹ À titre d'exemple, le code monétaire et financier ne connaît que les caisses de « crédit agricole mutuel » ; il s'agit alors d'un terme et non d'une dénomination propre.

5. Crédit Agricole + DA + TS : Crédit Agricole Alternative Investment Product Group Inc
6. Crédit Agricole + DA + TS : Credit Agricole Futures Inc
7. Crédit Agricole Indosuez + NE : Crédit Agricole Indosuez Cheuvreux
8. Crédit Agricole Indosuez Cheuvreux + LA + TS : Crédit Agricole Indosuez Cheuvreux Deutschland GmbH

Ces quelques exemples permettent d'illustrer également le fait que la combinatoire est contrainte : le type de société apparaît toujours à la fin, le pays d'exercice de l'activité en dernière ou en avant-dernière position. Par ailleurs, ces dénominations sont susceptibles d'évoluer par accumulation comme on peut le voir dans les exemples 1, 5 et 8, où la dénomination initiale se charge progressivement d'autres dénominations propres qui rendent possible à leur tour de nouvelles combinaisons, etc. Il peut résulter de ce processus des dénominations relativement longues, mais on remarque que celles-ci excèdent très rarement les sept membres dans le corpus. Il semble que l'on atteigne alors certaines limites en termes de « maniabilité¹ » de la dénomination. Cela explique probablement l'existence des sigles et de la suppression très fréquente de la mention du type de société dans les contextes syntactiques. On dispose alors pour les exemples 5 et 8 de raccourcis, tels que CA AIPG et CAIC Deutschland, auxquels l'émetteur recourt fréquemment.

Deux autres patrons génériques utilisent la dénomination propre « Crédit Agricole », sous les formes « X + Crédit Agricole » ou « X + Crédit Agricole + X ». Cependant, il faut aussi constater que ces constructions s'appliquent de façon remarquable à des sous-ensembles bien spécifiques de personnes morales. Autant le cas le plus fréquent qui a été illustré précédemment semble dédié à des sociétés privées appartenant au secteur financier international, autant les deux autres patrons semblent réservés soit à des composantes institutionnelles, soit aux entités historiques du groupe bancaire. Le premier cas est représenté par des dénominations du type « Fédération Nationale du Crédit Agricole », le second est dominé par les dénominations complètes des caisses régionales de Crédit Agricole, à l'instar de

¹ On reprend le terme proposé par R. Kocourek comme critère destiné à vérifier l'usage de termes complexes particulièrement longs (1991 : 140).

« Caisse régionale de Crédit agricole Alpes Provence ». Enfin, il faut noter que l'on retrouve les mêmes formants que pour le premier patron, qu'il s'agisse du type de société ou d'organisme (« Fédération Nationale », « Caisse régionale »), ou de la localisation géographique de l'activité (« Alpes Provence »).

Comme cela a été indiqué plus haut, la dénomination complète existe sous la forme du sigle « CA » ou « CAM ». Ces sigles entrent à leur tour dans la composition des sigles ou des acronymes des dénominations propres réutilisant « Crédit Agricole », comme dans les exemples donnés précédemment. On note aussi que l'utilisation de ces sigles tend parfois à leur conférer un rôle d'affixe, dans la mesure où la dénomination propre dans laquelle ils entrent ne se présente plus elle-même comme un sigle ou un acronyme.

Avant d'aborder ce point, il convient de présenter les patrons dans lesquels entrent ces deux sigles. Le sigle « CAM » est principalement utilisé dans les sigles ou les acronymes de dénominations complètes composées à l'aide de « Crédit Agricole Mutuel ». Dans les dénominations propres attestées on trouve, par exemple, « Cedicam », acronyme de « Centre d'échanges de données et d'informations du Crédit agricole mutuel ». Le sigle « CA » est quant à lui utilisé dans les sigles et les acronymes de dénominations complètes formées sur le modèle « Crédit Agricole + X ». On trouve, par exemple, « CA AM » pour « Crédit Agricole Asset Management ». Il est également utilisé dans la configuration rencontrée pour « CAM » lorsque ce dernier n'a pas été retenu. C'est le cas, entre autres, pour « CNCA », sigle de la dénomination propre « Caisse nationale de Crédit agricole ».

Les deux sigles « CAM » et « CA » sont par ailleurs utilisés à la manière des affixes lorsqu'ils entrent dans la formation de dénominations propres ne possédant pas de version développée. En effet, en l'absence de « version longue » ils continuent à jouer le rôle d'une particule marquant le rattachement du porteur de la dénomination au groupe bancaire Crédit Agricole. Il est probable que cette marque d'appartenance au groupe bancaire passe inaperçue chez la plupart des locuteurs. En revanche, on dispose d'indices relatifs à sa reconnaissance chez les salariés de la banque. Celle-ci s'est manifestée, par exemple, au moment de la création d'Indocam en 1997. On a alors été témoin de conversations au cours desquelles les avis

divergeaient sur la manière d'interpréter l'acronyme « Indocam » : certains interlocuteurs penchaient pour une interprétation reposant sur le patron « Indo[suez]¹ + Crédit Agricole Mutuel / CAM », Indocam étant né du rapprochement des gestions d'actifs de la banque Indosuez et du Crédit Agricole. On a vu plus haut que le patron « NE + NE » est relativement courant. D'autres locuteurs étaient d'un avis différent, et comprenaient l'acronyme comme une forme raccourcie de « Indo[suez] + c + A[sset] M[anagement] ». Le patron « NE + DA » constituant bien une formation possible, la présence inhabituelle de la lettre « c » étant expliquée par le souci de donner à la nouvelle dénomination une consonance harmonieuse. En admettant que les patrons de nomination utilisés acceptent aussi le recours à l'apocope, les deux interprétations peuvent être soutenues, avec peut-être un léger avantage en faveur de la première qui paraît plus « naturelle ». C'est pourtant la deuxième possibilité qui correspond à l'intention des créateurs de la dénomination². On ne peut cependant pas exclure que l'utilisation de « CAM » ait été perçue par ceux-ci comme offrant une ambiguïté opportune.

Plus récemment, la dénomination « Calyon » a été substituée à « Crédit Agricole Indosuez » à la suite de l'intégration des activités de banque d'investissement du Crédit Lyonnais. Alors que l'on retrouve le sigle « CA » couplé avec le résultat d'une apocope pratiquée sur « Lyonnais », la dénomination propre ne correspond pas à une forme développée qui pourrait être « Crédit Agricole Crédit Lyonnais ». Il en va de même pour « Pacifica » ou « Predica³ », etc⁴. Le sentiment linguistique dont il a été question dans le paragraphe précédent peut, par ailleurs, s'effacer devant une consonance heureuse, quand bien même celle-ci n'a pas été souhaitée par les créateurs de la dénomination. Lorsque le nom de la Caisse

¹ La dénomination « Indosuez » ou « Banque Indosuez » résulte d'une apocope pratiquée sur « Indochine » et de l'adjonction de « Suez ». Cette dénomination propre est apparue dans un contexte similaire, à la suite de la fusion dans les années 1970 entre la Banque de Suez et la Banque d'Indochine.

² On pourrait également trouver une double motivation correspondant à cette façon de voir : Indo[suez] + CA + AM, sur le patron NE + NE + DA qui est à notre connaissance moins fréquent que ceux qui ont été mentionnés plus haut. En revanche, le sigle CA AM, patron NE + DA, a été utilisé pour baptiser une deuxième fois « Indocam ».

³ Filiales d'assurance dommages et d'assurance-vie du Crédit Agricole.

⁴ La liste complète des dénominations réutilisant « Crédit Agricole », « CAM » et « CA » est donnée en annexe.

nationale de Crédit agricole est devenu, à la suite de son entrée en bourse, « Crédit Agricole S.A. », il ne faisait pas de doute pour les dirigeants du groupe bancaire que la version abrégée de la dénomination se prononcerait à la manière d'un sigle : « C-A-S-A », sur le modèle de « P-S-A » pour « Peugeot S.A. ». Quelques jours après la diffusion du nom, la prononciation [kaza] était néanmoins adoptée aussi bien par la presse économique que par les salariés de l'entreprise. Il n'en reste pas moins que la connaissance de ces patrons dénominatifs facilite le classement des dénominations propres. On donne ci-dessous, à titre d'illustration, les formes simples dotées d'une majuscule initiale, identifiées comme correspondant à des dénominations propres (simples ou abrégées sous la forme d'un sigle), ou à certains de leurs composants, et se terminant par « CA » ou « ca »¹ :

*Tableau 56.
Formes du corpus avec majuscule initiale et terminaison en « ca »*

Formes	Frq.	Formes	Frq.
CNCA	107	Predica	133
FNCA	44	Pacifica	89
PACIFICA	9	Banca	60
PREDICA	9	Unifica	19
UNIFICA	9	Sodica	10
CETCA	3	Etica	4
SODICA	3	Camca	3
IBCA	2	Batica	2
APCA	1	Africa	1
BANCA	1	Fireca	1
CLCA	1	Oreca	1
CNMCCA	1	Pratica	1
SICA	1	ca ²	1
SOFIPACA	1		

Sur les vingt-huit formes ainsi recensées, les trois-quarts possèdent la particule « ca » pour marquer qu'il s'agit d'entreprises ou d'organisations rattachées

¹ La liste des formes est obtenue à partir de deux expressions régulières « ca\> » et « CA\> » qui signifient que l'on recherche les chaînes de caractères se terminant par « ca » ou « CA ». Ces expressions régulières ont été automatisées dans le module de recherche de Lexico où il est possible d'utiliser un patron de recherche automatique « est la fin de ce que je recherche ».

² « CA » entre dans les sigles de dénominations propres utilisant « Crédit Agricole » dans leur version développée. La seule occurrence de « ca » est recensée dans une adresse de site : www.ca-ch@nge-enligne.com.

au groupe Crédit Agricole. Sur les sept formes qui ne peuvent pas être interprétées de cette manière, on compte trois formes étrangères (« BANCA », « Banca » - « banque » en italien -, et « Africa » en anglais), un sigle entrant dans la dénomination propre d'une agence de notation (« Fitch IBCA »), et trois sigles correspondant à des dénominations propres d'organismes agricoles (« APCA » - Assemblée Permanente des Chambres d'Agriculture -, « CNMCCA » - Confédération Nationale de la Mutualité, de la Coopération et du Crédit Agricoles – et « SICA » [Habitat Rural] – Société d'intérêt collectif agricole Habitat Rural).

La prise en compte d'une signalétique locale n'est donc pas dénuée d'intérêt pour la collecte, même si elle n'écarte pas à l'avance certaines exceptions (on comptera dans ce cas un bruit de 25 % représenté par les sept formes possédant bien une terminaison en « ca », mais ne correspondant pas à des dénominations d'entreprises rattachées au Crédit Agricole ; le bruit tombe à 4 % si l'on comptabilise les dénominations ou les parties de dénominations d'entreprises et d'organisations en général.). Cependant, la signalétique ne livre pas toujours les indications suffisantes pour classer les dénominations à partir d'une simple liste. Le fonctionnement propre à certaines unités nécessite que l'on complète la faiblesse des indices disponibles dans la dénomination candidate par un retour au contexte.

C. Les secteurs d'activité : existence d'un *continuum* homonymique

Le découpage de l'activité en secteurs ou domaines a déjà été abordé dans la deuxième partie, lorsqu'on a relevé l'utilisation particulière de certains termes présents dans le corpus. En effet, il est alors apparu que les termes servant à désigner des objets ou des types d'opérations bancaires étaient également utilisés pour caractériser le domaine ou le secteur dans lequel ces objets ou activités prennent place ordinairement. Ainsi, la « conservation de titres » ou l'« affacturage » renvoient-ils à des formes de l'activité financière mais aussi aux domaines correspondants. Cette double utilisation se marque, entre autres, par l'apparition de la majuscule sur la première partie de la dénomination ou sur chacun de ses membres. On relève pour une unité appartenant à la phraséologie des rapports d'activité du corpus la répartition suivante : « banque de proximité » (56 occ.), « Banque de proximité en France » (36 occ.) et « Banque de proximité » (10 occ.).

La présence de la majuscule permet fréquemment de repérer des rôles différents selon que l'unité apparaît dans des contextes de type syntactique ou non syntactique. Dans les premiers, la majuscule initiale peut s'expliquer par la présence de la dénomination dans un titre. Dans les seconds, l'apparition isolée du syntagme lexicalisé, soit dans une table des matières, soit dans certains organigrammes ou histogrammes, révèle qu'un rôle particulier lui a été dévolu dans le découpage de l'activité que propose les rapports d'activité.

Le fait que certaines unités terminologiques servent, en plus de leurs utilisations habituelles, à dénommer des secteurs d'activité, semble autoriser une opération métonymique supplémentaire. En effet, il est fréquent que ce découpage corresponde également à une subdivision de l'organisation ou de l'entreprise qui a en charge le suivi, le contrôle ou le développement des activités impliquées dans le secteur ou le domaine éponyme. Ces subdivisions portent ordinairement le nom de « directions » ou « départements », mais le même phénomène s'applique à tous les niveaux de l'organisation de l'entreprise. De ce point de vue, lorsque les rapports d'activité communiquent, par l'intermédiaire des organigrammes, les noms des directions et des départements de l'entreprise, ils restent au même niveau de généralité que celui qui caractérise la restitution de l'activité dans son ensemble. On conçoit bien que s'ils devaient donner le détail de l'organigramme de toutes les entreprises du groupe bancaire, il n'y aurait pas de place pour d'autres informations.

La collecte initiale, qui s'est étendue délibérément aux unités valorisées par une majuscule initiale, a permis de mettre au jour des séries homonymiques qui font partie intégrante du mode de fonctionnement des unités concernées¹. Leur étude montre en fait l'existence d'un *continuum* au sein duquel une unité est susceptible d'endosser différentes fonctions :

1. unité relevant d'un usage non spécialisé a priori
2. unité terminologique désignant un concept ou une notion
3. unité terminologique désignant un concept ou une notion et un secteur d'activité ou un marché

¹ La liste est donnée en annexes.

4. unité terminologique utilisée hors contexte syntagmatique avec majuscule initiale
5. dénomination d'une subdivision de l'organisation
6. dénomination d'un regroupement de subdivisions de l'organisation et/ou de domaines ou de secteurs d'activité.

On illustrera un premier type de circulation faisant passer une unité du statut 1 aux statuts 3, 4 et 5 avec le mot « particulier ». Tout d'abord, il est nécessaire de préciser que le mot se répartit en deux utilisations tranchées selon qu'il est au pluriel ou au singulier. Pour ce dernier usage il entre dans la locution adverbiale « en particulier » à hauteur de 44 occurrences sur 47, les trois occurrences restantes correspondant à l'adjectif. Lorsqu'il est au pluriel, le mot correspond au substantif pris au sens d'une collection de personnes privées, par opposition aux collectivités organisées. Les 160 occurrences sont utilisées pour caractériser un segment de clientèle du groupe bancaire. Il s'oppose à ce titre à d'autres marchés, comme dans cet exemple : « (...) *auprès des particuliers, des agriculteurs, des professionnels, des collectivités locales et des entreprises.*¹ ». Les 11 occurrences de la forme avec majuscule initiale sont majoritairement présentes en contextes à dominante non syntactique. On trouve, par exemple, plusieurs entrées de table des matières avec cette graphie : « Les Particuliers ». Mais l'utilisation dominante correspond à la présence de la graphie avec majuscule initiale comme nom de direction, ou comme partie d'une telle dénomination. Dans le premier cas, on a seulement « Particuliers » ; dans le second, on trouve, par exemple, « Marché des Particuliers et des Professionnels », qu'il faut comprendre comme « Direction du marché des Particuliers et des Professionnels ». Dans les usages discursifs, les régissants « direction » et « marché » sont fréquemment omis, et il n'est pas rare de rencontrer des formulations métonymiques du type « Les Particuliers n'ont pas assisté à la réunion. », qui doivent être comprises comme « Les représentants de la Direction du marché des Particuliers n'ont pas assisté à la réunion. »

¹ Rapport d'activité 2000, p. 29.

On avait observé précédemment¹ que le terme « moyen de paiement » sert à désigner « *tous les instruments qui permettent à toute personne de transférer des fonds, quel que soit le support ou le procédé technique utilisé*². » Ce terme est utilisé dans le corpus comme générique de « carte bancaire », ou « chéquier ». Toutefois, son utilisation principale réside dans la désignation d'un marché, comme dans ce début de phrase : « *Leader sur le marché des moyens de paiement, le Crédit agricole (...)* ». Un marché se caractérisant par une clientèle, des concurrents, des parts de marché, et des produits et services, il fonctionne aussi comme un domaine d'activité englobant l'activité du Crédit Agricole : « (...) *dans les domaines de l'assurance, des moyens de paiement, de l'affacturage (...)* ». Le terme est utilisé comme intertitre pour le découpage en rubriques des rapports d'activité, et il apparaît pour caractériser l'activité d'une filiale dans les contextes de type non syntactique : « *CEDICAM : Moyens de paiement* ». Dès lors, il devient assez probable de le rencontrer utilisé tel quel ou intégré dans une dénomination propre de subdivision de l'organisation, par exemple dans un organigramme : « *Jean - Pierre Ledru / Moyens de paiement* », qui doit s'interpréter dans l'organigramme de la CNCA tel qu'il est donné dans le rapport d'activité de 1999 : « Jean-Pierre Ledru est responsable de la direction des Moyens de paiement, rattachée au pôle Développement et marchés de la CNCA. » On voit par cet exemple, que le même terme (2) est utilisé, par le biais de métonymies successives, pour désigner un marché (3), un domaine d'activité (4), puis est employé comme dénomination propre pour une direction de l'entreprise CNCA (5). Les deux dernières utilisations produisent des unités complexes homonymes avec majuscule initiale, mais dans le premier cas il s'agit d'une simple valorisation graphique d'un terme haussé au rang de domaine d'activité, alors que le second correspond à une dénomination propre. Le repérage à partir d'une simple liste de dénominations ne permet pas de détecter ce type d'ambiguïté. Seul le recours au contexte permet de le faire, mais encore faut-il avoir pris connaissance de l'usage de cette forme particulière de signalétique qui s'appuie sur les variétés de

¹ Voir la deuxième partie.

² Code monétaire et financier, Art. L. 311-4.

langue et le processus métonymique pour faire de la dénomination propre d'une direction d'entreprise l'ultime avatar d'un syntagme terminologique.

En résumé, on a tenté dans ce chapitre de traiter par des moyens textométriques une problématique habituellement prise en charge par les traitements automatiques du langage naturel. Cependant, loin de constituer une régression par rapport à l'emploi de programmes plus sophistiqués, l'approche robuste présentée ici permet d'aborder de façon assistée par ordinateur l'extraction d'une catégorie d'unités à partir d'un corpus textuel. Les objectifs assignés à cette extraction – la constitution d'un référentiel terminologique adapté au contexte afin de documenter certaines activités d'information – autorisent l'usage de ressources textuelles dont la taille ne nécessite pas de « *maîtriser les déluges de données hétérogènes*¹ ». Par ailleurs, l'extraction des noms propres peut s'appuyer dans le cadre textométrique sur un certain nombre de facilités mises à contribution à chaque étape de la démarche proposée.

Au préalable, il a été nécessaire de clarifier la notion de nom propre retenue afin de réaliser la collecte. Le périmètre de celle-ci a été étendu sur plusieurs points. Par rapport au nom propre grammatical limité à la catégorie des substantifs, la notion de dénomination propre, qui a été retenue, embrasse toutes les sortes d'unités simples ou complexes, syntagmatiques ou phrastiques, utilisées comme formes fixes de dénomination d'un référent unique. En second lieu, la collecte porte sur les noms propres prototypiques (noms de personnes et de lieux), mais elle intègre aussi des dénominations dont le statut fait encore l'objet de débats, comme les noms de produits. Ces derniers, se comportant selon les usages comme des termes ou des dénominations propres, ont été rattachés à cette dernière catégorie. En effet – et c'est la troisième sorte d'ouverture adoptée pour cette collecte – on a pris le parti de sélectionner un sous-ensemble au sein d'un ensemble plus vaste composé des unités qui relèvent des capacités plus cognitives que linguistiques des interlocuteurs. L'usage de la majuscule permet de ce point de vue de marquer la frontière entre les unités collectées à ce stade et celles qui ne le sont pas. En effet, on a admis, pour la phase de repérage de la collecte, que des mots dotés d'une majuscule initiale, à la manière des noms propres, sans pour autant appartenir à cette catégorie, puissent

¹ On reprend le titre d'un article de G. Illouz, B. Habert, S. Fleury *et alii* (1999).

être également sélectionnés. L'adoption de ce périmètre relativement large a été motivée par le souci d'intégrer le plus d'unités possibles, qui soient d'une manière ou d'une autre en prise avec le référentiel à explorer, que la majuscule indique un nom propre prototypique ou une unité que le locuteur a souhaité distinguer des autres éléments de sa classe par ce procédé.

En second lieu, les traitements textométriques utilisés ont permis, d'une part, d'automatiser une partie de la collecte et, d'autre part, de procéder aux vérifications en contexte rendues nécessaires par les tâches de désambiguïsation. Par la segmentation des formes ainsi que la constitution automatique d'un dictionnaire, on a pu isoler automatiquement les formes simples dotées d'une majuscule initiale et les classer par ordre de fréquence décroissant. Le calcul des segments répétés a rendu possible le recensement de toutes les suites de formes répétées au moins deux fois dans le corpus, et leur dictionnaire a offert les mêmes possibilités de tri que celles qui ont été utilisées pour les formes simples. Le repérage des unités a consisté ensuite à utiliser chaque forme simple dotée d'une majuscule initiale pour rechercher la présence de dénominations complexes ou confirmer l'usage de la forme simple comme dénomination propre. Pour les formes simples dotées d'une fréquence élevée (on compte, par exemple, près de trois mille occurrences pour « Crédit »), la recherche des dénominations propres s'est effectuée à l'aide du dictionnaire des segments répétés. On a alors pu constater l'efficacité de l'approche, dès lors que les segments sont triés par ordre décroissant de fréquence, les dénominations bien formées étant généralement encadrées par celles qui sont incomplètes et celles qui sont associées à une particule étrangère captée par le segment répété. Pour les fréquences faibles, de une à cinq occurrences, une projection par lots de cinq à trente formes sur les concordances de Lexico a permis d'alléger un repérage, qui sans cela aurait été fastidieux. Le repérage des variantes a, par ailleurs, permis de rassembler autour d'une dénomination propre principale des dénominations annexes utiles pour la recherche d'information. Ces dénominations ont été consignées dans un tableur de manière à en faciliter le tri et les exploitations futures. Mais ce n'est pas le seul apport d'une telle collecte.

En effet, celle-ci a rendu possible une meilleure connaissance de la présence de près de trois mille dénominations propres dans le vocabulaire et les textes du

corpus. Pour plusieurs raisons (définitions différentes des dénominations propres¹, genres discursifs et époques dissemblables), ce nombre peut difficilement être comparé avec les informations déjà collectées à partir d'autres corpus. Cependant, lorsqu'il est rapporté au seul vocabulaire du corpus, il permet de mettre en évidence la présence des dénominations propres dans le vocabulaire du genre étudié, ce qui apporte une première réponse aux interrogations formulées en introduction de ce chapitre. Le recensement presque exhaustif des dénominations propres (on a précisé qu'il est de l'ordre de 98% pour la mesure P&R) a rendu possible un affinement de l'évaluation de la présence de cette catégorie d'unités.

On a ainsi pu voir que leur accroissement suit celui du vocabulaire du corpus pris dans son ensemble. Si les noms propres sont présents dans toutes les parties du corpus, leur répartition est susceptible de varier. Les évolutions majeures du groupe bancaire sont répercutées de manière indirecte dans les textes des rapports d'activité, ce qui peut provoquer un afflux ou un reflux de dénominations propres. Si ces dernières sont présentes dans la majorité des paragraphes des textes du corpus elles sont surreprésentées dans les contextes non syntactiques, ce qui confirme leur capacité à s'affranchir de la phrase pour être intégrées aux dispositifs non syntactiques.

Par ailleurs, les dénominations propres présentent une configuration originale si elles sont réparties en fonction des grandes catégories de composants de l'activité bancaire. On note sur ce point une surreprésentation de la catégorie des agents et des toponymes. Cela permet d'identifier une première limite concernant la portée de la collecte que l'on propose de réaliser. En effet, la présence des dénominations propres doit être considérée à la lumière de deux types de sélections : la sélection socialement normée des référents, qui a été présentée dans le premier chapitre, et la sélection parmi les composants de l'activité bancaire telle que celle-ci est restituée dans les rapports d'activité. Ces sélections sont doublées par une organisation

¹ Les remarques relatives au domaine textométrique peuvent être étendues à l'étude des registres de langue menée, par exemple, par D. Biber. Parmi les caractéristiques linguistiques prises en compte par cet auteur, figurent les noms propres grammaticaux (par ex. D. Biber 1993 et 1995), ce qui peut s'avérer trop restrictif, ou les expressions référentielles, ce qui revient à prendre en compte également les descriptions définies en plus des noms propres, ce qui est trop large (D. Biber, S. Conrad & R. Reppen, 1998 : 108-122).

hiérarchique de chaque classe de dénominations propres. En effet, les unités possèdent des fréquences très variables entre classes mais aussi à l'intérieur de chaque classe, où les fréquences faibles et les hapax sont toujours plus nombreux. Il semble que ce deuxième choix puisse être rapproché de l'importance prise par les référents dénommés dans le référentiel du groupe bancaire, tel qu'il est restitué dans le genre de discours des rapports d'activité. Cette collecte est donc de nature à préciser les relations qui s'établissent dans un genre de discours donné, entre dénominations propres et référentiel. En effet, on dispose ainsi d'indications, d'une part, sur les pans de référentiel qui sont privilégiés pour la restitution de l'activité bancaire, et d'autre part, sur la hiérarchie des entités à l'intérieur d'une sphère d'activité.

Enfin, cette collecte constitue une bonne introduction à l'étude du fonctionnement discursif des dénominations propres. Outre qu'elle fournit des indications chiffrées pour procéder à des sélections entre classes, et à l'intérieur de celles-ci, entre dénominations propres, afin d'orienter la collecte d'autres unités appartenant au vocabulaire des rapports d'activité, elle met en lumière certaines propriétés linguistiques des dénominations propres. La catégorisation manuelle des entités a ainsi révélé l'existence de patrons morphosyntaxiques pour la construction des dénominations propres, que l'on a appelés, à la suite de B. Bosredon, signalétiques. Certains patrons sont utilisés depuis une douzaine d'années dans le repérage et la catégorisation automatique des entités nommées, comme la présence de prénoms, abrégés ou non, devant un nom de famille pour identifier une personne. D'autres patrons sont cependant moins évidents et articulent autrement les composantes d'un référentiel. Il en est ainsi de la construction des noms d'entreprise incluant « Crédit Agricole ». Si les patrons peuvent probablement être généralisés, leurs ingrédients semblent étroitement liés à un contexte et donc une ontologie précis. Il en va de même pour les circulations homonymiques affectant certaines unités du vocabulaire des rapports d'activité, successivement termes désignant un type d'objet ou d'opération, nom de secteur d'activité et nom d'unité dans l'organisation du groupe bancaire. Ces déplacements métonymiques reposent sur un référentiel d'objets particulier dont les rapports d'activité sélectionnent certains éléments. Dans la mesure où l'on souhaite maintenant explorer la manière dont

l'activité bancaire est restituée par ce genre de discours, il devient nécessaire d'approfondir l'étude de l'usage discursif des dénominations propres. On va voir que cet usage est susceptible de varier de façon importante en fonction des classes de dénominations propres et selon que le l'énonciateur tient pour acquis que les unités qu'il utilise appartiennent ou non à un référentiel partagé avec son public destinataire. L'étude de l'usage des toponymes dans le chapitre 9 permettra d'étudier le fonctionnement de dénominations propres dont la connaissance est théoriquement partagée par les interlocuteurs.

CHAPITRE 9

L'ANCRAGE D'UN RÉFÉRENTIEL : LES TOPONYMES

Les toponymes constituent la deuxième classe par les effectifs, derrière les dénominations propres de personnes morales, à être bien représentés dans les rapports d'activité¹. À ce titre, ils forment un ensemble significatif que l'on peut utiliser afin de mettre en lumière une première série d'aspects relatifs au fonctionnement des dénominations propres en discours². Cette connaissance du fonctionnement discursif des dénominations propres est orientée par les besoins de la collecte. Or, on sait que celle-ci est à double détente : collecte des dénominations propres puis, à partir de cette collecte, extraction de certaines unités caractéristiques du vocabulaire du corpus. Dans les deux cas, il s'agit de mettre en évidence l'apport d'une prise en compte des dénominations propres dès lors que l'on souhaite capter les évolutions d'un référentiel spécifique à travers ses manifestations verbales. Deux aspects de la question doivent donc être envisagés. Tout d'abord, quelles informations l'étude des seules dénominations propres en discours permet-elle de capter relativement aux évolutions affectant un référentiel ? En second lieu, il s'agit d'évaluer la productivité de la collecte d'autres unités appartenant au vocabulaire du corpus, lorsqu'elle est réalisée à partir de telle ou telle classe de dénominations propres. Sur ce point, on fait le raisonnement que si toutes les classes de dénominations propres présentent a priori un intérêt pour ce type de collecte, il semble plus judicieux de retenir les classes qui permettront de capter dans ce vocabulaire des éléments exemplaires, voire centraux.

De ce point de vue, les toponymes offrent un positionnement particulier. Ils font très certainement partie des dénominations propres les mieux partagées par les interlocuteurs de la communication établie par les rapports d'activité. Ce référentiel partagé a pour contrepartie de favoriser une prise en charge minimale de la part de

¹ Voir chapitre 8.

² Dans sa thèse, où elle étudie la sémantique et le fonctionnement discursif des noms de pays, G. Cislaru (2005) souligne l'importance qu'il y a à adopter une approche discursive pour un tel sujet. C'est, du reste, la voie déjà proposée par certains, comme J.-L. Vaxelaire (2001) dans une perspective lexicologique. Le présent travail s'inscrit dans un tel courant. Toutefois, on n'a pas jugé nécessaire d'adopter la proposition selon laquelle le nom de pays correspondrait à un concept hybride oscillant entre un pôle /lieu/ et un pôle /pays/ (2005 : 520), pas plus d'ailleurs qu'il n'a semblé nécessaire d'adopter le point de vue de J.-L. Vaxelaire, tendant à assimiler le sens du nom propre à celui du nom commun (dans le cadre de la sémantique interprétative de F. Rastier). Sur cette approche, voir chap. 8 et 10.

l'énonciateur, qui peut tabler dans ce cas sur les connaissances de son destinataire. De fait, les indications déjà réunies dans le précédent chapitre semblent indiquer que les toponymes ne jouent pas un rôle central dans le vocabulaire utilisé pour la restitution de l'activité bancaire. On ne cherchera donc pas dans un premier temps à réaliser une collecte à partir de ces unités, mais plutôt à mettre en lumière les propriétés discursives permettant d'évaluer la manière dont elles participent à la verbalisation de la restitution de l'activité bancaire. Cela présuppose que l'on sache, par exemple, si les toponymes sont tous utilisés de la même manière dans le corpus. Si des usages différents sont constatés, il faudra les préciser au regard de l'intérêt qu'ils peuvent présenter pour une collecte d'unités. En effet, on peut supposer que l'importance des toponymes dans les discours est susceptible de varier en fonction du genre dans lequel ils apparaissent. Il suffit pour cela de comparer la place qu'ils occupent dans les guides touristiques¹, les ouvrages de géographie ou dans certaines oeuvres romanesques, avec celle qui leur revient dans une documentation comptable ou un annuaire international. Il convient donc de commencer par préciser leur usage dans les textes du corpus.

En revanche, et quand bien même les toponymes seraient réduits à ne jouer qu'un rôle de coordonnées spatiales, leur mention fait partie du référentiel qui est verbalisé dans et par les rapports d'activité. À ce titre, étudier leur usage devrait permettre de révéler la prise en compte de la dimension spatiale dans la restitution de l'activité bancaire. Dans un tout autre registre, Étienne Brunet a remarqué que les toponymes, qui forment une sorte de toile de fond dont les éléments sont récurrents d'un roman à l'autre, permettent de construire « *l'espace culturel de l'écrivain* »².

D'autre part, on a vu dans les chapitres précédents, en particulier lorsqu'il s'est agi de la terminologie bancaire, que la présence de certains noms d'entités était liée à l'importance de l'entreprise qu'ils servent à nommer dans le référentiel du groupe bancaire ; on verra plus loin que la fréquence de ces dénominations propres entretient un rapport avec leur implication dans la restitution de l'activité³. Dès lors, n'est-il pas légitime de considérer que, si des écarts de fréquence

¹ L. Mondada, 1994.

² É. Brunet (1985 : 296), dans *Le vocabulaire d'Emile Zola*.

³ Voir chapitre 10.

importants sont constatés entre toponymes, y compris à l'intérieur d'une même sous-catégorie, ce phénomène a un rapport avec le rôle qu'ils jouent dans la restitution de l'activité ? Cette question mérite d'autant plus d'être soulevée que le corpus étudié est une série chronologique dans laquelle les mêmes éléments sont théoriquement repris d'une année sur l'autre. Les variations de fréquence ne peuvent plus alors être attribuées seulement à des changements thématiques : il s'agit d'une évolution du référentiel lui-même. De manière plus générale, on va voir que l'explication de ces variations de fréquence exige que l'on rompe avec une conception utilitaire de l'emploi des toponymes.

Il conviendra donc de s'interroger, à partir des variations observées dans les fréquences et la répartition des toponymes dans le dispositif textuel, sur l'existence et la nature de rôles différents endossés par ces unités. On utilisera à des fins heuristiques les catégories communes dans lesquelles les toponymes sont habituellement classés : odonymes, villes et pays. On s'attachera pour chaque catégorie à observer son fonctionnement dans le corpus et à isoler, à partir des indications de fréquence et de répartition, les rôles particuliers tenus par certains éléments. L'enquête exploitera les possibilités d'exploration textuelle offertes par le programme Lexico (carte des sections où les paragraphes sont distingués et rattachés à des contextes à dominante syntactique ou non syntactique, eux-mêmes regroupés en fonction des principales parties du corpus), ainsi que la ventilation des fréquences des unités sur les différentes parties du corpus. C'est à la confluence de ces deux aspects que l'on recherchera l'existence de rôles particuliers, c'est-à-dire d'usages pragmatiques spéciaux de certaines dénominations propres ou classes de dénominations dans le genre de discours représenté par les textes du corpus.

I. LES TOPONYMES DANS LE CORPUS DE RAPPORTS D'ACTIVITÉ

A. Les principales fréquences

L'examen de la liste des toponymes impose un premier constat : à l'image des types d'unités déjà étudiés, le classement par ordre décroissant de fréquences de ces dénominations propres révèle d'importantes disparités. On a ainsi 150 unités dont la fréquence est de une ou deux occurrences, un autre groupe rassemblant plus de 200 unités dont les fréquences s'échelonnent entre 3 et 17 ; enfin, un troisième groupe peut être constitué avec les trente dénominations toponymiques les plus fréquentes. Les toponymes représentent près de 3700 occurrences dans le corpus, soit 18% des dénominations propres. Le tableau ci-dessous les reprend par ordre décroissant de fréquences.

Tableau 57.
Toponymes classés par fréquence décroissante supérieure à 20

Dénominations	Variante	Frq.
France	FRANCE	542
Europe	EUROPE	236
Paris		120
Asie	ASIE	103
Italie		96
Suisse		80
Etats Unis	USA	68
Espagne		68
Hong Kong		60
Grèce		55
Portugal		54
Londres		46
Monaco	Principauté de Monaco ; Principauté	37
Argentine		36
Allemagne		36
Pologne		34
Singapour		33
New York	NY	33
Afrique	AFRIQUE	30
Moyen Orient		30
Brésil		29
Japon		29
Amérique latine	Amérique Latine	27
Russie	ex-URSS	26
Luxembourg		25
Tokyo		25
Maroc		24
Liban		24
Royaume Uni		21
Chili		20

Les unités rassemblées dans ce tableau peuvent être classées en trois sortes de toponymes : zones géopolitiques (pour les 5 unités correspondant à « Europe », « Asie », « Afrique », « Moyen-Orient » et « Amérique latine »), pays (21 unités) et villes (4 unités correspondant à « Paris », « New York », « Londres » et « Tokyo »). Ces catégories ne sont pas tout à fait satisfaisantes, dans la mesure où l'on note la

présence de « villes-États¹ » comme Hong-Kong, Monaco, Singapour et Luxembourg. Tout en gardant ce classement, on peut tenter d'en préciser les caractéristiques. En effet, les fréquences semblent nettement privilégier certaines zones géographiques : « France » se détache des autres toponymes, ainsi que « Europe ». On peut chercher à vérifier si le poids attaché à ces dénominations correspond à l'importance de ces zones géographiques dans le référentiel véhiculé par le corpus.

B. Zones géographiques

Pour vérifier ce point, on a regroupé les différents toponymes par grandes zones géographiques. En procédant de la sorte on a ménagé une exception pour la France qui possède un profil atypique.

*Tableau 58.
Classement des unités toponymiques en fonction des principales zones géographiques*

Zone géographique	Unités	Occurrences
France	200	1096
Europe	76	1615
Amériques	42	372
Asie	37	382
Afrique	20	103
Moyen-Orient	13	102
Océanie	4	17

La lecture du tableau 58 ci-dessus permet de constater que la France est la mieux représentée en termes d'unités, et que parmi celles-ci la dénomination « France » occupe bien une place spécifique puisqu'elle est responsable à elle seule de la moitié des effectifs du groupe et de 30% de ses occurrences. De ce point de vue, « Europe » n'a pas le même profil, puisque la dénomination ne représente que 20% des effectifs du groupe. En revanche, c'est ce groupe qui l'emporte en termes d'occurrences. En anticipant un peu sur les développements qui suivent, on peut avancer une première explication à cette répartition. Étant le pays d'origine du

¹ Hong-Kong est une région administrative spéciale de la République populaire de Chine, Monaco et Singapour sont des cités-États, Luxembourg est un état.

groupe bancaire, la France y tient un rôle si considérable qu'elle fonctionne comme toile de fond dans le discours de l'émetteur qui n'a pas toujours jugé nécessaire de la mentionner. On voit, par exemple, que pour les coordonnées des entités du groupe localisées en France, le nom du pays ne figure pas, alors que c'est généralement le cas pour les entités localisées à l'étranger. Les toponymes se rattachant à l'Europe sont donc pour cette première raison plus systématiquement mentionnés. Mais une seconde raison peut expliquer leur diversité : ils font partie de la principale zone de développement à l'étranger de la banque française.

C. Types de toponymes

Afin de compléter ce tableau des toponymes dans les rapports d'activité, on peut chercher à vérifier la place tenue par chaque sous catégorie et vérifier ainsi si les noms de pays forment la principale catégorie présente dans le corpus. Lorsque les quatre cents dénominations toponymiques sont organisées à l'aide des catégories usuelles en ce domaine, on obtient une image plus complète des grandes tendances caractéristiques l'ensemble. En effet, on peut constater, d'une part, que toutes les catégories de toponymes ne sont pas présentes dans les mêmes proportions et, d'autre part, que certaines d'entre elles, absentes du tableau précédent, doivent pourtant être prises en compte.

*Tableau 59.
Classement des toponymes et apparentés présents dans le corpus*

Toponymes et apparentés	Unités
Villes	132
Voies	100
Pays	93
Zones géopolitiques	26
Régions (administratives et géographiques)	24
Lieux-dits	8
Immeubles	5
Ensembles économiques	4
Lieux publics	3
Départements	2

Les villes sont les mieux représentées dans le groupe avec 132 unités, suivies des odonymes (une centaine) et des pays avec 93 unités. On peut constater que la

liste des types de toponymes est plus riche que le laissait penser la liste des trente premières fréquences. Si l'on retrouve trois catégories déjà identifiées, la présence d'odonymes mérite d'être éclaircie, ce qui conduira à ce que l'on interroge les relations tissées dans les rapports d'activité entre toponymes et référentiel du groupe bancaire.

II. ODOMYMES : DE LA FONCTION DE « COORDONNÉES » À L'EFFET « INTERNATIONAL »

A. La fonction « coordonnées »

Les odonymes possèdent en général un classificateur intégré à leur dénomination ; cette catégorisation propre à la signalétique des noms de voies est réutilisée telle quelle dans le corpus des rapports d'activité, lorsqu'elle sert à communiquer les coordonnées de certaines entités appartenant au groupe bancaire. Ainsi qu'on va pouvoir le constater, la mention de ces coordonnées ne peut pas être limitée aux seuls aspects « pratiques ».

Parmi la centaine d'odonymes répertoriés dans le corpus, les plus fréquents présentent moins de 10 occurrences. On donne ci-dessous les premiers éléments de la liste classés par ordre de fréquence décroissante.

Tableau 60.
Extrait de la liste des odonymes répertoriés dans le corpus « rapports d'activité »

Odonymes	Frq.
avenue du Maréchal Foch	9
avenue Foch	9
avenue André Malraux	8
avenue de Kéranguen	8
avenue de la Libération	8
avenue Elisée Cusenier	8
avenue François Verdier	8
avenue Napoléon III	8
boulevard du Président Wilson	8
boulevard de Vanteaux	8
boulevard Guillet Maillet	8
chemin du Mas de Cheylon	8
place de la Gare	8
place Jeanne d Arc	8
route du Loch	8
rue Bergson	8
rue Case Nègre	8
rue Libergier	8
rue Paul Claudel	8
rue Pierre Bretonneau	8

Le tableau ci-dessus met en évidence un palier à la fréquence 8 qui se poursuit au-delà des seuls cas recensés ici à titre d'exemple. À l'intérieur d'une catégorie aussi délimitée que les odonymes, ce pallier fonctionne comme l'indice d'une utilisation homogène et récurrente sur plusieurs parties du corpus (qui en compte neuf).

On précisera l'indice fourni par les fréquences en observant la manière dont les odonymes se répartissent entre types de contextes, puisqu'on a vu dans la première partie que des sortes d'informations différentes y sont regroupées¹. Pour conduire cette exploration, on regroupe les toponymes désignant des voies² et on

¹ Voir les chapitres 2 et 3.

² Dans la mesure où cette liste comporte un nombre important de faibles fréquences, elle a été ramenée aux formes simples entrant dans les dénominations et comportant une majuscule initiale ou appartenant à un groupe restreint de mots désignant des sortes de voies et utilisés dans les dénominations complexes (allée, avenue, axe, boulevard, bd, chemin, faubourg, place, quai, route, rue). Par ailleurs, on a retiré de la liste des unités utilisées dans les dénominations d'odonymes mais appartenant également à d'autres sortes de dénominations ; dans la mesure où les fréquences de ces unités étaient nettement supérieures à celles observées pour les odonymes en moyenne (4,3 occurrences, avec un maximum de 9), les unités ont été retirées de la liste (il s'agit de

projette cet ensemble sur le graphique de ventilation des occurrences du programme Lexico.

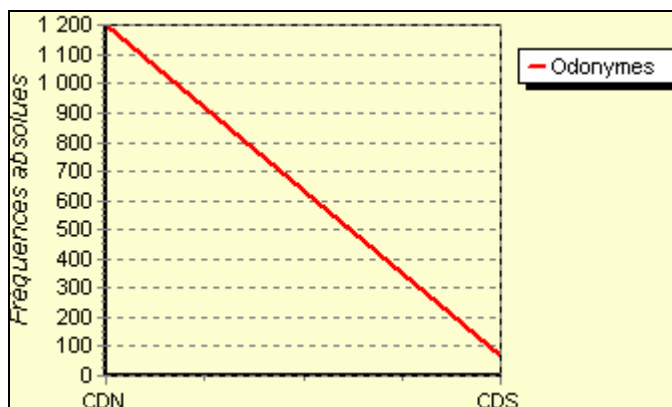


Figure 63.

Ventilations des odonymes entre contextes syntactiques et non syntactiques

Ce graphique met en évidence que les odonymes appartiennent de façon presque exclusive aux contextes à dominante non syntactique (le graphique utilise des fréquences absolues qui donnent le décompte exact des fréquences observées). Si cela ne fournit pas encore la clé de leur utilisation dominante, on peut néanmoins en tirer une première conclusion : ces unités ont un nombre réduit de rôles dans les rapports d'activité. Ces rôles peuvent être appréhendés globalement en recherchant les co-occurents les plus fréquents des odonymes. Pour cela, le groupe de formes est projeté sur la carte des sections du programme Lexico. Le calcul des spécificités présente ensuite les mots qui apparaissent fréquemment dans les paragraphes où figure au moins une des unités du groupe « odonymes ».

« Europe »(236), « François » (73), « Général »(485), « Louis »(37), « New » (33), « Paul » (86), « Pierre » (169), « Président » (724), « York » (33), et « place » (234 occurrences utilisées en général dans des locutions du type « première place »).

Tableau 61.
Liste des vingt premiers co-occurents des odonymes dans le corpus des rapports d'activité

Forme	Frq.Tot.	Fréquence	Coeff¹.
04	101	87	***
Total	688	617	***
MF	663	616	***
fonds	567	312	***
01	99	78	***
Bureaux	114	106	***
bureaux	223	204	***
02	112	91	***
Président	724	502	***
Directeur	724	460	***
avenue	79	79	***
Paul	86	73	***
Tél	449	417	***
05	86	80	***
bilan	394	308	***
Jean	399	184	***
locales	484	312	***
Général	485	320	***
03	91	82	***
propres	453	311	***

Les co-occurents les plus fréquents des odonymes appartiennent majoritairement au dispositif destiné à présenter les coordonnées des entités du groupe bancaire, et plus particulièrement les Caisses régionales de Crédit agricole. En effet, on reconnaît dans le tableau les indicatifs téléphoniques, les fonctions de président et de directeur général, la mention du nombre de bureaux et de caisses locales, ainsi que les indications chiffrées relatives aux fonds propres et aux bilans des banques régionales affiliées à l'organe central du Crédit agricole². Les fiches signalétiques des caisses régionales, qui sont des entités nombreuses (une cinquantaine), et régulièrement présentées dans les rapports, fournissent, à l'exception de l'année 2002, le plus grand nombre de formes co-occurentes pour les odonymes. L'interruption de l'année 2002 explique qu'un nombre important

¹ Les astérisques indiquent un coefficient positif supérieur à 50, ce qui signifie que tous les co-occurents regroupés dans le tableau apparaissent très fréquemment dans les mêmes contextes (paragraphes) que les odonymes.

² Un exemple de ce dispositif est donné dans le chapitre 2.

d'odonymes fréquents soient mentionnés huit fois et non neuf comme on aurait pu s'y attendre : c'est qu'ils sont insérés dans un dispositif textuel qui, à un moment donné, a pu être jugé facultatif par le destinataire des rapports d'activité.

Les observations réalisées dans la première partie au sujet de l'évolution du vocabulaire des rapports d'activité peuvent fournir une explication à cette éclipse de certains odonymes associés à la présentation des caisses régionales de Crédit Agricole. En effet, on a pu constater que le changement de statut de l'émetteur des rapports d'activité, intervenu en 2001, avait eu de multiples répercussions sur le traitement de la restitution de l'activité du groupe bancaire. Tout se passe alors comme si les rapports d'activité, devenant un outil de communication principalement au service de la partie cotée du groupe Crédit agricole, n'étaient plus liés par la tradition, remontant à 1946, de mentionner certains renseignements pratiques relatifs à la partie non cotée du Crédit Agricole¹. La réapparition des coordonnées des Caisses régionales dans le rapport de l'année suivante suggère que l'enjeu de leur présence dépasse les seuls aspects pratiques. Ces derniers semblent surdéterminés par des enjeux communicationnels relatifs à ce qui est restitué du référentiel commun : l'enracinement régional de la banque française ne pouvait pas être escamoté de cette manière.

On a ainsi pu cerner l'usage dominant des odonymes dans les rapports d'activité et avoir au passage confirmation des indications fournies par le palier de fréquence identifié dans le tableau 61. Il reste néanmoins deux points à approfondir. Le premier concerne la présence d'odonymes dans les contextes à dominante syntactique : s'agit-il d'un usage différent de celui qui vient d'être constaté ? En second lieu, la liste des odonymes² (voir annexes) fournit un certain nombre d'unités qui, du fait de leur consonance étrangère, ne peuvent pas appartenir aux coordonnées des Caisses régionales de Crédit agricole, implantées uniquement sur le territoire français.

¹ Il faut faire une exception pour la vingtaine de caisses régionales cotées en bourse et qui disposent de leurs propres moyens de communication financière.

² Voir annexes.

B. Formes partagées et emprunt de signalétique

Le graphique donné en début de section indique que certaines formes correspondant à des parties d'odonymes¹ apparaissent en dehors des contextes non syntactiques. Dans la mesure où elle présente les paragraphes du corpus regroupés par types de contextes, la carte des sections du programme Lexico permet de repérer rapidement les contextes dans lesquels apparaissent ces exceptions. On donne ci-dessous un exemple de la mise en évidence des contextes hors CDN.

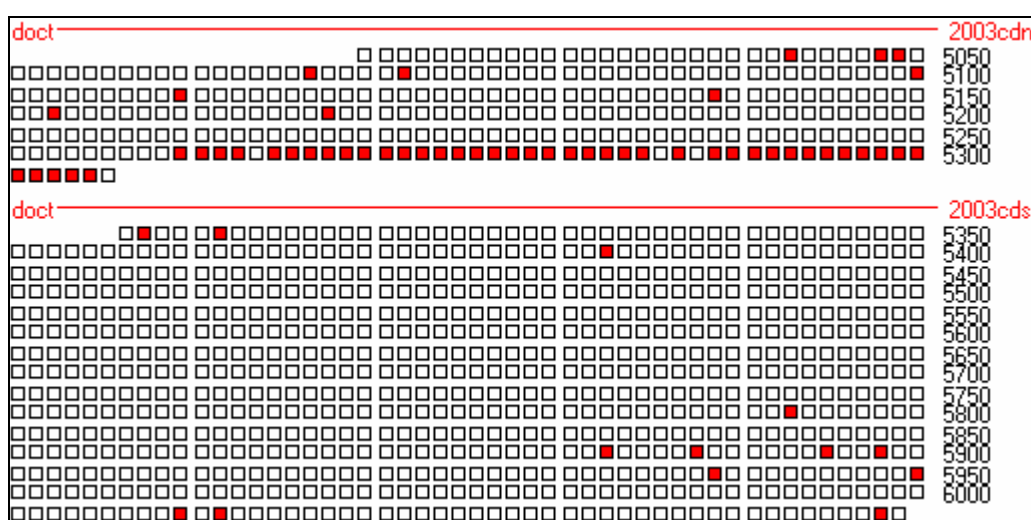


Figure 64.

Sections (paragraphes) des CDN et CDS de l'année 2003 dans lesquelles les carrés colorés en rouge indiquent la présence de formes appartenant à des odonymes

Les résultats fournis par l'examen des treize sections appartenant à des contextes syntactiques de la partie 2003, et dans lesquelles des composants d'odonymes apparaissent, sont à l'image de ceux qui caractérisent les autres parties du corpus. On rencontre, en effet, des formes que les odonymes partagent avec des syntagmes divers ou des dénominations. C'est ainsi que la forme « Trois » apparaissant dans la dénomination « chemin des Trois Cyprès » se rencontre également en début de phrase dans les contextes syntactiques. Pour les seuls

¹ On a indiqué précédemment que la liste des odonymes avait été réduite à un ensemble de formes simples entrant dans leur composition ; on a donné dans la section précédente les critères d'élimination pour certaines formes présentes dans les dénominations. On peut voir dans cette section que le tri n'a pas été exhaustif.

contextes CDS de 2003 on a ainsi « Trois facteurs... », « Trois Caisses régionales... », « Trois augmentations de capital... ». Les formes « Foch », « Via » et « axe » présentent des cas de figure similaires. Mais le partage peut également avoir lieu entre dénominations ; la « rue Saint Fuscien » prête son « saint » à « Saint-Gobain », grande entreprise ; de même, le « boulevard Pasteur » à Paris partage-t-il « Pasteur » avec l' « Institut Pasteur ». Ce partage de formes, qui a pour arrière-plan un référentiel français étendu dans lequel s'insère l'activité du groupe bancaire, peut aller dans certains cas jusqu'à un emprunt de signalétique.

On trouve en effet dans la partie 2003 l'exemple de « SAS Rue La Boétie » (« *Constituée sous la forme d'une Société par Actions Simplifiée, la SAS Rue La Boétie est un holding de contrôle regroupant les participations des Caisses Régionales au capital de Crédit Agricole S.A. au 31 décembre 2001*¹. »). On trouve un autre exemple dans le corpus avec « Rue Impériale de Lyon », holding de contrôle du groupe Lazard. Il est significatif que dans les deux cas il s'agisse d'une holding, c'est-à-dire d'une société qui possède rarement une identité propre, puisque son rôle est principalement d'assurer un contrôle capitalistique sur des sociétés qui peuvent être très différentes. Il ne s'agit donc pas d'un nom qui doit nécessairement fonctionner comme marque, ce qui explique la liberté prise avec la signalétique des noms de sociétés (il en sera question dans le chapitre suivant). En revanche, l'enracinement local à travers un nom de rue est de nature à faire sens pour les initiés. En effet, « SAS Rue La Boétie » reprend le nom de la rue parisienne où sont situés les locaux de la Fédération Nationale de Crédit Agricole, instance représentative des Caisses régionales de Crédit agricole, par ailleurs actionnaire principal de Crédit Agricole s.a. De même, la dénomination « Rue Impériale de Lyon » reprend-elle l'ancien nom d'une rue lyonnaise². Lorsque l'on compare les profils statistiques de ces deux entités avec celui de l'une des rues mentionnées dans le tableau 60, le décalage est immédiatement perceptible.

¹ Rapport d'activité 2001.

² La « rue Impériale » a été ouverte sous le Second Empire en 1854 ; elle est devenue « rue de Lyon » en 1871, puis « rue de la République » en 1878 (*Rues de Lyon à travers les siècles*, M. Vanario, Lyon, Éditions lyonnaises d'art et d'histoire, 2002, 333 p.).

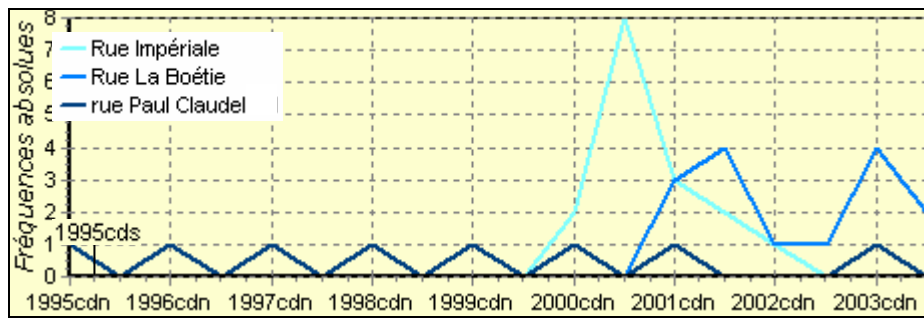


Figure 65.

Ventilation des fréquences de « Rue Impériale », « Rue la Boétie » et « rue Paul Claudel »

Lecture de la figure 65 : les fréquences des trois « odonymes » sont ventilées sur l'ensemble des parties du corpus, chacune de ces parties étant décomposée en contextes syntactiques (CDS) et non syntactiques (CDN). Par exemple pour l'année 1995, il y a une occurrence de « rue Paul Claudel » dans les contextes à dominante non syntactique (1995 CDN) et aucune dans les contextes à dominante syntactique.

Le graphique met en évidence le comportement atypique des deux faux odonymes « Rue Impériale de Lyon » et « Rue La Boétie ». En effet, on constate, d'une part, que ces derniers possèdent des fréquences nettement supérieures à celles enregistrées pour les autres odonymes, et d'autre part, qu'ils sont présents également dans les contextes syntactiques, ce qui suggère une utilisation différente de celle de simple repère spatial. Ce détournement de signalétique n'est cependant pas la seule manifestation des traces que la pragmatique des odonymes a laissées dans le corpus. En effet, parmi les nombreuses références culturelles présentes dans les odonymes et qui enracinent le référentiel du Crédit agricole dans le territoire français, la présence de noms de voies étrangères étonne et mérite un examen complémentaire : s'agit-il d'un indice de l'internationalisation du groupe bancaire ?

C. Odonymes au service d'un « effet international »

Parmi la centaine de dénominations de voies recensées, des toponymes étrangers sont manifestement présents : « California Street », « Platz der Republik », « Paseo de la Habana », etc. On peut se demander si ces dénominations fonctionnent comme un indice de l'internationalisation du groupe bancaire. En réalité, un sondage réalisé dans la carte des sections révèle que ces occurrences sont concentrées en

début de corpus et principalement dans l'année 1995, comme le montre la ventilation ci-dessous.

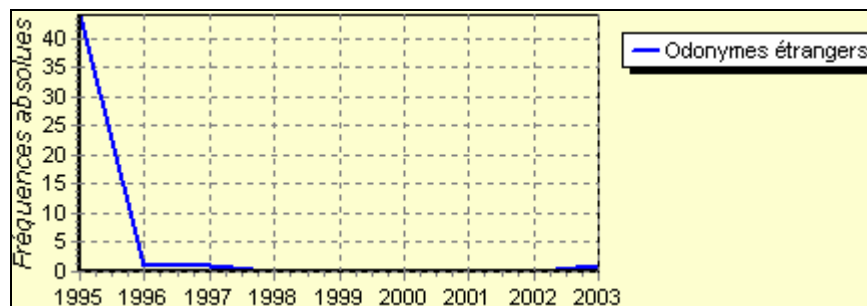


Figure 66.

Ventilation des toponymes étrangers sur les neuf parties du corpus RA9503

Le fait que ces toponymes étrangers soient concentrés dans la seule partie 1995 (la seule occurrence pour 2003 est due à l'homographie de « Via », toponyme italien dans « Via San Pietro all'Orto » en 1995, et préposition placée en début de phrase en 2003 dans « via ce partenariat (...) »), doit être rapproché d'un fait extérieur, l'acquisition de la Banque Indosuez l'année suivante. En effet, il semble qu'avant cette acquisition, l'émetteur des rapports d'activité ait assigné aux coordonnées de succursales ou de participations à l'étranger la fonction d'étoffer, et de rendre en quelque sorte plus tangible la dimension internationale du groupe bancaire. Une fois l'acquisition réalisée, cet artifice n'a plus paru justifié, tant la dimension internationale était devenue incontestable avec le rachat du réseau de la Banque Indosuez.

L'étude des toponymes présents dans le corpus a permis de montrer que leur rôle principal consiste à fournir des repères dans certains dispositifs textuels livrant les coordonnées d'entités du groupe bancaire Crédit Agricole. Ce rôle est associé à une norme mesurable : fréquences inférieures à 10, présence majoritaire dans les contextes à dominante non syntactique. Le profil atypique de deux noms d'entreprises ayant emprunté leur forme à la signalétique des toponymes vient étayer cette observation : malgré la ressemblance formelle, la différence de nature de l'entité dénommée produit un profil textométrique différent. Par ailleurs, l'utilisation des toponymes comme simples repères mentionnés dans les textes n'hypothèque pas l'existence d'un autre rôle intervenant sur un plan différent. En effet, l'apparition ou

la disparition de certains odonymes ne semble pas pouvoir s'expliquer simplement par des raisons de commodités pratiques. La disparition d'odonymes étrangers à partir de 1996, ou l'éclipse de certaines dénominations françaises appartenant à la même classe peuvent être mises en relation avec des perturbations du référentiel du groupe bancaire, comme l'acquisition d'une véritable dimension internationale ou la cotation de son organe central. On voit ainsi que les fréquences des odonymes doivent aussi être interprétées à la lumière de l'usage communicationnel qui est fait du dispositif textuel dans lequel ces unités apparaissent. Avec les noms de villes, c'est un autre aspect de l'usage d'une unité qui devra être pris en compte : ses rôles peuvent varier en fonction du type de configurations discursives dans lesquelles elle apparaît.

III. NOMS DE VILLES : D'UNE GÉOGRAPHIE DE L'ACTUALITÉ SECONDAIRE AUX PLACES FINANCIÈRES INTERNATIONALES

A. Le rôle de repère et les dispositifs de « coordonnées » dans les CDN

Avec les noms de villes, autre catégorie de toponymes bien représentée dans les rapports d'activité, on peut étudier les rôles différents qu'une même unité peut tenir dans un corpus rassemblant des textes appartenant au même genre de discours. Cette notion de rôle est importante, car elle permet de regrouper des aspects différents du fonctionnement discursif d'une dénomination propre : explicitation de son contenu par le discours, relations établies en général avec les autres unités linguistiques et, en définitive, type de réseau pouvant être construit dans le cadre d'un référentiel terminologique. Par exemple, on a pu voir dans la section précédente que les odonymes utilisés dans un dispositif de type « coordonnées » ne font l'objet d'aucune explicitation et que les relations de voisinage entretenues avec d'autres unités sont déterminées par le contenu habituel du dispositif dans lequel ils se trouvent insérés.

Pour rendre compte de l'utilisation des noms de villes dans les rapports d'activité, on peut partir des unités les plus fréquentes, car elles sont susceptibles de se rencontrer dans des contextes multiples, où différents rôles ont pu leur être

attribués. On partira donc ici des vingt noms de villes dont les fréquences sont les plus élevées.

*Tableau 62.
Vingt noms de villes les plus fréquents attestés dans le corpus RA9503*

Noms de villes	Frq.
Paris	120
Londres	46
New York	33
Tokyo	25
Chicago	17
Toulouse	13
Bordeaux	12
Genève	11
Madrid	11
Poitiers	11
Francfort	10
Milan	10
Lyon	10
Nantes	10
Rennes	10
Abymes	9
Chartres	9
Grenoble	9
Lille	9
Meaux	9
Quimper	9

Sur les 132 noms de villes recensés dans le corpus, « Paris » possède une fréquence deux fois et demie supérieure au deuxième nom (Londres, avec 46 occurrences). Le groupe composé de « Londres », « New York », « Tokyo » et « Chicago » se distingue par des fréquences nettement supérieures à 10. Enfin, un troisième groupe rassemble des villes françaises et étrangères dont la fréquence oscille entre 13 et 9. Pour contrôler la distribution des fréquences entre sortes de contextes, on peut constituer quatre ensembles : « Paris », les villes étrangères dont la fréquence est supérieure à 15, ou groupe 2 (« Londres », « New York », « Tokyo » et « Chicago »), les villes françaises dont la fréquence est inférieure à 15, ou groupe 3 (« Toulouse », « Bordeaux », « Poitiers », « Lyon », « Nantes », « Rennes », « Abymes », « Chartres », « Grenoble », « Lille », « Meaux », « Quimper »), et les villes étrangères dont la fréquence est inférieure à 15, ou groupe 4 (« Genève », « Madrid », « Francfort », « Milan »). On donne ci-dessous en

fréquences absolues la répartition des occurrences des quatre groupes entre les deux sortes de contextes, à dominante syntactique et non syntactique.

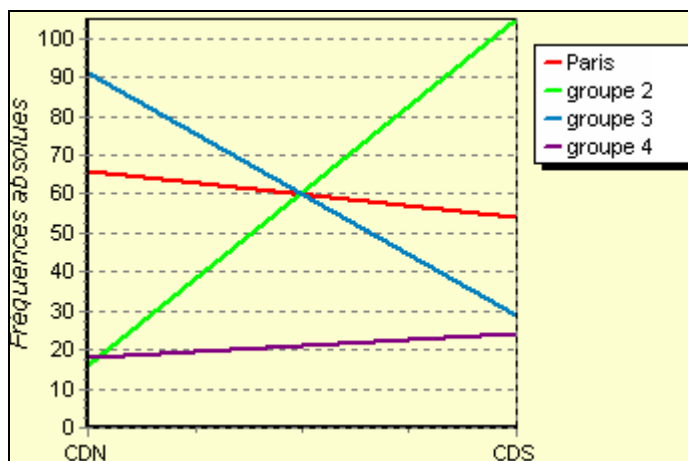


Figure 67.

Fréquences absolues de quatre groupes de noms de villes dans les deux sortes de contextes CDN et CDS

La ventilation des quatre groupes de formes montre des situations assez contrastées. Trois cas de figure sont à distinguer : le groupe 3 est très majoritaire dans les CDN, ce qui l’oppose au groupe 2 qui, dans des proportions voisines, est majoritairement représenté dans les CDS. Enfin, Paris et le groupe 4 (noms de villes étrangères dont la fréquence est inférieure 15) ont une répartition de leurs fréquences assez équilibrée dans les deux sortes de contextes.

Malgré les différences assez marquées de ces profils statistiques, l’examen des contextes de type CDN à l’aide de la carte des sections indique plus de points communs entre les quatre groupes que l’on aurait pu l’imaginer au premier abord. En effet, tous les noms de villes sont mis à contribution dans des dispositifs de coordonnées, ce qui rapproche leurs caractéristiques de celles des odonymes. On trouve, comme pour les odonymes, des échanges de formes entre catégories de dénominations : « Paris » est inclus dans les odonymes « Route de Paris » ou « Avenue de Paris » ; on le trouve également dans les dénominations de certaines sociétés, comme « Crédit Agricole de Paris et de l’Ile de France », de même que « Madrid », réutilisé dans la dénomination « Indocam Madrid ». Enfin, l’année 1995 confirme son caractère atypique au regard des coordonnées étrangères, puisqu’elle est la seule année où des noms de villes appartenant aux groupes 2 et 4 figurent dans

les contextes à dominante non syntactique. On va voir que l'étude des contextes syntactiques révèle des situations différentes.

B. À villes moyennes, actualité de second ordre

Présentes dans les deux sortes de contextes avec des proportions similaires, respectivement 19 et 22 sections, les formes des groupes 3 et 4 fournissent un cadre pour la présentation d'une actualité de second ordre. Il s'agit, par exemple, des villes françaises pilotes pour le porte-monnaie électronique Monéo, ou d'infrastructures de collectivités locales financées par une filiale du Crédit agricole, comme les incinérateurs de déchets. La participation de la banque à certaines manifestations d'élus locaux, ainsi que la présentation d'ouvrages ou de monuments dont la restauration a été partiellement prise en charge par la Fondation du Crédit agricole « Pays de France », sont l'occasion de mettre en scène un enracinement local.

Ce qui caractérise la mention des quatre villes européennes appartenant au groupe 4, c'est le signalement de l'ouverture d'implantations modestes, comme des succursales, des bureaux de représentation ou des agences. Ces têtes de pont du groupe bancaire délimitent, en même temps qu'elles consolident, un développement de l'activité à l'étranger centré sur l'Europe de l'Ouest. Il est frappant de voir que dans les deux cas, villes françaises et villes européennes, il y a une certaine correspondance avec l'importance accordée aux faits d'actualité rapportés. Pour présenter ces faits secondaires, le locuteur semble avoir « naturellement » adopté un cadre plus local que national. On note que cette forme de traitement de l'actualité étrangère cesse pour les années 2002 et 2003, peut-être à cause de son caractère trop anecdotique ; en revanche, lorsque des thématiques nouvelles d'importance nationale comportent une dimension locale, les villes françaises maintiennent leur présence, que ce soit comme cadres dans lesquels ont eu lieu des réunions d'information à destination des actionnaires, ou comme théâtres d'expériences relevant du développement durable. Il reste maintenant à examiner les propriétés en contextes syntactiques des villes appartenant aux deux autres groupes.

C. Places financières

Les quatre noms de villes étrangères du groupe 2 (Londres, New York, Tokyo, Chicago) présentent de nombreux points communs avec Paris. En ce qui concerne les contextes non syntactiques, la présence de ces noms de villes est directement liée aux coordonnées des entités du groupe bancaire. Pour les quatre villes étrangères, on retrouve la configuration décrite à propos des odonymes. L'année 1995 a un statut particulier sur lequel on ne revient pas, si ce n'est pour préciser que les quatre villes sont absentes des CDN pour toutes les autres parties du corpus. Ce n'est pas le cas de Paris qui est présent dans ces contextes sur l'ensemble de la période couverte par le corpus. Les pics de fréquence constatés pour les trois années consécutives 1999, 2000 et 2001 sont directement liés à l'apparition des coordonnées complètes des filiales du groupe bancaire. Ces filiales spécialisées ayant majoritairement leur siège à Paris, cela a entraîné une augmentation mécanique de la présence de la forme. Pour illustrer le fonctionnement de ces formes dans les contextes syntactiques, on s'appuiera sur le cas exemplaire de Paris. Afin de conduire cette exploration dans l'ensemble du corpus, on procède à un calcul des co-occurents que l'on lance pour chaque sorte de contexte dans lequel le toponyme apparaît.

Tableaux 63 et 64.

Principaux co-occurents de la forme « Paris » dans les contextes CDN (tableau 63) et CDS (tableau 64)

Forme	Frq.Tot.	Fréquence	Coeff.	Forme	Frq.Tot.	Fréquence	Coeff.
01	99	41	***	Paris	120	54	***
Paris	120	65	***	Londres	46	22	26
Tél	449	58	43	York	33	13	15
Président	724	68	41	New	33	13	15
Directeur	724	56	30	Tokyo	25	10	12
Général	485	46	28	Singapour	33	10	11
quai	17	14	26	Bourse	32	10	11
61	49	18	23	FP	27	9	10
Cedex	20	13	22	CAL	27	9	10
23	115	22	21	Centre	41	9	8
43	84	19	20	AIPG	7	5	8
Téléphone	34	13	18	Fitch	12	5	7
75015	8	8	17	rang	55	9	7
Fax	38	13	17	places	30	7	7
44000	8	8	17	Chicago	17	6	7
55	87	16	16	Lazard	55	8	6
Paul	86	16	16	Place	14	5	6
75012	7	7	15	notation	14	5	6
02	112	17	15	conférences	3	3	6
Nantes	10	8	15	installée	4	3	6

Si l'on compare les deux tableaux, l'opposition entre les sortes d'informations réunies dans chaque contexte apparaît nettement. Le tableau 63, qui présente les vingt co-occurents les plus fréquents de « Paris » dans les contextes non syntactiques, rassemble des formes dont certaines sont aussi des co-occurents du groupe des odonymes. Il s'agit dans les deux cas de formes apparaissant dans des dispositifs de coordonnées : indicatifs téléphoniques, fax, codes postaux, titres des responsables des entités, et membres de dénominations d'odonymes comme « Route de Paris ». Paris est donc répertoriée comme ville de siège social. La lecture du tableau 64 ne permet pas d'arriver à la même conclusion.

En effet, dans les contextes de type CDS, la forme « Paris » entre dans le voisinage d'autres villes dont celles qui appartiennent au groupe 2 – Londres, Tokyo, New York, Chicago –, mais également la « ville-État » de Singapour. Le point commun qui rassemble ces noms de villes est fourni par la liste des co-occurents elle-même. Il s'agit dans tous les cas de places financières

internationales¹. Ces villes apparaissent dans des contextes de ce type : « *Présentes sur les principales places financières aux Etats-Unis (New York et Chicago), en Europe (Paris, Londres, Madrid) et en Asie (Singapour, Tokyo), les équipes de Carr Futures offrent un service global, tant sur les futures financiers que sur les matières premières*². » Il faut également noter que le terme de « place » sert à construire certaines dénominations complexes « Place de Paris » (3 occ.) et « Place de cotation Euronext Paris » (2 occ.). Le seul mot « Place » est également utilisé pour désigner par catachrèse la « Place de Paris ».

Le mot « place », qui est aussi un terme banalisé ou en cours de banalisation³, désigne une ville « *où se font des opérations de banque, de commerce; ensemble des banquiers, des commerçants, des négociants qui exercent leur activité dans une ville*⁴. » La phraséologie des rapports d'activité propose ainsi un certain nombre de syntagmes : « places financières », « places financières internationales », « places financières mondiales », « places internationales », « places boursières », « places principales à vocation internationale ». Ces places, où s'effectuent un grand nombre d'échanges financiers, sont également déclinées par zones géographiques. On a alors « places asiatiques » ou « places européennes ». Dans la mesure où ces expressions sont figées, elles méritent d'être prises en compte comme formes de catégorisation des référents de certains noms de villes. L'extrait du rapport de 1997 cité précédemment permet d'éclairer un autre point sur lequel on reviendra plus en détail dans le chapitre suivant. En effet, on remarque que la catégorisation intervient lorsqu'un aspect particulier du référent « ville » est sélectionné. Ce cas

¹ D'autres formes de la liste suggèrent également cette interprétation. C'est le cas de « FP », « CAL », « AIPG », « Lazard » et « Fitch » qui correspondent à des dénominations propres de sociétés ou aux membres de certaines dénominations. À partir de ces éléments, il est possible de reconstituer les dénominations complètes : « CA AIPG » (Crédit Agricole Alternative Investment Products Group) et « CAL FP » (Crédit Agricole Lazard Financial Products). « Fitch » correspondant au nom d'une agence de notation. L'activité de ces sociétés est principalement tournée vers la gestion d'actifs financiers. Ce point sera abordé plus loin, lorsqu'il sera question des dénominations de personnes morales.

² Rapport d'activité 1997, p. 31.

³ Voir chapitre 5.

⁴ *Nouveau Petit Robert électronique*, 2001. Dans le *Lexique de la banque et des marchés financiers*, le terme est défini ainsi : « *Le terme peut signifier la localité où s'effectuent des opérations sur instruments financiers (place boursière) ou celui où s'effectuent des opérations de banque (place bancaire).* » Dans le corpus de rapports d'activité, « place » est utilisé pour « place boursière ».

particulier conforte la norme d'utilisation des noms de villes dans les rapports d'activité, c'est-à-dire une forme de repère géographique censément partagé par tous, et qui ne nécessite donc pas d'être catégorisé.

Dans les rapports d'activité, certains noms de villes sont amenés à jouer trois rôles : ils servent d'abord à désigner des repères géographiques intégrés aux dispositifs de coordonnées présents dans les contextes à dominante non syntactique. Ils font également office de cadres géographiques permettant d'ancrer la relation de faits ou d'événements de second ordre dans les contextes à dominante syntactique. Enfin, dans ces mêmes contextes, ils sont susceptibles de jouer le rôle de dénominations propres de places financières internationales, évoquant par métonymie un aspect caractéristique de certaines villes. Ce dernier usage est associé à la présence de classificateurs, ceux-ci pouvant être intégrés, comme pour les noms de voies, à la dénomination propre elle-même. Comme pour les odonymes, on a pu constater que la fréquence et la ventilation des occurrences constituent de bons indicateurs de l'existence éventuelle d'usages diversifiés pour une même unité. Par ailleurs, on voit comment il est possible d'accéder par le biais des noms de villes à la partie du référentiel du groupe bancaire relatif à sa « géographie financière ». L'étude des noms de pays va permettre d'aborder sous un autre angle la géographie dans laquelle se déploie l'activité du groupe bancaire.

IV. PAYS : LA « DIMENSION INTERNATIONALE » ET LA STRUCTURATION GÉOGRAPHIQUE DE L'ACTIVITÉ

A. La dimension internationale du groupe bancaire

On a vu dans une précédente section que les noms de pays avaient supplanté les odonymes pour la localisation des entités du groupe bancaire, la France ayant néanmoins un statut particulier sur lequel on reviendra. S'agit-il de la seule utilisation des noms de pays dans le corpus des rapports d'activité ? Pour répondre à cette interrogation et afin de poursuivre les investigations des sections précédentes sur l'utilisation des toponymes dans le corpus étudié, on partira de l'examen de la liste des principales fréquences.

*Tableau 65.
Noms de pays les plus fréquents dans le corpus de rapports d'activité*

Noms de pays	Variantes	Fréquences
France	FRANCE	542
Italie		96
Suisse		80
Etats Unis	USA	68
Espagne		68
Hong Kong		60
Grèce		55
Portugal		54
Monaco	Principauté de Monaco	37
Argentine		36
Allemagne		36
Pologne		34
Singapour		33
Brésil		29
Japon		29
Russie	ex URSS	26
Luxembourg		25
Maroc		24
Liban		24
Royaume Uni		21
Chili		20

Ce tableau met en évidence le statut particulier des graphies de « France » dans le corpus, ce toponyme possédant une fréquence cinq fois plus élevée que celle du second toponyme le plus fréquent (« Italie », avec 96 occurrences). Ensuite, on voit que derrière la France, la zone géographique la mieux représentée en termes de noms de pays et d'occurrences est l'Europe (Italie, Espagne, Grèce, Portugal, Allemagne, Pologne, Luxembourg, Royaume Uni, pays auxquels il est possible d'ajouter la Suisse et Monaco si l'on s'en tient à une délimitation géographique), avec 389 ou 506 occurrences, suivie par les Amériques (Etats-Unis, Argentine, Brésil, Chili), avec 153 occurrences, et l'Asie (Hong Kong, Singapour, Japon), avec 122 occurrences. Dans cette liste, le Maroc et le Liban sont les seuls représentants des ensembles formés par l'Afrique et le Moyen-Orient. Du fait de sa position géographique centrale, la Russie peut être rattachée à l'un ou l'autre des trois ensembles, sans que cela modifie le classement général.

Si les noms de pays les plus fréquents suggèrent une répartition géographique inégale, ils indiquent également que l'activité du groupe bancaire possède une dimension internationale. Acquis pour une large part avec l'achat de la

banque Indosuez en 1996, cette dimension nouvelle a rapidement été mise en valeur dans le dispositif communicationnel des rapports d'activité. En 1996, la couverture des rapports annuels montre en filigrane une carte du monde et, dès 1997, une carte sur deux pages intitulée « Le Crédit agricole dans le monde » présente la nouvelle géographie des implantations du groupe bancaire :



Figure 68.
Reproduction de la carte des implantations du Crédit agricole donnée dans le rapport d'activité de 1997

Dans la mise en valeur de la dimension internationale de l'activité de la banque, les noms de pays prennent le pas sur les autres toponymes que l'on a déjà rencontrés. Ce phénomène peut être mis en évidence par la projection sur le graphique de répartition de Lexico de trois groupes de formes rassemblant odonymes, noms de villes et noms de pays étrangers.

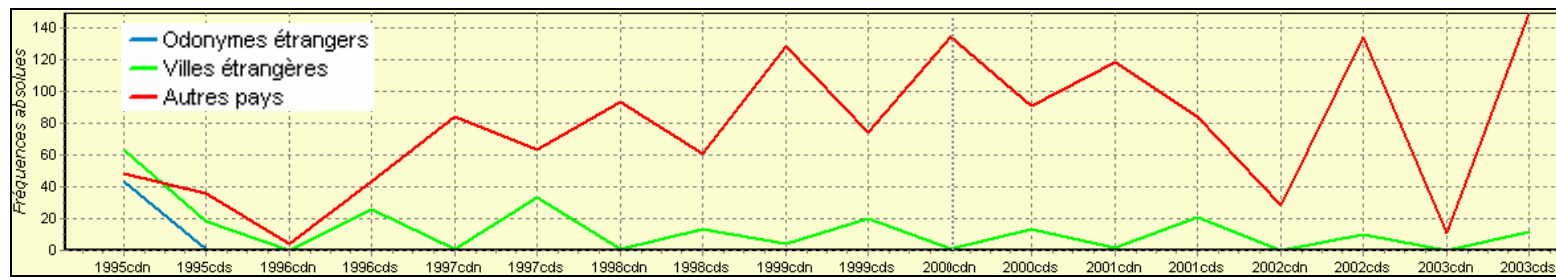


Figure 69.
Ventilation sur les neuf parties du corpus découpées en CDN et CDS des fréquences des odonymes, des villes et des pays étrangers (France non comprise)

Sur le graphique 70 ci-dessus, deux phénomènes sont mis en lumière. Le premier est celui auquel il vient d'être fait allusion, c'est-à-dire la disparition des odonymes étrangers et la diminution de la présence des noms de villes étrangères au profit des noms de pays après 1995. Ce mouvement de substitution correspond en fait à l'abandon du dispositif « coordonnées » pour les entités du groupe localisées à l'étranger. Or, l'augmentation de la présence des noms de pays dans les contextes non syntactiques suggère la mise en place d'un nouveau dispositif. Celui-ci semble être utilisé de 1997 à 2001, date à partir de laquelle ces noms diminuent de façon importante dans les contextes non syntactiques. Il n'en va pas de même pour les contextes syntactiques, dans lesquels les noms de pays semblent tenir une place croissante de 1997 à 2003¹. Si l'on fait l'hypothèse que les noms de pays constituent probablement un meilleur indicateur que les odonymes ou les noms de villes étrangers pour évaluer la montée de l'internationalisation des activités du groupe bancaire, il sera néanmoins nécessaire de préciser la manière dont cet indicateur est utilisé par les rapports d'activité dans leurs principaux contextes. Ces points vont faire l'objet des investigations qui suivent, de même que la fréquence très atypique de « France ».

B. La structuration géographique du développement international

Afin d'aborder ces points, on procèdera encore à l'examen de la répartition des noms de pays entre les deux sortes de contextes à dominante syntactique ou non syntactique, afin de mieux cerner les sortes de rôles qu'ils sont susceptibles de jouer dans les rapports d'activité.

¹ Dans le graphique x infra, on peut voir que la légère baisse des fréquences absolues pour les contextes à dominante non syntactique de 2001 correspond plutôt à une stabilisation lorsqu'elle est rapportée aux fréquences relatives qui tiennent compte de la taille de la partie.

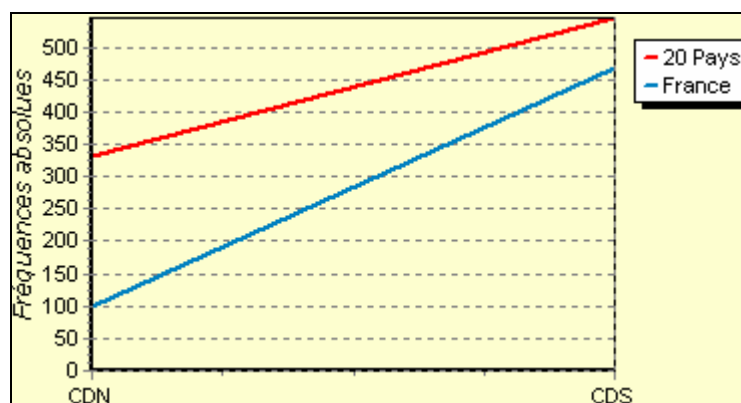


Figure 70.

Ventilation en fréquences absolues des noms de pays et des graphies de « France » entre contextes à dominante syntactique et non syntactique

La répartition des fréquences absolues observées pour le groupe « 20 noms de pays » (rassemblant les pays figurant dans le tableau donné en début de section à l'exception de la France), et pour le groupe rassemblant les deux graphies « France » et « FRANCE », indique un profil général partagé entre les deux contextes CDN et CDS ; néanmoins, il faut noter que le groupe « France » possède une répartition proportionnellement mieux équilibrée que le groupe de 20 pays entre les deux sortes de contextes, ce que mettent en lumière les fréquences relatives :

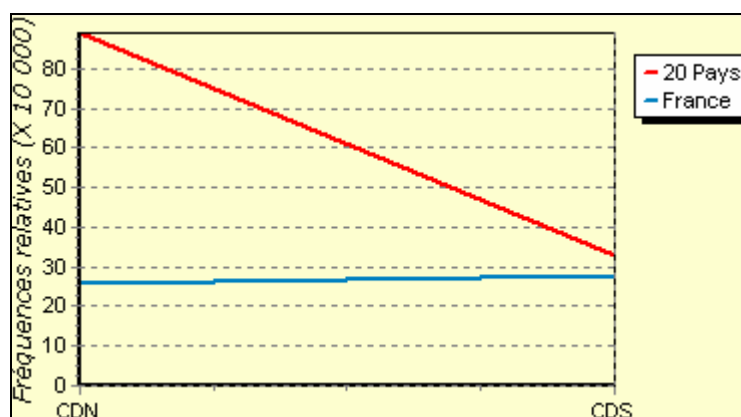


Figure 71.

Ventilation des noms de pays et des graphies de « France » entre contextes à dominante syntactique et non syntactique, en fréquences relatives

1. Le classement des entités du groupe bancaire

On cherchera donc à savoir si cette différence de répartition correspond aussi à un rôle différent qui serait dévolu à la dénomination « France ». Afin d'identifier les utilisations des noms de pays dans ces deux sortes de contextes, on a cherché à avoir une vision globale de leurs co-occurents dans le corpus. On commencera par présenter les co-occurents des noms de pays étrangers avant d'aborder le cas particulier de la dénomination « France ». Le calcul des co-occurents est du même type que celui qui a été pratiqué pour les odonymes et les noms de villes. On présente ci-dessous son résultat pour les contextes à dominante non syntactique.

Tableau 66.
Principaux co-occurents du groupe des 20 noms de pays dans les contextes de type CDN

Forme	Frq.Tot.	Frq.	Coeff.	Forme	Frq.Tot.	Frq.	Coeff.
Fax	38	35	***	I	106	26	24
Carr	138	59	***	Espagne	68	22	23
Agricole	1714	152	***	SAL	14	12	21
Banco	167	52	***	Monaco	37	17	21
SA	83	45	***	Italie	96	23	21
Ltd	108	74	***	PTE	21	14	21
Indosuez	827	166	***	Grèce	55	19	21
Téléphone	34	34	***	Maroc	24	14	20
Crédit	2992	167	43	Banque	435	41	20
Asia	31	24	38	Indocam	139	26	20
Inc	27	23	38	Allemagne	36	16	20
Luxembourg	54	27	34	España	12	11	20
Futures	70	29	34	Singapore	10	10	19
Portugal	54	27	34	Management	175	27	19
Kong	60	27	33	Cheuvreux	114	23	19
Hong	60	27	33	Française	48	17	19
Securities	78	28	31	ETRANGER	11	10	18
W	69	25	28	Chili	20	12	17
Liban	24	18	28	Japan	13	10	17
Suisse	80	25	26	Asset	161	23	16

Parmi les quarante premiers co-occurents des noms des vingt pays dont la fréquence est supérieure à 20, on distingue quatre ensembles de formes de taille inégale. Le premier rassemble deux formes appartenant à des dispositifs de type « coordonnées » (« Téléphone » et de « Fax »). Cependant, à la différence de ce qui avait été constaté pour les autres toponymes, ce type d'unités est nettement minoritaire. Cette observation conforte celle qui a été faite plus haut concernant

l'abandon du dispositif du type « coordonnées », mais elle pose la question d'un nouveau dispositif pour lequel les noms de pays seraient mobilisés dans les contextes à dominante non syntactique.

Un deuxième groupe rassemble des composants de dénominations complexes de sociétés. Afin d'éclairer ce point, on donne ci-dessous un tableau dans lequel on a fait correspondre aux formes simples typiques des CDN les dénominations complètes auxquelles elles appartiennent.

*Tableau 67.
Formes simples co-occurentes des noms de pays et principales dénominations
d'entreprises auxquelles elles se rattachent*

Formes	Dénominations propres
Agricole	Crédit Agricole Indosuez, Crédit Agricole S A
Asia	Crédit Lyonnais Securities Asia, Crédit Agricole Indosuez Merchant Bank Asia Ltd
Asset	Crédit Agricole Asset Management
Banco	Banco Bisel, Banco del Desarrollo, Banco Espirito Santo
Banque	Banque Indosuez, Banque de Financement et de Trésorerie, Banque de Gestion Privée, Banque Commerciale de Grèce
Carr	Carr Futures, Indosuez W I Carr Securities
Cheuvreux	Crédit Agricole Indosuez Cheuvreux
Crédit	Crédit Agricole Indosuez, Crédit Agricole S A
España	Carr Futures España SA, Crédit Agricole Indosuez Cheuvreux España
Française	Banque Française de l'Orient, Banque Française Commerciale Antilles Guyane
Futures	Carr Futures
I	Indosuez W I Carr Securities
Inc	Carr Futures Inc, Crédit Agricole Indosuez Securities Inc
Indocam	Indocam
Japan	Indosuez W I Carr Securities Japan Ltd, Indocam Japan Ltd
Ltd	Indosuez W I Carr Securities Ltd, Carr Futures Ltd
Management	Crédit Agricole Asset Management
PTE	Carr Futures PTE Ltd
SA	Lukas SA, Crédit Agricole Indosuez Suisse SA
SAL	Crédit Agricole Indosuez Liban SAL, Banque Libano Française SAL
Securities	Indosuez W I Carr Securities
Singapore	Indosuez W I Carr Securities Singapore PTE Ltd
W	Indosuez W I Carr Securities

Ce que font ressortir les principaux co-occurents listés dans le tableau ci-dessus, c'est l'apparition, dès 1997, d'un classement des entités du groupe – dont le

nombre s'est considérablement accru¹ –, en fonction d'une unité géographique, le pays, dont l'échelle a été jugée plus adaptée à la nouvelle dimension du groupe bancaire. On donne ci-dessous un exemple de ce nouveau dispositif qui a succédé à la présentation des coordonnées :

Amérique du Nord	
Etats-Unis	● Crédit Agricole Indosuez
	◆ Carr Global Advisor
	◆ Carr Futures Inc.
	◆ Indosuez Ventures
	◆ Indosuez Capital Securities Inc.
Amérique centrale	
Dominique	◆ Banque Française Commerciale
Guadeloupe	Antilles-Guyane
Martinique	
Mexique	◆ Indosuez Mexico SA de CV

Figure 72.

Présentation des entités du groupe bancaire localisées à l'étranger dans le rapport d'activité de 1997

Le tableau 66 comporte deux autres groupes dont il peut être intéressant de cerner de plus près les propriétés. En effet, on constate qu'une partie seulement des noms de pays projetés sur les contextes CDN se retrouve dans la liste des principaux co-occurents. Le calcul des spécificités étant appliqué à des fenêtres découpant de petits îlots au sein des textes du corpus, on peut s'attendre à ce que les formes utilisées pour repérer ces contextes particuliers – il s'agit ici des paragraphes comportant au moins l'un des noms de pays du tableau 67 –, soient relativement bien placées parmi les formes spécifiques. En effet, les noms de pays eux-mêmes font alors partie d'un groupe de formes, peut-être relativement peu nombreuses, qui apparaissent fréquemment dans le voisinage de ces noms. Cependant, cela suppose que la répartition entre les deux sortes de contextes soit au moins équivalente, ce qui n'est pas le cas, comme cela a déjà été identifié pour l'ensemble du groupe des toponymes (voir figure 63). La sélection des seuls paragraphes CDN dans lesquels

¹ Au 31 décembre 1995, le périmètre du bilan consolidé du Crédit agricole compte 180 filiales auxquelles s'ajoutent 60 caisses régionales. Après l'intégration de la banque Indosuez, au 31 décembre 1997 ce périmètre comprend 276 filiales auxquelles s'ajoutent 56 caisses régionales ainsi que 2775 caisses locales.

au moins un nom de pays appartenant à la liste est présent, a eu pour effet d'opérer une répartition de ces noms en deux sous-ensembles : le premier comporte les noms de pays qui sont nombreux dans les CDN et peut-être proportionnellement aussi nombreux ou moins nombreux dans les CDS. Ces noms de pays se trouvent ainsi bien placés dans la liste des formes spécifiques apparaissant dans les paragraphes CDN qu'ils ont permis de sélectionner. En revanche, les noms de pays qui sont plus fréquents dans les CDS sont moins bien placés dans la liste et n'apparaissent pas parmi les 40 premiers co-occurents. Il s'agit de « Singapour », « États-Unis », « Argentine », « Brésil », « Japon », « Russie » et « Royaume-Uni ». À ce stade, on peut faire l'hypothèse que la différence de répartition observée pour ce sous-groupe de pays correspond au fait qu'ils sont utilisés dans les contextes syntactiques comme cadres pour la restitution de l'activité, voire comme objets de discours, plutôt que comme organisateurs des entités du groupe dans les contextes non syntactiques.

2. Les formes du développement international

Afin de poursuivre cette comparaison, il est utile de disposer d'une vision globale des co-occurents du groupe des pays étrangers dans les contextes syntactiques.

Tableau 68.
Principaux co-occurents du groupe des 20 noms de pays dans les contextes de type CDS

Forme	Frq.Tot.	Frq.	Coeff.
Italie	96	73	***
Banco	167	86	41
Suisse	80	55	36
Espagne	68	46	30
Etats	62	44	30
Unis	62	42	28
en	3641	576	27
Europe	236	84	26
Grèce	55	36	23
Banca	60	35	20
Intesa	80	41	20
Pologne	34	26	20
pays	159	57	18
filiale	146	53	17
Hong	60	32	17
Kong	60	32	17
Singapour	33	23	16
Asie	103	40	15
En	490	108	15
Argentine	36	23	15

Forme	Frq. Tot.	Frq.	Coeff.
Bisel	42	25	15
a	2275	350	15
Portugal	54	27	14
Luxembourg	54	27	14
banques	135	46	14
détient	64	29	13
Japon	29	19	13
Russie	23	17	13
Bank	76	30	12
avec	1004	173	12
New	33	19	12
au	936	164	12
Brésil	29	18	12
York	33	19	12
Monaco	37	20	11
Allemagne	36	19	11
Lukas	43	21	11
IntesaBci	25	15	10
actionnaire	40	19	10
Tokyo	25	15	10

Comme on pouvait s’y attendre, les pays absents des quarante plus fortes spécificités des contextes CDN sélectionnés apparaissent dans le tableau ci-dessus, à l’exception du Royaume-Uni¹. On retrouve également dans ce tableau la moitié des pays signalés par les spécificités CDN (Italie, Suisse, Espagne, Hong Kong, Grèce, Portugal, Monaco, Allemagne). Les spécificités propres aux CDS dans lesquels apparaissent au moins l’un des noms de pays recensés en début de section, font aussi émerger de nouvelles formes. Un premier groupe rassemble des dénominations ou des parties de dénominations de sociétés, dont certaines sont également présentes parmi les formes caractéristiques des CDN (Banco, Bisel) ; cependant, elles ne sont pas majoritaires. En effet, « Banca », « Intesa », « Bank », « Lukas » et « IntesaBci » n’apparaissent que dans le tableau des spécificités CDS. On note, par ailleurs, un groupe d’organismes spatiaux « en », « En » et « au » qui suggèrent que le

¹ Les occurrences de « Royaume-Uni » étant relativement faibles et à peu près équitablement réparties entre CDN (9 occurrences) et CDS (12 occurrences), le calcul des spécificités n’a pas rattaché ce nom de pays plus particulièrement à l’un ou à l’autre des deux contextes.

classement par pays des CDN trouve une correspondance dans les CDS avec une présentation de l'activité pour laquelle le cadre géographique a également été utilisé. Dans le même ordre d'idées, les deux zones géographiques « Europe » et « Asie » et le substantif « pays » (utilisé dans les paragraphes comme anaphorique), trouvent peut-être un équivalent dans les CDN avec « ETRANGER ». On reviendra sur ces deux points dans la section suivante.

Enfin, un dernier groupe peut être composé avec les formes co-occurentes ne relevant a priori ni de la thématique géographique au sens large, ni des dénominations ou parties de dénominations propres d'entreprises. Il s'agit de « filiale », « banques », « a », « avec », « détient » et « actionnaire ». Avec cet ensemble, on accède indirectement aux formes d'activités pour lesquelles noms de pays et noms de sociétés sont réunis dans le corpus. Pour étendre et préciser la collecte d'informations relatives à ces relations, on rassemble ces éléments dans un groupe de formes, à l'exclusion de « a » qui est susceptible d'entrer dans tous les contextes syntactiques et qui n'est pas discriminant pour cette raison.

On projette le contenu du groupe sur la carte des sections de Lexico et on lance le calcul des spécificités sur les 949 paragraphes sélectionnés. Le résultat fait ressortir les formes suivantes¹ : « partenariat », « marché », « accord », « % », « liens », « accords », « contrats », « partenariats », « capital », « participation », « liaison », « relations ». L'examen de quelques contextes fait rapidement apparaître que ces mots appartiennent au un champ lexical du développement international du groupe bancaire. En effet, le développement du réseau du groupe bancaire peut revêtir plusieurs formes : accords avec des banques étrangères pour obtenir des délégations et prises de participation dans le capital de banques étrangères, en général assorties d'accords de partenariat pour développer localement une ou plusieurs formes d'activité. Ce développement peut prendre encore une autre forme, avec les filiales ou les succursales de filiales (ou du Crédit agricole), créées ou acquises à l'étranger. Cette dernière forme étant en général préparée plusieurs

¹ Parmi les formes ayant un coefficient de spécificité supérieur ou égal à 8, on a exclu les membres du groupe de formes projetées, et les grammaticaux, particules verbales et adverbes (a, le, un, en de, une, la, sur, sa, son, dans, à, l, plus, les, également, est, désormais).

années à l'avance, sa mise en place donne lieu à un certain nombre d'échanges (relations, liaison, liens) entre le groupe bancaire et ses partenaires étrangers.

On peut synthétiser les informations collectées dans cette section au sujet de la restitution du développement international du groupe bancaire en répartissant les pays étudiés en trois ensembles. Le premier ensemble regroupe les noms de pays fortement présents à la fois dans les contextes CDN et CDS : « Italie », « Suisse », « Espagne », « Hong Kong », « Grèce », « Portugal », « Monaco », « Allemagne » et « Luxembourg ». Cette forte présence dans les deux contextes correspond à deux phénomènes étroitement liés, mais que l'on doit néanmoins distinguer. La présence en CDN est directement liée, on l'a vu, à la présentation des implantations du groupe bancaire. L'utilisation d'un pays dans ce type de dispositif suppose que l'implantation, quelle que soit sa nature (bureau de représentation, succursale, filiale ou participation), possède déjà quelques années d'ancienneté, n'est pas menacée de disparition à court terme et revêt une importance significative qui lui vaut un tel signalement. Les noms de pays fonctionnent alors comme simples repères géographiques, rôles qu'ils partagent avec les odonymes et les noms de villes. La mention d'un pays dans un dispositif de type CDS est liée à la restitution de l'activité, c'est-à-dire ce qu'il convient de porter à la connaissance des destinataires en termes de faits remarquables apparus durant l'année écoulée. La présence d'un nom de pays dans ce type de contexte doit donc être liée à la restitution de l'activité économique en cours : les pays du groupe correspondent au périmètre européen de développement de l'activité internationale du Crédit agricole. Quant à Hong-Kong, il s'agit de la plus ancienne implantation du groupe en Asie, autre terrain de développement de l'activité. On peut déduire de ces observations que les pays dont les noms figurent principalement dans les CDN correspondent à des implantations mineures ou dont le développement n'est pas à l'ordre du jour ; sont ainsi regroupés « Maroc », « Liban » et « Chili ». On voit ainsi que seuls certains noms de pays possèdent, en plus de leur rôle de repère, celui de cadre de restitution de l'activité.

Le troisième groupe est plus hétérogène et rassemble des pays fort différents les uns des autres : « États-Unis », « Argentine », « Pologne », « Singapour », « Brésil », « Japon » et « Russie ». Il s'agit en quelque sorte de terrains de conquête et, à ce titre, ils n'ont pas la stabilité du périmètre européen. On peut distinguer

parmi ces terrains ceux qui sont liés aux principales places financières mondiales : les Etats-Unis, Singapour et le Japon sont également présents dans de nombreux contextes sous la forme de « New York », « Singapour » ou « Tokyo », qui valent à la fois comme places boursières et comme états. L'Argentine, la Pologne, le Brésil et la Russie correspondent aux pays dans lesquels le groupe bancaire a tenté de développer son activité. L'Amérique du sud et les anciens pays de l'Est sont des territoires de conquête mais aussi de revers. Les principaux co-occurents de ces noms de pays dans les CDS sont éloquentes : outre les noms de banques ou de sociétés financières, on rencontre « pays », « crise », « émergents » qui sont autant d'échos aux crises économiques ayant frappé ces pays émergents à la fin des années 1990.

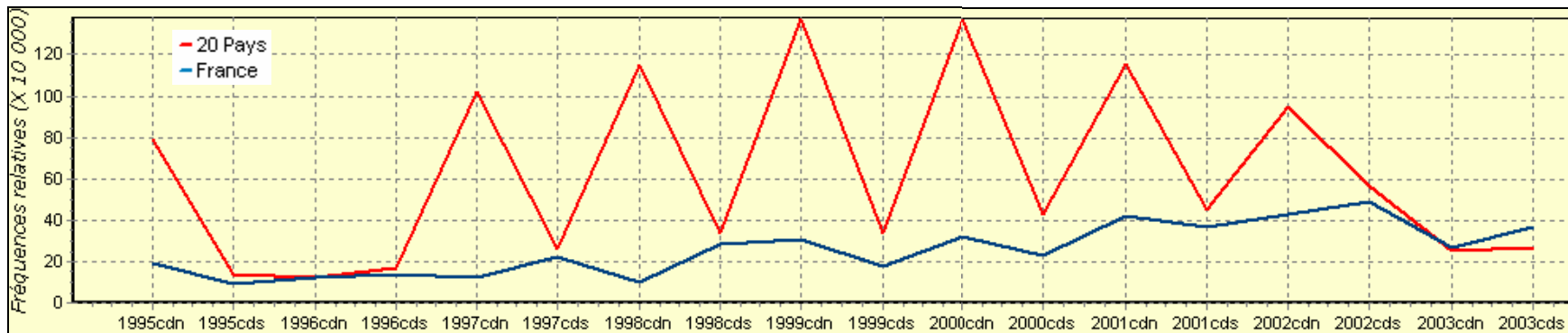
Dans la mesure où l'on a su tirer profit de la répartition des informations des rapports d'activité dans deux sortes de contextes, les noms de pays sont apparus comme un indicateur pertinent du développement international du groupe bancaire. On ne peut pas escompter le même type de résultat à partir de l'étude du vocable « France ». Il convient donc de procéder à l'examen du fonctionnement discursif de ce toponyme, et l'on cherchera en particulier à voir de quelle manière le développement international a influé sur la manière de restituer l'activité bancaire en France.

C. La structuration géographique de l'activité

Comme on vient de le voir, une certaine correspondance peut exister entre le classement géographique des entreprises rattachées au groupe bancaire et la description de l'activité dans les rapports d'activité. Il convient d'examiner ce point plus en détail en rapprochant les observations obtenues à propos des noms de pays étrangers de celles que l'on va chercher à réaliser pour le toponyme « France ». Par ailleurs, il s'agit également de repérer les unités qui sont plus particulièrement utilisées dans la restitution de l'activité lorsque la géographie est appelée à jouer un rôle structurant. On se demandera alors s'il s'agit dans tous les cas de la même « géographie ».

Afin de rapprocher les précédentes observations de celles que l'on compte conduire au sujet du toponyme « France », il a semblé utile de comparer la manière

dont les occurrences de ce vocable et celles des noms de pays dont il vient d'être question sont réparties dans l'ensemble du corpus.



*Figure 73.
Ventilation en fréquences relatives des groupes de formes « 20 pays » et « France » sur les parties du corpus découpées en fonction des contextes de type CDN et CDS*

L'allure générale du graphique donné ci-dessus appelle une première remarque relative aux extrémités de la courbe des fréquences des pays¹. On remarque, en effet, que cette présence est relativement importante dans les contextes à dominante non syntactique de la partie centrale du graphique. En revanche, pour les années 1996 et 2003 on observe des fréquences nettement plus basses. On peut penser que les deux événements majeurs intervenus au cours de ces deux années et mentionnés dans les rapports d'activité, c'est-à-dire l'achat de la banque Indosuez en 1996 et celui du Crédit Lyonnais en 2003, ont eu des répercussions si importantes qu'ils ont provoqué une sorte de « remise à zéro » en matière de positionnement géographique. Le système de coordonnées du référentiel a donc été mis momentanément entre parenthèses, le temps que les éléments nouveaux y soient convenablement intégrés. Il reste que l'on ne constate pas le même mouvement pour les fréquences de « France » ; c'est le point qu'il convient d'examiner maintenant.

Dans le graphique 73 la comparaison de la ventilation des occurrences du groupe « France » (comportant également la graphie « FRANCE ») avec celle des vingt pays dont les noms sont les plus fréquents révèle que ces derniers sont trois fois plus fréquents dans les CDN que les graphies de « France ». Cette différence de traitement s'explique par le fait que dans les dispositifs de coordonnées ou dans les présentations des entités du groupe organisées par zones géographiques, la localisation « France » fonctionne largement sur le mode de l'implicite. Par exemple, on retrouve dans toutes les parties du corpus, excepté pour 2002, les coordonnées des Caisses régionales de Crédit agricole, toutes localisées en France ; cependant, l'émetteur a jugé superflu de préciser qu'il s'agissait à chaque fois

¹ L'allure générale de la courbe représentant la ventilation des occurrences du groupe de formes « 20 pays » est différente de celle du graphique 69 donné plus haut (comparaison de la ventilation des occurrences des odonymes, des noms de villes et de pays étrangers). Cette différence tient au fait que la courbe représente ici les fréquences relatives du groupe de formes, c'est-à-dire que les points du graphique sont calculés à partir de la fréquence absolue ramenée à la taille de la partie considérée. Autrement dit, pour une fréquence équivalente, plus la partie en question est largement dimensionnée, plus les fréquences des noms de pays risquent de se trouver diluées dans le reste du vocabulaire de la partie. Ce phénomène est surtout sensible pour la partie 2003CDS, dont le vocabulaire est environ deux fois plus important que celui des autres parties CDS. Alors que les fréquences absolues indiquent une hausse de la présence de ces unités, ramenée à la taille de la partie, les fréquences relatives indiquent une baisse assez importante.

d'adresses situées en France puisque, par définition, les Caisses régionales ont leur siège dans ce pays. Cet implicite explique pour une bonne part la faible présence relative du groupe « France » dans les CDN. Cependant, il reste à expliquer pourquoi ces graphies ne sont pas absentes des CDN. On reviendra sur ce point un peu plus loin, dans la mesure où un troisième et dernier phénomène important doit encore être commenté.

En effet, si les courbes des deux groupes de formes ont des allures très contrastées pour ce qui est des contextes à dominante non syntactique, elles sont relativement proches pour les CDS. Globalement, sur la période couverte par le corpus, on assiste à une progression régulière de la fréquence des noms de pays, y compris de « France ». Même si les fréquences ne progressent pas de façon linéaire, la tendance depuis 1997 est bien à une augmentation parallèle ; et la baisse enregistrée en 2003 se fait aussi dans des proportions voisines pour le groupe des noms de pays et pour les graphies de France. On peut maintenant chercher à mettre au jour les différentes formes prises par cette dimension géographique lorsqu'elle est utilisée afin d'organiser la présentation de l'activité du groupe bancaire.

1. Réutilisation d'une signalétique

Il s'agit tout d'abord d'expliquer la présence paradoxale des occurrences de « France » dans les CDN. En effet, on a vu que la dénomination « France » est peu utilisée dans les CDN parce qu'il s'agit d'une référence géographique implicite que l'émetteur des rapports d'activité a estimé pouvoir omettre. Comment, dans ce cas, expliquer la présence d'un certain nombre d'occurrences dans ces contextes ? Afin d'éclaircir ce point, on calcule les co-occurents les plus fréquents des paragraphes à dominante non syntactique dans lesquels apparaissent les graphies « France » ou « FRANCE ».

*Tableau 69.
Quarante premiers co-occurents des graphies de « France » dans les contextes
CDN*

Forme	Frq.Tot.	Frq.	Coeff.
France	543	78	***
FRANCE	23	20	36
Bastide	15	12	21
Président	724	46	19
73	48	15	18
Téléphone	34	13	17
Libération	8	8	16
Pierre	169	22	16
Fax	38	13	16
63000	8	8	16
Directeur	724	42	16
Général	485	33	15
MF	663	38	15
Banque	435	31	15
Ferrand	9	8	15
Clermont	9	8	15
Rapée	7	7	14
Administrateur	32	11	14
VAL	7	7	14
Boutet	7	7	14

Forme	Frq. Tot.	Frq.	Coeff.
75012	7	7	14
Total	688	38	14
PARIS	7	7	14
28000	7	7	14
proximité	199	20	13
Chartres	9	7	13
30	186	20	13
57	40	11	13
22	107	15	12
Edouard	10	7	12
quai	17	8	12
Daniel	13	7	11
Centre	41	10	11
Foncaris	5	5	10
Tél	449	25	10
DE	111	13	10
26	92	12	10
CENTRE	28	8	10
Gestion	135	14	10
Maurice	25	8	10

Le calcul des spécificités permet d'établir que les occurrences du groupe « France » sont majoritairement dues à l'utilisation du nom de pays comme composant à l'intérieur de dénominations complexes. On a ainsi les Caisses régionales « Paris et Ile-de-France », « Centre France », « Val de France » et « Nord de France » qui totalisent une trentaine d'occurrences. Trois de ces dénominations propres sont antérieures au développement international. Cette utilisation du toponyme rentre dans le cadre d'une signalétique banale où « France » assure un double effet d'ancrage et de valorisation de la réalité dénommée, comme c'est par exemple le cas pour « Air France », « France Télécom » ou « Gaz de France ».

À partir de 1999 apparaît une deuxième série de dénominations constituée par l'ajout de « France » aux dénominations de filiales ou de groupes affiliés. Cette fois, il semble bien que cette adjonction puisse être imputée à l'internationalisation de l'activité du groupe bancaire. En effet, ces entités ayant développé une partie de leurs activités à l'étranger, leur présentation dans les CDN a entraîné l'apparition du distinguo « Groupe Sofinco France » et « Groupe Sofinco Etranger » ; il en va de même pour Predica ou Ucabail, autres filiales du groupe bancaire. C'est la même

influence du développement international que l'on peut déceler dans la fabrication de certaines dénominations propres recensées dans la section B. On peut remarquer dans ces exemples que le toponyme est parfois donné dans la langue à laquelle la dénomination propre complète se rattache : « Indosuez W I Carr Securities Japan Ltd » ou « Indosuez W I Carr Securities Singapore PTE Ltd ». Cependant, des statuts mixtes sont également utilisés comme pour « Crédit Lyonnais Securities Asia », mais aussi avec « Carr Futures España SA ». La réutilisation d'une signalétique existante de noms de sociétés correspond toujours au souci d'ancrer la dénomination dans un contexte national précis, mais plutôt que la recherche d'une valorisation indirecte de l'entité nommée, comme pour « France Télécom », on devine dans ces noms de sociétés étrangères la manifestation d'une volonté de classement, proche des regroupements par pays repérés dans les contextes non syntactiques du corpus. En effet, les noms de pays permettent de distinguer les sociétés tout en maintenant un lien de filiation grâce à la réutilisation de la dénomination d'origine. Ainsi, Carr Futures España SA « doit » être la filiale de Carr Futures en Espagne.

Lorsque l'activité est présentée de manière thématique, la distinction entre la part de l'activité exercée en France et celle qui est exercée à l'étranger est également utilisée. On a, par exemple, une séparation entre la « Gestion d'actifs France » et la « Gestion d'actifs Etranger », de même a-t-on les « Participations bancaires », « En France » et « À l'Etranger ». Le syntagme « Banque de proximité en France » apparaît une douzaine de fois dans les contextes CDN sur un total de 36 occurrences dans l'ensemble du corpus. Enfin, cette distinction a également gagné, semble-t-il, les dénominations des directions de l'organe central où apparaissent des intitulés comme « Marchés de proximité en France et coordination de l'offre », « Planification stratégique et participations France ». C'est dire que l'on dispose ainsi d'un autre indice de l'internationalisation du groupe bancaire. Son développement à l'étranger a eu pour effet d'activer la réutilisation d'une signalétique reposant sur l'utilisation de la distinction « en France » – « hors-de-France » au sein même de certaines dénominations propres existantes. Il reste à prendre la mesure de ce phénomène dans le texte suivi des contextes à dominante syntactique.

2. France VS monde : une nouvelle organisation de la restitution de l'activité bancaire

Afin de disposer d'une vue d'ensemble sur les principaux co-occurents des graphies de « France » dans les paragraphes des contextes à dominante syntactique, on lance un calcul des spécificités sur les sections qui comportent au moins une occurrence du toponyme. On donne ci-dessous le résultat du calcul pour les formes les plus spécifiques.

Tableau 70.

Les quarante premiers co-occurents du toponyme « France » dans les CDS

Forme	Frq.Tot.	Frq.	Coeff.
France	543	464	***
en	3641	820	50
proximité	199	72	16
Fondation	30	21	13
effectifs	42	25	12
CDI	24	17	11
les	2512	450	10
Télécom	16	13	10
Association	18	13	9
positions	107	37	8
cadres	22	14	8
Calyon	54	24	8
et	4486	734	8
Europe	236	62	8
Sofinco	165	45	7
monde	94	31	7
pour	1026	192	7
Loire	27	14	7
Centre	41	18	7
banque	528	111	7

Forme	Frq.Tot.	Frq.	Coeff.
Régionales	176	46	6
UAF	11	8	6
patrimoine	39	17	6
Pays	39	17	6
amplifier	16	10	6
source	34	15	6
Maires	6	6	6
Cadres	6	6	6
1979	6	6	6
Nord	44	18	6
préretraites	6	6	6
Finaref	30	14	6
position	76	25	6
bancaire	229	56	6
poste	18	11	6
Non	7	6	5
Réseau	7	6	5
UTP	5	5	5
de	12744	1870	5
femmes	6	5	5

Au moins trois regroupements de formes peuvent être constitués à partir des données fournies par ce tableau qui rassemble les principales spécificités positives des graphies « France » et « FRANCE » dans les CDS du corpus. Dans la présente section, on traitera du premier ensemble, les deux autres figurant dans les sections suivantes.

Un premier groupe peut être constitué à l'aide de la seule forme « en » qui apparaît en deuxième position parmi les principales spécificités positives. On se souvient que les spécificités positives du groupe des vingt dénominations de pays avaient également mis en évidence cette forme, ainsi que « En » et « au ». On fait

l'hypothèse que dans ces trois cas, l'usage majoritaire est celui d'une préposition qui, placée devant un nom propre de lieu, indique que les faits rapportés ont pour cadre la région ou le pays mentionnés. Afin de vérifier ce point on a recours aux inventaires distributionnels qui, entre autres dans Lexico, fournissent le contexte droit de la forme retenue. On sait que ces inventaires permettent d'avoir une vision plus synthétique que les seules concordances, lorsque le nombre d'occurrences à examiner est élevé ; or, c'est bien le cas ici, puisque « en » possède 3641 occurrences, « au » 936 et « En » 490. Dans la mesure où « en » présente le cas le plus riche, on n'étudiera ici que les distributions dans lesquelles « en » apparaît devant une ou plusieurs autres formes avec une fréquence supérieure à 10. On obtient alors soixante-dix huit distributions débutant par « en », totalisant 2844 occurrences sur 3641, soit près de 80% des occurrences¹. Les principales fréquences sont rassemblées par ordre décroissant dans le tableau ci-dessous.

Tableau 71.

Premières fréquences de l'inventaire distributionnel droit de la forme « en »

Distributions	Frq.
en France	318
en Europe	145
en 2002	118
en 2003	106
en 1996	105
en place	104
en 1995	89
en 2001	84
en 1997	80
en 1998	72
en hausse de	65
en matière de	61
en Asie	56
en progression	56
en 1999	55
en oeuvre	52
en 2000	46
en progression de	46
en particulier	44

¹ Cette liste est donnée en annexe.

Les principales fréquences permettent d'avoir une idée assez précise des sortes de contextes dans lesquels la préposition « en » est majoritairement utilisée. Soit celle-ci est placée devant une date ou un nom de pays, ou de région géopolitique, utilisés comme cadres temporels ou spatiaux pour les faits rapportés, soit elle est intégrée dans diverses locutions. Pour ce qui est des toponymes on retrouve les cadres « France » et « Europe », suivis d'assez loin par « Asie ». Lorsque l'on additionne les fréquences concernant seulement les toponymes on obtient 659 occurrences (dont la moitié environ concerne la France), sur 2844, soit approximativement un quart des occurrences. Si la préposition n'est pas exclusivement utilisée devant des noms de pays ou de régions, il reste que ce cas de figure est loin d'être négligeable. Ce type de marqueur discursif de la spatialité (et il en existe d'autres qui pourraient être également pris pour exemples), contribue à la structuration des discours sur l'activité du groupe bancaire. On reviendra sur ce point dans le dernier chapitre.

Par ailleurs, il apparaît que ce découpage géographique ne vise pas seulement à reproduire une partition statique et conventionnelle du monde. En effet, il semble intégré au projet même de développement qui anime le groupe bancaire, comme le suggèrent certaines coordinations entre « France » et un toponyme ou un mot apparenté. C'est ainsi que l'inventaire des distributions permet de repérer les syntagmes « en France et dans le monde » (18 occ.), « en France et à l'international » (12 occ.) ou encore « en France et en Europe » (11 occ.). On peut alors chercher à savoir si cette présence correspond aussi à une évolution dans la série chronologique, une sorte d'internationalisation du discours des rapports d'activité.

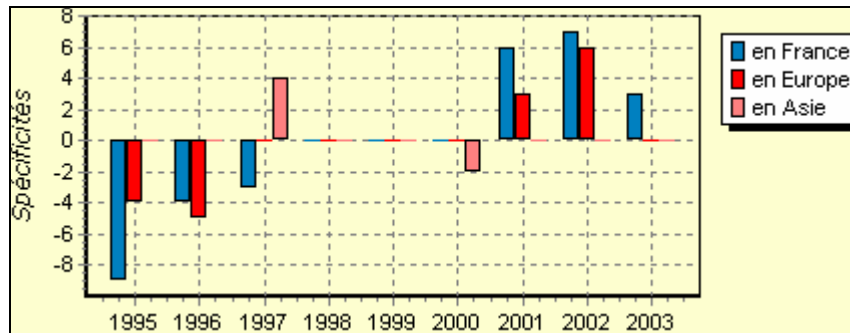


Figure 74.

Spécificités des segments répétés « en France », « en Europe » et « en Asie »

La répartition des spécificités des trois segments ayant pour tête la préposition « en » suggère un double mouvement. Dans la période 1995-1996 les préoccupations relatives au développement international du groupe ne font pas l'objet de commentaires fournis ; cette relative discrétion correspond à un implicite très fort : l'ancrage essentiellement national du groupe bancaire. Les remarques formulées précédemment, au sujet des odonymes étrangers et de la « remise à zéro » des coordonnées qui caractérisent 1996, vont dans le même sens. L'année 1997 correspond à l'intégration de nouvelles coordonnées géographiques dans le référentiel du groupe bancaire. La forte présence en Asie de la banque Indosuez donne le prétexte pour exposer les nouvelles zones d'activité du groupe bancaire. Les années suivantes sont des années caractérisées par des crises boursières et économiques importantes touchant, entre autres, l'Asie. On peut alors penser que les spécificités enregistrent un double mouvement, de banalisation pour les fréquences, et de discrétion au niveau de l'énonciation. Enfin, la période 2000-2003 semble être caractérisée par l'apparition d'une nouvelle géographie, recentrée sur l'Europe, ce qui est confirmé par les précédentes analyses relatives aux pays étrangers.

Le deuxième mouvement que l'on peut suivre en partant du graphique est une sorte de tension entre un espace d'ancrage historique, la France, et un espace de conquête commerciale, l'Europe¹. La géographie est ici au service de l'énonciation

¹ Ce que les fréquences révèlent est également pris en charge par le découpage général du texte des rapports d'activité en rubriques de sommaires. En effet, c'est en 1997 que la rubrique « Le Crédit

d'un projet de développement économique. Quoique de façon un peu différente, cette problématique se dégage aussi des résultats du graphique ci-dessous.

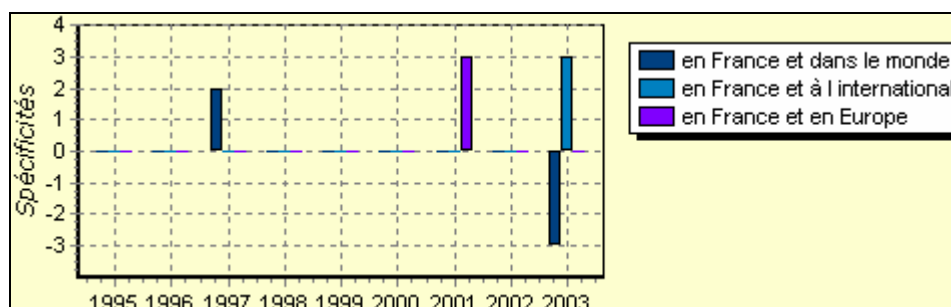


Figure 75.

Spécificités des segments répétés¹ « en France et dans le monde », « en France et à l'international » et « en France et en Europe »

Les spécificités de trois segments répétés ayant « en France et » pour tête suggèrent une lecture complémentaire du phénomène qui vient d'être décrit. Tout se passe comme si, une géographie fantasmée, représentée par l'expression récurrente « en France et dans le monde » avait cédé la place à la fin de la période à une géographie plus pragmatique, recentrée sur l'origine, la France, et ce qui constitue son pourtour immédiat, l'Europe, voire l'Afrique. De ce point de vue « en France et dans le monde » s'oppose à « en France et en Europe », alors que le syntagme « en France et à l'international », qui reprend la forme banale « international » (elle fait partie du vocabulaire commun), apparaît comme un credo tardif propre aux dernières parties du corpus. L'analyse factorielle des correspondances vient appuyer ces analyses en proposant une « carte » de l'évolution de la géographie de l'activité du groupe bancaire sur une dizaine d'années.

agricole dans le monde » fait son apparition ; son titre est remplacé en 2001 par « Un acteur majeur en France et en Europe ».

¹ « en France et dans le monde » (18 occ.), « en France et à l'international » (12 occ.) et « en France et en Europe » (11 occ.).

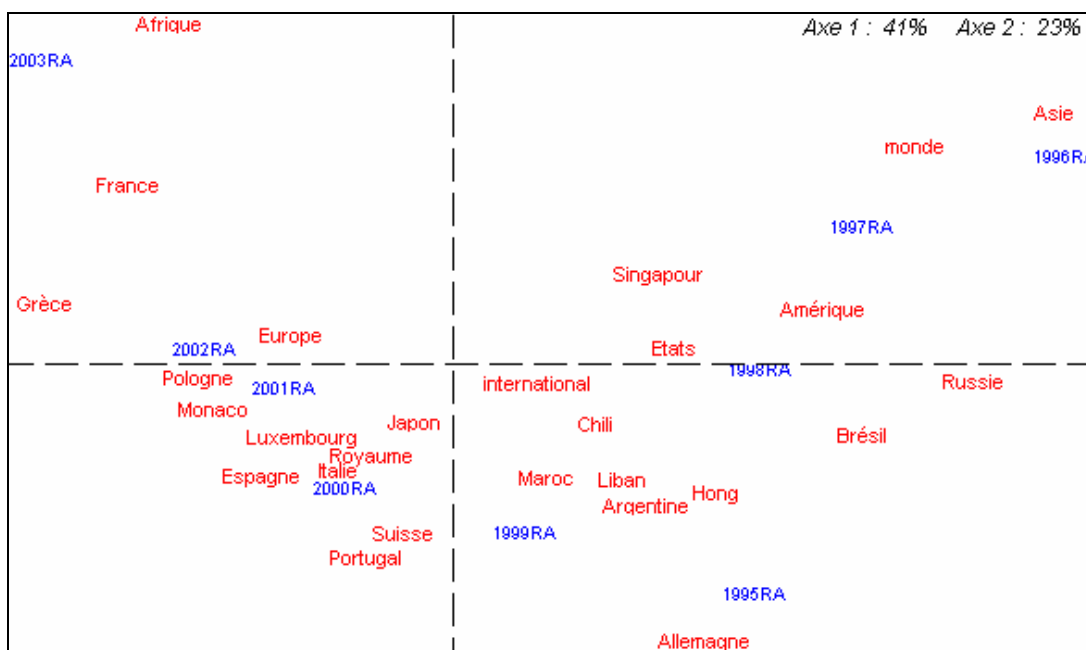


Figure 76.
Analyse factorielle des correspondances des fréquences d'une liste de 27 toponymes¹ ou apparentés du corpus Rapports d'activité (RA)

Lecture de la figure 76 : cette figure représente le résultat d'une analyse factorielle des correspondances ayant porté sur la répartition des fréquences dans les neuf parties du corpus d'une liste de 27 formes correspondant à des toponymes ou apparentés (monde, international). Les formes apparaissent en rouge et les parties du corpus apparaissent en bleu. L'axe horizontal correspond au facteur 1 (41%) et l'axe vertical au facteur 2 (23%). Cette analyse s'appuyant sur des formes simples et non des segments répétés, les noms de pays composés ont été ramenés à la forme simple ayant la même fréquence ou la fréquence la plus proche de la forme composée : « Royaume » correspond à « Royaume-Uni », « Hong » à « Hong Kong » et « Etats » à « Etats-Unis ».

L'analyse factorielle des correspondances² permet ici de comparer la manière dont les occurrences de certains noms de pays et apparentés se répartissent dans les neuf parties que compte le corpus. Cette comparaison prend la forme de plusieurs facteurs classés par ordre décroissant de quantité d'informations prises en compte.

¹ Cette liste est composée des noms des 21 noms de pays les plus fréquents auxquels ont été ajoutés, afin de tenir compte des analyses précédentes, certains noms de régions géopolitiques (Afrique, Amérique, Asie, Europe), ainsi que de deux substantifs désignant des espaces très généraux (monde, international) que l'on rencontre dans les contextes où des noms de pays sont également présents.

² La liste de formes et le graphique ont été obtenus à l'aide du logiciel Hyperbase (version 5.1). Il s'agit du même corpus (RA9503) qui est par ailleurs exploité avec le logiciel Lexico 3.

Ces informations proviennent d'un extrait du tableau lexical entier qui reprend toutes les formes présentes dans le corpus pour constituer les lignes d'un tableau, dont la deuxième entrée correspond neuf aux parties du corpus. Chaque case de ce tableau enregistre pour une forme donnée sa fréquence dans une partie donnée. Ici, les noms de pays et apparentés sont organisés dans le graphique en fonction des deux premiers facteurs, c'est-à-dire ceux qui généralement totalisent la plus grande partie de l'information extraite du tableau lexical. L'extrait du tableau lexical est constitué ici par les 27 lignes correspondant chacune à une forme graphique représentant un nom de pays, dont les fréquences sont réparties sur les neuf parties du corpus. Le résultat de l'analyse factorielle des correspondances rapproche ou éloigne les noms de pays en fonction des ressemblances ou des différences existant entre la ventilation de leurs occurrences dans les neuf parties du corpus. Ces dernières se positionnent les unes par rapport aux autres en fonction des ressemblances ou des différences qui existent entre les valeurs rassemblées dans leurs colonnes.

Les résultats de l'analyse factorielle des correspondances appellent plusieurs lectures complémentaires. Tout d'abord, on constate que l'AFC appliquée à cette série de toponymes met en évidence les propriétés d'une série textuelle chronologique dans laquelle les parties se succèdent selon l'ordre chronologique sur une courbe traversant les quadrants découpés par le croisement des deux premiers facteurs en leur point d'inertie. Ce croissant part de « 1996 », situé dans le quadrant supérieur droit, pour se terminer dans le quadrant supérieur gauche avec « 2003 », en passant par « 1999 » situé à proximité de l'axe 2, dans le quadrant inférieur droit. Les parties du corpus sont globalement organisées en fonction du premier axe qui représente la manière dont cet extrait du vocabulaire du corpus s'est renouvelé dans le temps.

On note cependant une exception pour l'année 1995, qui se situe en dehors de ce mouvement général. Cette position confirme les observations réalisées au début du chapitre, selon lesquelles la première partie du corpus semble clore une période antérieure – remontant peut-être à 1988 –, plus qu'elle ne se rattache à la période nouvelle inaugurée par l'acquisition de la banque Indosuez. On voit ici que le changement de la géographie du groupe bancaire que l'on peut faire débiter en

1996-1997 est nettement marqué. On remarque par ailleurs, que l'année 2003 s'écarte de la série 1997-2002 pour au moins trois raisons qui tiennent toutes à l'élargissement de la zone d'activité du groupe. De plus, on va voir que les trois toponymes qui occupent des positions voisines représentent chacun une forme particulière de cette expansion. L'Afrique apparaît dans la géographie du groupe bancaire du fait de l'acquisition du Crédit Lyonnais. En effet, celui-ci disposait d'un réseau relativement important dans cette partie du monde, ce qui n'était le cas ni pour le Crédit agricole ni pour la banque Indosuez. Mais l'achat du Crédit Lyonnais a surtout eu pour effet de renforcer le Crédit agricole sur des zones où il était déjà implanté, à commencer par la France. À la suite du rapprochement des deux groupes bancaires, Crédit Agricole s.a. s'est trouvé à la tête de deux réseaux de banque de détail concurrents : celui des Caisses régionales de Crédit agricole, son actionnaire majoritaire via la société holding SAS Rue La Boétie, et celui du Crédit Lyonnais, dont Crédit Agricole s.a. est l'actionnaire majoritaire. Autrement dit, à partir de 2003 la « banque de proximité en France » est représentée par deux réseaux dont le rapport d'activité doit rendre compte, ce qui multiplie les occasions de rappeler le cadre national du développement de la banque de détail. Par ailleurs, le nombre de commentaires mentionnant la France a également été augmenté en raison du rapprochement des deux banques dans d'autres formes de l'activité bancaire entendue au sens large, comme la gestion d'actifs ou la banque d'affaires qui, elles aussi, sont exercées en France à titre principal. Enfin, la Grèce entre dans le champ des développements du groupe bancaire en Europe à travers des partenariats et autres accords tissés avec des établissements bancaires étrangers.

Enfin, on remarque que la figure 76 met en évidence des oppositions dont il a été question plus haut. On note ainsi sur l'axe horizontal, qui oppose les dernières parties du corpus aux premières parties, le clivage entre « France » d'une part, et « monde » et « Asie » d'autre part, « Europe » et « Amérique », mais aussi « Pologne » et « Russie ». Dans ce recentrage européen le Japon fait figure d'intrus ; en fait, il se rapproche du centre de la figure, marqué par l'intersection des deux axes, et qui correspond à une position neutre, c'est-à-dire à une répartition relativement homogène des fréquences dans les différentes parties du corpus. On sait que ce pays correspond à une implantation historique de la banque Indosuez,

fortement intégrée au référentiel de coordonnées géographiques, contrairement aux terrains de conquêtes représentés par certains pays émergents (Brésil, Russie), dont les occurrences ont eu du mal à résister aux crises financières ayant marqué la fin des années 1990. On note le même positionnement central pour « international », dont on a vu qu'il correspond majoritairement à un substantif désignant de façon générale les marchés extérieurs au marché français¹ : malgré les revers économiques, la volonté de développement international du groupe bancaire est affirmée de façon relativement constante. Comme on l'a vu plus haut, celle-ci connaît même une augmentation de fréquence dans la dernière partie du corpus.

Ce faisant, on a mis en évidence que les noms de pays ne fonctionnent pas seulement comme un cadre extérieur sur lequel s'appuie la restitution de l'activité, mais qu'ils incarnent aussi une stratégie de développement économique, dans laquelle le monde est synonyme de marché international. De ce point de vue, la France est progressivement mise en scène dans les rapports d'activité comme étant le point d'origine de cette dynamique d'expansion économique.

3. Les pays vus comme des marchés

D'autres co-occurents de « France » dans les contextes CDS vont permettre d'éclairer cette géographie du marché. En effet, les formes « proximité », « positions », « banque », « patrimoine », « source », « position », « bancaire » et « réseau », surtout lorsqu'elles sont incluses dans des locutions repérées à l'aide des segments répétés permettent, chacune à leur manière, de caractériser l'activité du Crédit agricole qui est présenté à satiété comme « occupant le terrain » bancaire. La métaphore spatiale se rencontre dans l'expression « banque de proximité en France » qui fonctionne comme un lointain écho du slogan apparu en 1976 du « bon sens près de chez vous ». Une « position de leader » (30 occurrences) est toujours relative à un marché donné, en l'occurrence le marché bancaire français, et les « positions commerciales » (10 occ.), même s'il s'agit de « positions de premier plan » (15 occ.) doivent être défendues en permanence. Le « premier réseau bancaire français » (19

¹ Voir chapitre 4.

occ.) se doit en outre « d'amplifier ses développements » (7 occ.) dans des secteurs comme la « gestion de patrimoine » (18 occ.).

Ces affirmations pourraient facilement passer aux yeux du lecteur pour des démonstrations d'autosatisfaction un peu gratuites. Afin de prévenir ce mouvement de défiance, l'émetteur semble avoir recherché l'appui d'un regard tiers, crédité a priori de plus d'objectivité, afin de cautionner la présentation de ses réussites. Dans les dernières parties du corpus, chaque affirmation du type de celles que l'on a données en exemple est accompagnée de sa « source » d'information, en général un journal économique ou une agence de notation ayant pignon sur le marché. De référence implicite et un peu étriquée à l'ère de l'économie mondialisée, la France est présentée dans la dernière période des rapports d'activité comme une sorte de base arrière commerciale à partir de laquelle les filiales et les autres entités gravitant autour du groupe bancaire sont lancées à la conquête de nouveaux marchés « en Europe » (165 occ.) et « à l'international » (73 occ.)¹.

En effet, les spécificités positives du tableau 70 livrent les formes « Calyon », « Sofinco », « Finaref » qui sont des filiales dont l'activité s'exerce de manière historique à l'étranger (« Calyon² »), ou bien sur un marché en plein essor en Europe, le crédit à la consommation. Dans les extraits que l'on donne ci-dessous pour chacune des filiales du groupe bancaire, on voit comment la dynamique géographique du développement est reprise et mise en scène dans les rapports d'activité. Ainsi pour Calyon, cette dynamique prend-elle nettement la forme d'une série de cercles concentriques :

« Avec un coeur de cible historique en France, Calyon réaffirme ses ambitions de croissance dynamique sur le marché européen. Dans le reste du monde, la stratégie de développement de Calyon est ciblée et adaptée à ses marchés, à leur potentiel et

¹ Lorsque l'on compare les spécificités associées à certains syntagmes lexicalisés, on retrouve cette tendance générale au recentrage de l'activité. Ainsi, le syntagme « marchés émergents » (13 occ.) est-il spécifique de l'année 1998, date de la crise financière déjà mentionnée, et les unités complexes « marché français » (36 occ.) et « marché européen » (9 occ.) sont caractéristiques de l'année 2001, la première des deux unités étant sous représentée en 1995.

² Calyon résulte du rapprochement entre la « Banque de Financement et d'Investissement » du Crédit Lyonnais et Crédit Agricole Indosuez, nom donné à l'époque à la banque Indosuez après son achat par le Crédit agricole en 1996.

*aux besoins de sa clientèle locale*¹. ». Pour Sofinco, autre filiale du groupe bancaire, ce développement prend un tour plus précis avec la présentation d'une série d'implantations qui évoquent, comme des petits drapeaux piqués sur une carte d'état major, les positions conquises ou en passe de l'être :

*« Cette nouvelle filiale a su s'appuyer sur son intégration dans le Crédit agricole pour accentuer son développement en France, avec le lancement de la marque Viaxel, dédiée au crédit à l'automobile et la signature d'un accord de partenariat exclusif avec les magasins Décathlon, numéro un français de la vente d'articles de sport. (...) La société a également poursuivi son expansion internationale. Elle a ouvert, en Hongrie, sa sixième filiale internationale, Credigen, et signé un accord de coopération avec FinConsum, société spécialisée du groupe bancaire espagnol Caixa, portant sur le développement du crédit à la consommation en Espagne. Cet accord se concrétise notamment par l'entrée de Sofinco au capital de FinConsum à hauteur de 45 %. Sofinco a ainsi élargi son cadre d'intervention à l'étranger, où elle a déjà développé des partenariats en Allemagne, aux Pays - Bas, en Italie, au Portugal et au Maroc*². »

Le passage utilisé pour Finaref vient rappeler que ce qui est vu depuis chaque filiale comme un développement de sa propre activité correspond au niveau du groupe bancaire à un renforcement global de la position qu'il occupe sur un marché donné dans le cadre européen :

*« Parallèlement, l'intégration de Finaref - acquis en février 2003 - s'est déroulée dans d'excellentes conditions et permet au Groupe de se placer au premier rang des acteurs sur le crédit à la consommation en France et de renforcer sa place en Europe*³. »

Les noms de pays cumulent ainsi les rôles de repères spatiaux, de cadres servant à découper la restitution de l'activité bancaire, et enfin, de marchés, chaque État étant alors perçu comme une forme de débouché. Cette géographie des marchés sur lesquels se déploie l'activité bancaire occupe une place importante dans les rapports d'activité. Cependant, elle ne doit pas masquer un autre usage pragmatique

¹ Rapport d'activité 2003.

² Rapport d'activité 1999.

³ Rapport d'activité 2003.

des toponymes qui, comme pour les odonymes, se situe sur un autre plan communicationnel. Cette autre utilisation de la géographie, peut-être plus subtile, sera abordée à l'aide des deux derniers groupes de formes co-occurentes de « France ».

4. Une « géographie » pour le marché : développement durable et mutualisme

Cette géographie pour le marché rassemble deux groupes de formes présentes dans le tableau 70. Le premier est formé par le regroupement de « effectifs », « CDI », « cadres », « Cadres », « préretraites », « poste », « UTP », « femmes ». La plupart de ces unités se trouvent dans la dernière partie du corpus et leur présence repose au moins sur deux raisons principales liées à l'achat du Crédit lyonnais. En effet, le rapprochement des deux banques a donné lieu à un plan de réduction d'« UTP » (pour « Unités Temps Plein »). Cette diminution des effectifs du personnel a été détaillée dans la mesure où elle représentait un enjeu particulièrement suivi par les marchés financiers, qui voyaient dans la manière dont cette question serait traitée une indication de la viabilité économique du projet de rapprochement des deux banques.

La deuxième raison tient probablement au traitement du développement durable dans les rapports d'activité à partir des années 2000. En effet, la problématique du développement durable, apparue à la fin des années 1990¹, fait son entrée dans les rapports d'activité des banques françaises une dizaine d'années plus tard². Parmi les points relevant du développement durable, il y a les conditions de travail des salariés. Ce point fait l'objet d'une notation par des agences spécialisées qui, chaque année, donnent une note aux établissements cotés. Cette notation étant

¹ Le « rapport Brundtland » (rapport également intitulé « Our Common Future ») a été présenté à l'Assemblée générale des Nations Unies en août 1987. La notion de développement durable y est définie en ces termes : « § 27. *Humanity has the ability to make development sustainable to ensure that it meets the needs of the present without compromising the ability of future generations to meet their own needs. (...)* » *Report of the World Commission on Environment and Development*, United Nations, 1987, p. 24.

² Les « ressources humaines » n'étaient pas tout à fait absentes des rapports d'activité des principales banques françaises avant l'année 2000, mais leur présence était plutôt discrète. Si l'on s'en tient au seul indicateur de la fréquence on note des fréquences oscillant entre 4 et 10 avant 2000, et entre 10 et 30 après l'année 2000 dans les rapports d'activité de BNPParibas, Société générale, Crédit lyonnais et Crédit agricole.

susceptible d'avoir des répercussions sur le cours de bourse (sélection d'un investissement « socialement responsable », fonds « éthiques », etc.), les entreprises concernées fournissent d'elles-mêmes les chiffres les plus caractéristiques relatifs à leur personnel : pyramide des âges, types de contrats, recrutement, mixité, intégration des personnes handicapées, formation, etc. Lorsque le groupe est international, les indications relatives à la France et à l'étranger sont dissociées, ce qui explique les cooccurrences mises en évidence dans le tableau. On a donc ici une utilisation du cadre géographique dans un contexte nouveau, celui de la présentation des ressources humaines du groupe bancaire.

Les problématiques liées au développement durable fournissent dans les dernières parties du corpus une autre série de co-occurrences de « France » associés eux aussi à une nouvelle utilisation de la géographie. Ce groupe rassemble les formes « Fondation », « Association », « Loire », « Centre », « Régionales », « patrimoine », « Pays », « Maires », « 1979 », « Nord ». Ces formes isolées appartiennent aux dénominations propres de différentes sortes de personnes morales. Il s'agit de fondations (« Fondation du Crédit Agricole Pays de France », « Fondation Solidarité Mutualiste », « Fondation de France »), d'associations (« ADIE - Association pour le Droit à l'Initiative Economique », « France Active », « France Initiative Réseau », « Association des Maires de France »), ou de banques régionales (« Loire Haute-Loire », « Centre-Loire », « Centre France », « Centre-Est », « Nord de France », « Caisses Régionales »). Le point commun qui justifie le rassemblement de ces dénominations d'entités, outre l'inclusion fréquente de « France » dans leurs noms, réside dans le fait que la restitution de leur activité sert à illustrer les actions du groupe bancaire dans le domaine du développement local, forme du développement durable enracinée dans les terroirs.

Alors que les Caisses régionales sont généralement présentées de façon collective dans les CDS et qu'elles font l'objet d'une fiche signalétique dans les CDN, dans les dernières parties du corpus certaines caisses sont mentionnées individuellement car elles financent ou participent à des financements à vocation sociale ou culturelle, en partenariat avec certaines associations (celles dont les dénominations ont été mentionnées plus haut). Cette timide manifestation contraste

avec l'implication du groupe bancaire dans la vie locale qui ne repose pas seulement sur différentes formes de mécénat, comme la publications d'ouvrages en collaboration avec l'Association des maires de France, le financement d'écomusées ou la rénovation du patrimoine rural à travers la fondation « Pays de France », créée en 1979. En effet, l'organisation mutualiste du Crédit Agricole suppose une participation de fait à la vie locale, avec plus de 2500 caisses locales dans lesquelles sont répartis près de six millions de sociétaires¹. C'est pourquoi on peut s'étonner que cette forme particulière d'organisation ne soit pas mieux mise en avant. Par ailleurs, l'argument selon lequel le traitement du fonctionnement institutionnel du Crédit agricole à l'échelle locale serait en décalage avec une restitution générale de l'activité dans les rapports ne peut pas être invoqué, car il est contredit par les exemples qui viennent d'être évoqués. Il faut donc chercher ailleurs l'explication de cette relative discrétion. Il semble que celle-ci puisse être imputée à la difficulté qu'il y aurait à faire état dans les rapports d'activité de la coexistence de modèles économiques différents, le mutualisme et le capitalisme, au sein du même groupe bancaire. En effet, ce genre de discours économique, comme cela a été indiqué dans la première partie, n'est pas destiné à évoquer les sujets qui font débat, que ce soit à l'intérieur ou à l'extérieur de l'institution².

Il semble cependant que les débats relatifs à la place du local dans le modèle de développement du groupe bancaire, que ce soit du point de vue institutionnel, économique, sociétal ou communicationnel, aient laissé des traces dans les discours. Celles-ci sont ténues, mais elles semblent néanmoins affleurer dans les variations de fréquences affectant certaines unités. Ces variations, comme on va le voir, sont dans deux cas provoquées par l'apparition soudaine de la majuscule sur l'un des composants d'une dénomination complexe. En fonctionnant comme un indicateur de valorisation de la réalité désignée, la majuscule révèle une modification de point de vue. On propose d'interpréter ce changement « minuscule » comme une traduction indirecte des tensions nées autour des débats relatifs à la cotation du groupe

¹ Rapport d'activité 2006.

² En effet, le groupe bancaire a tenté de concilier les deux modèles dans une construction originale qui avait soulevé une certaine incompréhension parmi les analystes financiers au moment de la cotation de Crédit Agricole s.a. en 2001. On renvoie sur ce point aux chapitres 1 et 3.

mutualiste. Afin d'illustrer ce point, on a présenté dans le graphique ci-dessous les fréquences d'unités entretenant un rapport étroit avec cette thématique.

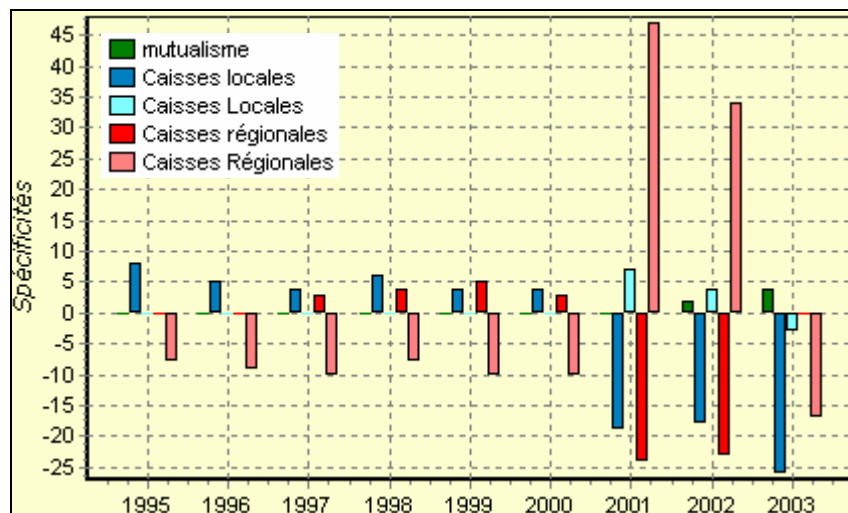


Figure 77.

Ventilation des formes et groupes de formes¹ « mutualisme », « Caisses locales », « proximité », « Caisses Régionales » et « Caisses régionales »

On constate tout d'abord que le mutualisme institutionnel, surtout rappelé jusqu'en 2000 par le décompte des caisses locales rattachées à chaque Caisse régionale dans les contextes non syntactiques, disparaît en 2001. De façon simultanée, le terme fait son entrée dans les contextes syntactiques, gratifié d'une double majuscule en 2001 et 2002. C'est la même double majuscule qui affecte le terme « Caisses régionales » au même moment, et l'on voit également en 2002 les noms des Caisses régionales disparaître des CDN et faire leur entrée, pour certains d'entre eux, dans les CDS. Tout se passe comme si l'émetteur avait cherché à compenser cet escamotage par une valorisation graphique à l'aide de la majuscule appliquée à « locales » et à « régionales », la lettre étant en quelque sorte chargée d'exprimer ce que l'esprit devait taire.

C'est une opération alchimique similaire de transformation de la valeur qui a été pratiquée sur les mots « proximité » et ceux de la famille de « mutualisme ». On

¹ Le groupe de formes « mutualisme » reprend les formes « mutualiste » (15 occ.), « mutualistes » (9 occ.), « mutualisme » (3 occ.) ; segments répétés « Caisses locales » (372 occ.) et « Caisses Locales » (17 occ.) ; forme « proximité » (199 occ.), segment « Caisses Régionales » (176 occ.) et segment « Caisses régionales » (478 occ.).

a vu que pour le premier, cette proximité est principalement utilisée afin d'évoquer un type de marché bancaire, la banque de proximité (les segments « Banque de proximité » et « banque de proximité » sont les plus fréquents parmi les inclusions de « proximité », et ils représentent une centaine d'occurrences). Ce terme, qui a pour équivalents « banque de détail » ou « *retail banking* » en anglais, est une réutilisation marketing du concept de « proximité » à des fins de séduction de la clientèle. L'apparition, dans les dernières parties du corpus, de formes appartenant au groupe « mutualisme », et plus particulièrement la forme « mutualistes » dont sept occurrences sur neuf sont concentrées en 2003, est également révélatrice. En effet, dans cette dernière partie du corpus il est principalement question de « valeurs mutualistes », présentées comme une sorte de preuve venant renforcer la légitimité des actions du groupe bancaire en matière de développement durable. Autrement dit, il semble que l'émetteur des rapports d'activité ait perçu que la réalité mutualiste transformée en « valeurs mutualistes » avait atteint un degré d'abstraction qui la rendait audible pour le marché. Sur ce point, comme sur d'autres présentés précédemment, les tensions perceptibles dans la géographie du groupe bancaire sont organisées en fonction des deux pôles, international et local.

Les toponymes appartiennent à un système de référence qui n'est pas propre à un référentiel particulier. Par ailleurs, la fonction de ce système de référence repose sur une stabilité adaptée pour fournir un cadre statique et relativement bien partagé par les interlocuteurs. Ces deux propriétés sont de nature à conférer une certaine neutralité aux toponymes dans la restitution de l'activité d'un groupe bancaire. Dès lors, on peut se demander si l'étude de leur fonctionnement discursif n'est pas condamnée à n'apporter que des informations de deuxième ordre dans le cadre d'une collecte, telle qu'elle a été envisagée au début de ce travail. Les observations qui ont pu être réalisées dans ce chapitre tendent à montrer l'inverse, non seulement en ce qui concerne l'étude de cette classe de dénominations propres prise en elle-même, mais aussi relativement aux questions soulevées par une collecte réalisée à partir des toponymes, afin de collecter d'autres éléments appartenant au vocabulaire du corpus.

En effet, si les toponymes appartiennent bien à un système de référence général, leur étude en corpus vient rappeler qu'un référentiel spécifique se définit d'abord par la sélection particulière de certains des éléments appartenant à un plus vaste ensemble. Par ailleurs, tous les toponymes sélectionnés ne sont pas dotés de la même importance. Enfin, il faut aussi constater que dans le cadre d'une même série textuelle et à l'intérieur d'une même catégorie de toponymes, les utilisations sont susceptibles de varier de façon importante. Il faut donc admettre qu'en ce qui concerne cette catégorie d'unités, la stabilité du système de référence auquel ils appartiennent contraste fortement avec la diversité des usages constatés en corpus. On a ainsi pu montrer que l'évaluation de l'intérêt que présentent les toponymes par rapport aux buts que l'on s'est fixés doit passer par la compréhension des rôles qu'ils jouent dans les textes du corpus.

Afin de rendre compte de cette diversité des usages, on a été conduit à regrouper les toponymes en catégories, dont trois d'entre elles ont fait l'objet d'une étude détaillée : les odonymes, les noms de villes et les noms de pays. Ce découpage conventionnel n'a pas empêché l'adoption d'un cadre d'analyse commun reposant sur les données textométriques disponibles. Celles-ci correspondent à la fréquence et la récurrence de la dénomination et de ses variantes, à la répartition de leurs

occurrences entre les deux principales sortes de contextes identifiées dans le corpus (contextes à dominante syntactique et non syntactique) et, enfin, au voisinage du toponyme sur les axes syntagmatique (segments répétés) et paradigmatique (co-occurents). La combinaison des indications fournies par ces quatre critères a permis d'identifier des rôles qui correspondent à autant d'utilisations d'une catégorie de toponymes.

On peut illustrer ce point à l'aide d'un exemple simple emprunté à la catégorie des odonymes. Pour une série importante d'odonymes on a pu constater une fréquence faible (inférieure ou égale à 9 occurrences), accompagnée d'une récurrence forte (les occurrences sont généralement réparties sur plusieurs parties consécutives dans le corpus), ainsi qu'une répartition majoritaire dans des contextes à dominante non syntactique. Le calcul des co-occurents a permis d'identifier que le voisinage des odonymes est essentiellement constitué par des unités apparaissant ordinairement dans des adresses postales. Le rôle des odonymes dans les rapports d'activité consiste donc principalement à fournir des repères spatiaux dans un dispositif textuel particulier de type « coordonnées ». D'autre part, le fait de disposer d'une série textuelle chronologique a permis de mettre en évidence que ce dispositif ne pouvait pas être réduit à une simple fonction utilitaire de type « informations pratiques ». En effet, cette explication fonctionnelle n'est pas capable à elle seule de rendre compte de variations de fréquence mises en évidence en ce qui concerne certains odonymes. Celles-ci doivent d'abord être rapportées à des modifications du dispositif textuel de type « coordonnées » devant à leur tour être interprétées en fonction des changements importants ayant affecté la géographie de l'activité du groupe bancaire, et donc de son référentiel. De façon plus générale, il ressort de ces analyses que la nature des relations entre référentiel et unités toponymiques doit être établie à partir de la compréhension de la manière dont le discours accomplit telle ou telle visée communicationnelle particulière.

Cette visée intervient dans les modifications du dispositif textuel et elle trouve son origine dans les changements affectant le champ de la situation de communication qui sert de cadre à la production des rapports d'activité. On vient de voir que les « renseignements pratiques » relèvent d'une catégorie de sens commun dont le fonctionnement est déterminé, au-delà des motifs utilitaires, par la manière

de restituer l'activité bancaire qui est jugée comme étant la mieux adaptée par l'énonciateur. On a ainsi pu voir disparaître un « effet international » qui n'avait plus de raison d'être à partir du moment où le groupe bancaire avait acquis une réelle dimension internationale. De même, lorsque le centre de gravité de l'activité restituée se déplace du groupe Crédit Agricole, au groupe Crédit Agricole s.a., les coordonnées des Caisses régionales peuvent apparaître un temps superflues à l'énonciateur, qui devra néanmoins revenir sur son choix. Ce qui vaut pour l'escamotage de certaines rubriques des rapports d'activité semble aussi pouvoir s'appliquer à l'introduction de nouveaux développements. Ainsi, l'apparition au début des années 2000 des problématiques du développement durable dans les rapports d'activité des principales banques françaises s'est-elle traduite, dans les rapports d'activité du Crédit Agricole, par la mise en valeur de certains aspects de son action économique au niveau local (en France). Cet effet de mode affectant le genre de discours « rapports d'activité » est pour une part responsable de l'augmentation de la fréquence de certains toponymes présents dans le corpus.

Que les toponymes fonctionnent aussi comme indices de modifications intervenant au niveau de la situation de communication ne doit pas faire perdre de vue le rôle important qu'ils jouent dans l'organisation des discours. En effet, on a pu constater que les noms de pays servent à organiser les entités du groupe bancaire en fonction de leur cadre d'activité économique. D'une manière plus générale, ils permettent de structurer la restitution de l'activité du groupe à l'étranger. La plupart des formes d'activité sont susceptibles de recevoir deux présentations complémentaires : ce qui s'est passé en France et ce qui concerne tel ou tel pays. Cette organisation de la présentation de l'activité du groupe se fait de plus en plus prégnante au fur et à mesure que l'on avance dans la série chronologique. Les manifestations de la dimension internationale de l'activité s'introduisent même à l'intérieur des dénominations, qu'il s'agisse de termes, comme « banque de détail à l'étranger », ou de dénominations propres, comme « Crédit Agricole Indosuez Suisse s.a. ». Les toponymes assument donc un deuxième rôle dans les rapports d'activité, qui consiste à structurer la restitution de l'activité bancaire. Cette contribution se manifeste à deux niveaux au moins : au niveau textuel, par

l'organisation de la restitution de l'information, et au niveau des dénominations, en réduisant l'extension de certains noms de secteurs d'activité.

Si le recours à un découpage géographique de l'activité se développe durant la période couverte par le corpus, c'est aussi que les noms de pays, qui constituent l'unité principalement utilisée dans ce cas, subissent une transformation. De simples repères territoriaux, fournissant un cadre pour présenter le développement de l'activité à l'étranger, ils ont de plus en plus tendance à être utilisés pour désigner des marchés, au sens de secteurs économiques. Cette évolution est particulièrement sensible pour « France » qui, de référence implicite en début de période, va être opposée en termes de cadre d'expansion économique à « monde », puis à « Europe ». Si dans les deux cas il s'agit d'incarner le développement du groupe bancaire à l'étranger, l'opposition « France-monde », progressivement remplacée par « France-Europe », marque le passage d'un développement économique souhaité à une réalité en train de se mettre en place. Là encore, l'utilisation d'une série textuelle chronologique permet de suivre cette évolution de la géographie de l'activité économique du groupe bancaire, telle que les rapports d'activité la restituent. De ce point de vue, la position occupée par « France » dans la figure 76 ne peut s'expliquer que si le nom de pays sert aussi à désigner un espace économique particulier, sorte de camp de base utilisé pour la conquête de nouveaux marchés à l'échelle de la planète. Un troisième rôle a donc été identifié pour certains toponymes présents dans le corpus, les pays-marchés.

On peut tirer deux enseignements majeurs de l'étude du fonctionnement discursif des toponymes pour la collecte que l'on souhaite réaliser. Le premier réside dans l'éclairage que les toponymes apportent sur l'évolution du référentiel du groupe bancaire dont les textes du corpus présentent l'activité. Sur une période d'une dizaine d'années seulement, ils illustrent le passage d'un développement économique reposant sur une géographie fantasmée, à un recentrage assumé sur la France et une Europe élargie, permettant de maintenir une présence significative sur les principales places financières mondiales. L'étude des toponymes présente donc un intérêt en soi, dès lors que l'on cherche à saisir les évolutions d'un référentiel à travers le fonctionnement discursif d'une classe de dénominations propres. Mais l'enseignement majeur pour la collecte que l'on souhaite réaliser réside dans la mise

au jour de l'importance des rôles que les dénominations propres sont susceptibles de revêtir dans les discours, soit individuellement, soit collectivement, lorsque le rôle caractérise l'usage d'une classe entière. Ces rôles ont été définis comme la manifestation, pour un genre de discours, d'une pragmatique propre à l'énonciateur, relativement à une ou à plusieurs classes de dénominations propres. Ils sont déterminés non seulement par le genre de discours dans lequel ils apparaissent, mais aussi par l'usage communicationnel qui est fait des contextes dans lesquels les dénominations sont présentes. Ce faisant, ils permettent de caractériser la manière dont ces unités sont intégrées aux discours à travers les relations qu'elles entretiennent avec les autres composantes du vocabulaire.

L'étude du fonctionnement discursif des toponymes a ainsi permis de dégager une gradation parmi les rôles endossés par une même unité, ou réservés à une série entière. Les odonymes caractérisent un premier palier dans lequel la dénomination propre est seulement mentionnée dans un dispositif récurrent et clos sur lui-même, les adresses postales. Dès lors que le toponyme est utilisé dans un contexte de type syntactique, certaines caractéristiques de son référent peuvent être sélectionnées. La catégorisation, en venant préciser explicitement les propriétés du référent sélectionnées en rapport avec le propos, réalise une opération d'incorporation de l'unité au reste du discours. Les noms de villes (places boursières) et de pays (marchés) bénéficient de ce traitement, quoique de façon non systématique. Enfin, seuls les noms de pays sont susceptibles de revêtir un troisième rôle, en plus des précédents, qui consiste à ouvrir des univers de discours servant de cadre à la restitution de l'activité bancaire. Alors que les deux premiers rôles endossés par les toponymes dans les textes du corpus ne laissent pas entrevoir une productivité importante pour la collecte, il n'en va pas de même pour le rôle d'organisateur de la restitution d'une partie au moins de l'activité bancaire. Il conviendra donc de tenir compte de cette dernière utilisation et d'en proposer une évaluation dans la perspective de réaliser une collecte en prise avec le vocabulaire du corpus.

CHAPITRE 10

CONTENU DES DÉNOMINATIONS PROPRES ET VOCABULAIRE DU CORPUS

Le précédent chapitre a permis de montrer l'intérêt qu'il y a à prendre en compte les principaux rôles attribués aux dénominations propres dans le genre de discours que l'on étudie, ainsi que l'usage communicationnel des dispositifs textuels dans lesquels ces unités apparaissent. L'étude du fonctionnement discursif de la classe des toponymes, réalisée à l'aide des principaux outils offerts par la textométrie (graphique de ventilation des fréquences entre contextes ou parties du corpus, segments répétés, carte des sections d'un texte, calcul des co-occurents et analyse factorielle des correspondances) a rendu possible une première série d'observations. Cependant, dans la mesure où les toponymes constituent probablement la seule classe de dénominations propres à appartenir au référentiel partagé des interlocuteurs, leur utilisation dans les discours se fait largement sur le mode tacite. De ce fait, les relations entretenues avec les autres composantes du vocabulaire s'en trouvent relativement limitées.

D'autre part, il est apparu que lorsqu'elle repose sur l'opposition entre contextes à dominante syntactique et non syntactique, la notion de rôle permet de clarifier l'usage qui est fait des dénominations propres. Les contextes à dominante non syntactique correspondent à une présentation statique du groupe bancaire, qui n'est que faiblement en prise avec une restitution commentée de l'activité. Dans ces contextes (il s'agit, par exemple, de coordonnées postales ou d'organigrammes), les relations entre unités se font essentiellement par des juxtapositions dont le sens, rarement verbalisé, doit être dégagé à partir de certains artifices liés à la mise en page et à l'utilisation de divers codages typographiques. Ces observations laissent penser que l'utilisation de dénominations propres présentes majoritairement dans de tels contextes risque de produire une collecte pauvre, voire circulaire. Il est en effet probable que celle-ci enregistrera principalement des relations entretenues entre dénominations propres¹ au détriment de celles qui pourraient relier ces unités aux autres composantes du vocabulaire².

¹ On a vu dans la première partie que les contextes CDN sont majoritairement composés de mots outils liés à certains dispositifs récurrents et de dénominations propres.

² Cette approche est non seulement possible mais utilisée, par exemple, pour la veille économique. Dans ce cadre, la mise en relation d'un nom de société et d'un nom de personne physique présente un

En revanche, l'apport de l'identification des rôles est moins évident pour ce qui concerne les contextes syntactiques. La présence de deux contextes particuliers offre un éclairage utile, mais elle ne permet pas de préciser finement les sortes de relations établies par une classe de dénominations propres avec les autres composantes du vocabulaire du corpus dans un contexte syntactique. Il est alors difficile d'évaluer les potentialités d'une collecte à partir de telle ou telle classe de dénominations propres. Par ailleurs, la question se pose en des termes différents pour les classes de dénominations propres qui ne font pas a priori partie du référentiel partagé par les interlocuteurs.

Le projet terminographique que représente la mise en place de référentiels terminologiques adaptables au contexte nécessite la prise en compte de la recherche du meilleur rapport entre temps passé pour réaliser la collecte, d'une part, et pertinence de l'information collectée, d'autre part. Même s'il est parfaitement possible de réaliser une collecte avec une ou plusieurs classes de dénominations propres, de façon exhaustive ou sous la forme de sondages, sans tenir compte de ce critère, la mise en œuvre du « cahier des charges » présenté au début de ce travail rend nécessaire d'envisager cet aspect de la collecte, la construction de ressources terminologiques opérationnelles étant généralement coûteuse en temps¹, au point de s'avérer parfois rédhibitoire, dès lors que des mises à jour régulières doivent être réalisées. Par conséquent, on a cherché à dégager dans ce chapitre une vision synthétique des relations que les principales classes de dénominations propres entretiennent avec les autres unités présentes dans les textes du corpus, afin d'être en mesure de sélectionner la ou les classes les plus pertinentes pour l'exploitation du corpus.

L'étude de ces relations soulève un certain nombre de questions. Si l'on admet que de telles relations existent, quelles formes prennent-elles dans les discours ? Quel statut faut-il leur accorder ? Constituent-elles autre chose qu'un

intérêt comme indice d'un futur changement de stratégie, lié au renouvellement de l'équipe de direction. Mais dans ce travail, on a assigné une finalité différente à la collecte d'informations réalisée à partir des dénominations propres. Il s'agit, en effet, de capter certains éléments caractéristiques d'un parler d'entreprise dans une situation de communication donnée.

¹ Ce sujet a été abordé, entre autres, dans D. Gouadec (1993 : 25-34), intervention de E. Lebreton et discussion de la table ronde qui a suivi, et dans C. Wiegandt (1995).

assemblage indéterminé et changeant ? Dans le cas contraire, comment caractériser cet assemblage ? Est-il amené à varier dans le temps ? Est-il formalisable et transposable dans un référentiel terminologique ? En définitive, quel rôle la schématisation de cet assemblage peut-elle être amenée à jouer dans le cadre de la collecte d'informations ? Voilà autant de questions auxquelles on va tenter de répondre dans ce chapitre et dans le suivant, qui constituera une mise en œuvre des principes que l'on va chercher à définir maintenant.

On commencera par s'interroger, à la suite des observations empiriques réalisées dans le précédent chapitre, sur la nature du fonctionnement discursif des dénominations propres. On sera alors conduit à caractériser le contenu des dénominations propres, entité sémantique particulière produite dans et par les discours. Il conviendra alors de caractériser la structure type du contenu des principales classes de dénominations propres présentes dans le corpus : dénominations de personnes morales, toponymes, anthroponymes et dénominations d'objets. On exposera la méthode employée afin de construire les schémas correspondant à ces structures. C'est à partir de ces observations que l'on proposera les orientations que devra suivre la réalisation de la collecte de deuxième niveau, en sélectionnant les classes et les relations jugées pertinentes en vue d'une réutilisation des données collectées à des fins d'organisation de l'information.

I. LE FONCTIONNEMENT DISCURSIF DES DÉNOMINATIONS PROPRES

A. Une candidate possible au thème de discours ?

Afin de d'aborder les questions relatives au fonctionnement discursif des dénominations propres, on commencera par examiner la manière dont ce sujet a déjà été envisagé selon deux approches différentes : les thèmes de discours et les objets de discours. On s'interrogera ensuite sur le statut qu'il convient d'accorder aux informations qui sont associées dans les discours aux dénominations propres.

Dans la mesure où l'on cherche à capter, à partir des dénominations propres, la manière dont une partie des informations contenues dans les textes est organisée, il convient de préciser en quoi cette approche diffère de celle des « termes-pivots ». Les « termes-pivots » sont des unités linguistiques utilisées comme mots-clés pour procéder à une analyse de discours. Inspirée des travaux de Z.S. Harris¹, l'approche française de l'analyse de discours a fait l'objet de critiques recensées par D. Maingueneau². L'utilisation de propositions décontextualisées (il s'agit des contextes situés immédiatement à gauche et à droite de la forme pivot), s'avère insuffisante dès lors qu'il s'agit de caractériser le contexte réel d'une unité. Par ailleurs, l'élimination des traces laissées par les opérations énonciatives (du fait des transformations opérées sur les contextes pour les réduire à une forme de proposition simple), constitue une perte d'information importante. Enfin, cette approche assimile le texte à une donnée compacte émise sur la même longueur d'onde, ce qui ne correspond pas la réalité observable. On a vu dans les développements qui précèdent que des réponses ont déjà été apportées à ces critiques : le recours au texte d'origine est pratiqué autant que nécessaire pour préciser le fonctionnement d'une unité ; la complexité native de ces contextes est conservée pour l'analyse ; enfin, le texte a

¹ Z.S. Harris (1969).

² D. Maingueneau (1991 : 81 et suiv.) et *Dictionnaire de l'analyse de discours* (2002 : 378-379).

d'abord été considéré comme un discours doté d'une autonomie de fonctionnement susceptible d'évoluer dans le temps.

Cependant, ces points ne concernent pas la critique majeure adressée à l'approche des termes-pivots : celle de reposer sur un savoir a priori utilisé pour sélectionner les termes qui vont ensuite guider l'analyse. Lorsqu'elle est rapportée à la problématique présente dans cette étude, une telle critique peut être décomposée en deux points : d'une part, les dénominations propres constituent une classe d'unités déterminée a priori, d'autre part, les dénominations propres sont posées, également a priori, comme thèmes de discours. Pour ce qui est du premier point, on a tenté de montrer, dans la première partie, que la sélection des noms propres comme unités utilisées pour explorer les textes correspondait à deux hypothèses qui devaient être vérifiées. La première concerne les relations susceptibles d'être établies entre un référentiel spécifique et les noms propres qu'il incorpore, ainsi que l'évaluation de la présence significative des noms propres dans les textes d'un corpus. En ce qui concerne le premier point, on a considéré que l'étude du fonctionnement pragmatique des noms propres apportait des éléments de réponse, certes incomplets, mais suffisants pour poursuivre les investigations dans le cadre de ce travail. Par ailleurs, les deux précédents chapitres ont permis d'évaluer de façon positive la présence des dénominations propres dans le corpus. La seconde hypothèse porte sur la productivité d'une collecte d'informations réalisée à partir des noms propres présents dans les textes d'un corpus. C'est l'objet de ce chapitre et du suivant que d'apprécier ce point. Le choix des noms propres est donc bien réalisé a priori, mais dans le cadre d'hypothèses devant être vérifiées, ce qui ne peut être assimilé à une forme de résultat acquis par anticipation. En procédant de cette manière, il n'est aucunement nécessaire d'assimiler les noms propres, ou dénominations propres, à des thèmes de discours. Il s'agit plutôt d'interroger les implications d'une telle assimilation.

Les relations pouvant être établies entre thème de discours et dénominations propres seront examinées à partir de la notion de contenu descriptif d'un thème de

discours, telle qu'elle a été discutée par J.-M. Marandin¹. Cet auteur propose une définition du contenu descriptif d'un thème de discours comme résultat d'un processus discursif :

« (i) un thème est un individu composite (descriptivement (sic) : un individu relativement à un texte) ;

(ii) le CD [contenu descriptif] du terme qui le nomme est représentable comme un agrégat subsumant d'autres individus dans leurs interrelations, telles qu'elles sont introduites dans les énoncés, reconstruites dans la compréhension et constitutives d'une interprétation.

Dans le cas d'un thème configuré :

(iii) le nom du thème est le SN en position de fermoir d'une chaîne-objet² ;

(iv) les composants du contenu descriptif sont les segments de la chaîne-objet.

Dans le cas d'un thème inféré :

(iii) aucune contrainte textuelle ne pèse sur le nom du thème ;

(iv) les composants du CD sont abstraits de la suite d'énoncés.

Dans les deux cas :

(v) la tête nominale du SN projette une interprétation sur les composants du CD³. »

La distinction introduite par J.-M. Marandin entre les deux sortes de thèmes tient à l'espace de donation du nom : « le texte pour le thème configuré, la réception pour le thème inféré⁴. »

Pour illustrer son propos, J.-M. Marandin a forgé un texte très court dans lequel apparaît le personnage de Lancelot. La dénomination propre est présentée comme une candidate possible pour dénommer le thème du discours dans différentes variantes du texte. Il reste que dans les deux cas de figure du thème configuré et du thème inféré, le choix du nom du thème de discours relève d'un processus qui est à la charge de l'interlocuteur. Même si l'existence de configurations discursives

¹ J.-M. Marandin (1988).

² La chaîne-objet ou la chaîne de référence correspond à une suite d'items renvoyant à un même référent. Le « fermoir » correspond à un syntagme nominal qui a la capacité de condenser la chaîne. On a vu plus haut que pour C. Schnedecker le fermoir correspond à la reprise du même nom propre, ce qui n'est qu'un cas de figure possible dans la perspective de J.-M. Marandin qui admet que ce rôle soit tenu également par un syntagme nominal démonstratif. *Ibid.* (1988 : 74).

³ J.-M. Marandin (1988 : 82).

⁴ *Ibid.* (1988 : 77).

permet de pointer certaines unités, il ne s'agit que de propositions destinées à guider la lecture. Celles-ci ne peuvent ni se substituer aux processus de compréhension et d'interprétation, ni sélectionner sans arbitraire l'unité ou les unités censées être les plus représentatives des thèmes présents dans un texte. En effet, comme l'affirme également F. Rastier, il y a une différence de nature entre les unités lexicales et les thèmes : les premières relèvent des signes alors que les seconds correspondent à des unités de contenu. Il ne peut donc y avoir que des lexicalisations privilégiées pour certains thèmes, et non pas de correspondance pré-établie entre lexème et thème¹.

On a vu plus haut que la démarche adoptée pour la collecte ne requiert pas que les dénominations propres fonctionnent également comme noms de thèmes de discours. On vient de voir que ce statut n'est pas donné, au sens où certaines unités seraient désignées explicitement comme dénominations de thèmes, mais qu'il résulte d'une double construction de la part du locuteur et de l'interlocuteur. En revanche, la notion de thème configuré suggère que le travail d'élaboration thématique produit par l'énonciateur laisse des traces dans l'organisation du discours. L'approche communicationnelle proposée dans le cadre de la logique naturelle semble fournir des éléments susceptibles d'approfondir ce point.

B. Un moyen d'accéder aux objets de discours

L'approche communicationnelle développée par J.-B. Grize² dans le cadre de la logique naturelle propose de concevoir l'interaction verbale comme une schématisation : « *Une schématisation a pour rôle de faire voir quelque chose à quelqu'un, plus précisément, c'est une représentation discursive orientée vers un destinataire de ce que son auteur conçoit ou imagine d'une certaine réalité*³. » Cet objet discursif, à la fois processus et résultat⁴, fait l'objet d'une double construction de la part du locuteur et du lecteur. C'est pour étudier la genèse de la schématisation

¹ F. Rastier (2001 : 191 et suiv.).

² J.-B. Grize (1990 et 1996).

³ *Ibid.* (1996 : 50).

⁴ « *Le concept de base est celui de schématisation. Le terme lui-même, comme toute nominalisation, renvoie à la fois à un processus et à son résultat.* » *Ibid.* (1996 : 69), voir également (1996 : 143).

à travers les traces que sa production a laissées dans le texte, que J.-B. Grize a proposé un ensemble d'opérations logico-discursives.

Une schématisation repose dans le discours sur la présence d'objets complexes que sont les classes-objets. Celles-ci sont composées d'objets de discours et d'ingrédients. L'objet de discours est d'abord un objet de pensée qui se présente sous la forme d'une notion primitive dont l'existence est postulée. Selon cette conception, le discours est ancré dans une notion primitive¹ à l'aide de deux opérations logico-discursives distinctes. L'opération α appliquée à une notion primitive va engendrer un nom d'objet, tandis que l'opération η va produire un couple prédicatif. On a, par exemple, α (/HERBE/) \rightarrow {foin} et η (/HERBE/) \rightarrow \pm Faucher, qui se lit : l'opération α appliquée à la notion primitive /HERBE/ a permis l'extraction du nom « foin » (mis entre accolades parce qu'il initie une classe-objet), et l'opération η appliquée à la même notion donne le couple prédicatif faucher / ne pas faucher. Lorsque l'opération α a été exécutée, il en résulte l'apparition dans le discours d'un nom correspondant à l'objet. Ce mot est introduit dans le discours avec un faisceau d'aspects qui lui sont associés en raison de sa signification, mais aussi des caractéristiques du référent qu'il permet de désigner, et de manière plus générale, de savoirs plus ou moins partagés par les interlocuteurs. Ces aspects comprennent des propriétés, des relations ainsi que des schèmes d'action². La classe initiée par l'objet de discours s'enrichit ensuite de nouveaux ingrédients provenant des faisceaux de l'objet. Ceux-ci sont introduits dans le discours par l'intermédiaire de nouvelles opérations logico-discursives.

On se limitera ici à un rappel succinct des opérations susceptibles d'accroître les ingrédients d'une classe objet, ou de donner naissance à une nouvelle classe. Les opérations sur les faisceaux font apparaître dans le discours des éléments appartenant habituellement aux faisceaux de l'objet. Par ailleurs, les opérations sur

¹ « Je dirais donc que les notions primitives « sont des systèmes de représentations complexes de propriétés physico-culturelles, c'est-à-dire des propriétés d'objet issues de manipulations nécessairement prises à l'intérieur de cultures » (Culioli, 1981, p.65). (...) Cela n'empêche toutefois pas de concevoir que les notions primitives ont quelque chose comme « des noyaux conceptuels qui relèvent plutôt de l'usage qu'une culture fait d'un lexème et de la vision pré-scientifique ou scientifique que cette culture a de l'unité culturelle à laquelle le lexème se réfère » (Eco, 1972, p. 98-99). » J.-B. Grize (1996 : 82-83).

² *Ibid.* (1996 : 83-84).

les domaines sollicitent des éléments révélant certains aspects de l'objet lorsque celui-ci est présenté dans tel ou tel domaine. Ces opérations, qui s'appliquent à l'objet, sont complétées par une série d'opérations portant sur son nom, les opérations de reprise. Enfin, deux autres séries d'opérations permettent à l'intérieur d'un même discours d'engendrer une nouvelle classe à partir d'une prédication ou d'un énoncé antérieurs ; elles correspondent, dans une conception étendue de l'anaphore nominale, à l'anaphore associative pour la première et à l'anaphore conceptuelle, ou résomptive, pour la seconde¹.

Dans la perspective de la logique naturelle, la schématisation occupe un rôle central, car la mise au jour de sa genèse vise à objectiver les bases sur lesquelles reposent l'interprétation et la compréhension d'un texte. Pour la collecte d'informations, cette compréhension globale n'est pas le but recherché, ce dernier consistant dans le recensement de certaines informations contenues dans le texte et se rattachant aux dénominations propres. La portée de la collecte ne s'étend donc pas au texte entier, mais se limite à certains de ses sous-ensembles. Ces derniers correspondent à des agrégats ayant pour centre, ou plutôt pour tête, un objet de discours. C'est pourquoi la notion de classe-objet constitue une représentation assez proche de ce que l'on cherche à caractériser. Comme le contenu discursif d'un thème de discours, celle-ci est construite dans et par le discours. Mais à la différence du thème, elle ne véhicule pas de problématique liée à l'identification d'une quelconque hiérarchie. Au contraire, pour J.-B. Grize, toutes les classes paraissent se valoir, dans la mesure où, comme il l'affirme lui-même, il « *est indéniable qu'il y a une grande part d'arbitraire dans le choix des classes-objets*². » La question n'est donc plus d'identifier des thèmes principaux ou secondaires, mais d'accéder aux classes objets construites dans le discours.

C'est un autre apport de la notion de classe-objet que de spécifier que ce rassemblement d'ingrédients est organisé, même si cette organisation n'a pas le caractère formel des classes distributives, mais se rapproche plutôt des classes

¹ On renvoie sur ce point à M. Riegel, J.-C. Pellet et R. Rioul (2004 : 614-615).

² J.-B. Grize (1996 : 136).

méréologiques¹, sans en posséder toutes les propriétés. Ce faisant, la recherche des ingrédients dont une telle classe se compose est censée donner accès à la façon dont le locuteur a cherché à élaborer son discours autour d'un objet donné. Il ne s'agit donc pas d'une organisation plaquée de l'extérieur, mais de la restitution d'une élaboration discursive. Ce faisant, la classe-objet est à même de capter des agrégats d'unités dont l'organisation a été abordée par d'autres approches comme les paradigmes désignationnels, présentés dans la deuxième partie, ainsi que les chaînes de référence, ou encore le contenu descriptif des thèmes de discours configurés.

Enfin, on notera que la classe-objet peut être « remplie » de deux manières différentes. La première est propre à la logique naturelle, qui est moins centrée sur le matériau linguistique que sur les mécanismes cognitifs qui sous-tendent les discours. Cela a pour conséquence de reléguer au second plan la spécificité des unités susceptibles d'être collectées. Ainsi, lorsqu'il présente l'analyse des schématisations d'un corpus de textes journalistiques, J.-B. Grize est-il amené à faire la remarque suivante : « *D'autre part, chaque journaliste a son style propre et, à côté de quelques expressions identiques – sans doute celles des agences de presse – les notions sont souvent exprimées par des lexèmes différents. Comme mon propos n'est pas de nature linguistique mais qu'il est cognitif, je me servirai de mots relativement neutres qui parfois ne figurent dans aucun des articles*². » Dans l'approche retenue pour ce travail, on considère que les unités linguistiques présentent un intérêt en soi, puisqu'elles sont destinées à fournir les matériaux d'un référentiel terminologique. Cette deuxième manière d'aborder la classe-objet, dans laquelle les unités linguistiques originales sont conservées telles quelles, a déjà été mise à profit dans d'autres travaux reprenant à leur compte la notion d'objet de discours³.

Même si on a adopté une conception de la classe-objet respectueuse des mots réellement utilisés dans le discours, il reste que cette notion oblige à considérer qu'une dénomination propre est susceptible d'occuper deux places : celle de nom de

¹ J.-B. Grize (1984 : 209 et suiv.). Voir la contribution de Denis Miéville intitulée « Logique naturelle et méréologie ».

² J.-B. Grize (1996 : 132).

³ On peut citer, à titre d'exemple, L. Mondada (1994 : 61 et suiv.), ou F. Sitri (2003 [1998]).

l'objet ou celle d'un ingrédient quelconque rattaché à l'objet ayant initié une classe. Ces deux positions soulèvent deux sortes de questions qu'il s'agit de mentionner maintenant, et auxquelles on tentera d'apporter une réponse dans les sections suivantes. Tout d'abord, lorsqu'une dénomination propre occupe la place de l'objet initiant une classe, cela suppose une opération d'ancrage préalable. Cet ancrage se faisant dans une notion primitive, il est nécessaire de se demander à quoi peut correspondre une telle notion dans le cas des dénominations propres¹. Autrement dit, il convient d'examiner la compatibilité d'une telle conception avec l'option relative au sens des dénominations propres que l'on a retenue dans ce travail. En second lieu, lorsqu'une dénomination propre se trouve appartenir aux ingrédients d'une classe-objet, il est nécessaire de s'interroger sur la portée d'une telle classe. Plus précisément, cela semble poser, avec plus d'acuité que dans le cas précédent, la question des limites de la classe et des sortes d'unités qu'il devient possible de capter à partir d'une dénomination propre « simple ingrédient ».

C. Le contenu de la dénomination propre

Une autre manière d'aborder le fonctionnement discursif des noms propres consiste à examiner cette question d'un point de vue sémantique. Dans le souci d'établir un sens linguistique pour le nom propre, M.-N. Gary Prieur a été amenée à distinguer nettement deux plans. Le sens du nom propre, comme on l'a vu plus haut, peut être limité à l'instruction d'aller chercher un référent précis dans la mémoire stable des interlocuteurs². Il ne comporte sur ce premier plan aucun élément

¹ On note que dans les exemples réels fournis par J.-B. Grize 1990 et 1996, les dénominations propres ne sont pas absentes mais ne servent pas non plus d'objet initiant une classe, ce rôle étant toujours dévolu à des lexèmes ou des termes. On trouve en revanche des exemples construits dans lesquels l'utilisation d'une dénomination propre dans cette position ne semble pas faire problème : « *Ainsi dans : Jean écrit une lettre, on aurait $\alpha_1 \rightarrow \{Jean\}$, $\alpha_2 \rightarrow \{lettre\}$ et $\eta \rightarrow \pm Ecrire$.* » J.-B. Grize (1996 : 87). Dans l'approche proposée par J.-M. Marandin, le nom propre de personne « Lancelot » est susceptible de constituer le thème d'un petit texte forgé à titre d'exemple. Il note cependant que la compréhension d'un thème dans la perspective de l'intelligence artificielle met en évidence le choix souvent implicite en faveur d'un mode catégorial de la compréhension qui « *interdit de prendre un nom propre comme thème (...)* ». J.-M. Marandin (1988 : 84-85).

² On peut donner également la position de K. Jonasson : « *Dans les sections précédentes nous avons essayé de rendre compte de divers types de connaissances au sujet du porteur d'un Npr, auxquelles il faut avoir recours pour bien comprendre un énoncé comportant le Npr considéré. Nous avons supposé qu'une chaîne causale vient associer le Npr directement à un particulier en vertu d'un lien*

susceptible de décrire le référent désigné. Cette description est portée par une composante d'une autre nature, le contenu du nom propre. Celui-ci est défini comme « (...) un ensemble de propriétés attribuées au référent initial de ce nom propre dans un univers de croyance¹. » On présentera tout d'abord le contexte d'élaboration de la notion de contenu du nom propre pour en proposer ensuite une extension.

La notion de contenu du nom propre a pour cadre les énoncés et repose sur la notion de référent initial : « *Le référent initial d'un nom propre dans un énoncé est l'individu associé par une présupposition à cette occurrence du nom propre en vertu d'un acte de baptême dont le locuteur et l'interlocuteur ont connaissance*². » Il se trouve cependant un certain nombre de cas pour lesquels le référent visé n'est pas, ou n'est pas directement, le référent initial. Dans des exemples tels que « *Paris, c'est Beyrouth* » ou « *Le Paris de Doisneau* » M.-N. Gary-Prieur considère que l'interprétation de la métaphore, dans le premier cas, ou la compréhension de l'allusion à une représentation de la capitale, dans le second cas, doivent s'appuyer sur certaines propriétés du référent, sélectionnées spécialement par le locuteur. En effet, on perçoit bien que la compréhension de ces énoncés ne dépend pas d'un ensemble de connaissances encyclopédiques standard sur les villes de Beyrouth ou de Paris. Au contraire, les images de ville en ruine ou d'atmosphère bon enfant, si elles prennent bien appui sur des propriétés du référent initial, ne requièrent que les informations utiles à la bonne interprétation de l'énoncé. Elles se distinguent en cela des propriétés encyclopédiques générales, dont elles ne constituent qu'une sélection *ad hoc*, mais aussi des connotations, qui correspondent à des associations reposant sur une exploitation culturelle du signifiant du nom propre.

Lorsque les propriétés du référent sont suffisamment partagées par les interlocuteurs – c'est-à-dire, pour reprendre la formulation de M.-N. Gary Prieur, lorsqu'elles se situent à l'intersection de leurs univers de croyance –, elles font partie

dénominateur, impliquant en même temps l'existence et l'unicité du porteur du Npr. Cette association fournit au Npr un modèle interprétatif comparable à celui des Nc, constituant un savoir non pas général mais spécifique, qui se reflète dans la capacité de fournir la description d'un porteur du Npr instauré dans notre mémoire stable. » (1994 : 150).

¹ M.-N. Gary-Prieur (1994 : 51). La notion d'univers de croyance est empruntée à R. Martin : « on appellera « univers de croyance » ou « univers » l'ensemble indéfini des propositions que le locuteur, au moment où il s'exprime, tient pour vraies ou qu'il peut accréditer comme telles. » (1992 : 38).

² *Ibid.* (1994 : 29).

des présupposés du discours et ne bénéficient donc pas d'une prise en charge énonciative particulière. C'est le traitement que l'on a pu constater pour les toponymes présents dans les textes du corpus. En revanche, lorsqu'elles doivent être sélectionnées dans l'univers de croyance du locuteur, ou dans une image d'univers¹, ces propriétés font généralement l'objet d'un apport d'information explicite, comme dans cet exemple cité par M.-N. Gary-Prieur : « *La ville de Skodor, cette Vendée de l'Albanie, intensément catholique, où la résistance au régime fut toujours vive*² ».

On peut alors se demander s'il n'en va pas de même pour toutes les situations dans lesquelles le locuteur prend en charge les lacunes présumées de l'interlocuteur, en injectant dans le discours les apports cognitifs nécessaires à la réussite de la communication. Or, le risque de déséquilibre entre les connaissances des interlocuteurs peut porter également sur la présupposition de base dans l'usage d'un nom propre, c'est-à-dire lorsque son référent, et non plus l'une quelconque de ses propriétés, n'est pas encore connu par le destinataire³. On propose donc d'élargir le cadre fourni par le contenu discursif des noms propres aux éléments explicitement apportés par le locuteur, afin de renseigner son interlocuteur sur le référent dont il parle. Néanmoins, et compte tenu de ce qui a été avancé précédemment, cet élargissement se fait dans le cadre d'une classe-objet initiée par une dénomination propre ou contenant une dénomination propre au titre d'ingrédient.

En abordant le fonctionnement discursif des dénominations propres, on a été conduit à ne pas retenir la sorte de raccourci qui poserait ce type d'unités comme dénominations de thèmes de discours. En revanche, l'appartenance aux classes-objets et la notion de contenu du nom propre permettent, semble-t-il, une articulation plus productive, entre fonctionnement discursif et collecte d'informations. En effet, ce qui est visé alors, ce sont certains aspects du contenu des dénominations propres

¹ « Au lieu de conférer lui-même à une proposition une valeur de vérité, le locuteur peut aussi la situer dans quelque univers qu'il évoque. On appellera image la représentation d'un univers dans le discours. Il y a image d'univers dès lors que, épistémiquement, le locuteur renvoie, dans son discours, à un univers de croyance. » R. Martin (1992 : 47).

² *Libération*, 30-31/3/1991, cité par M.-N. Gary-Prieur (1994 : 49).

³ « On notera d'abord qu'il est anormal d'employer un nom propre si l'on ne pense pas que ce nom « dit quelque chose » à l'interlocuteur, si donc l'interlocuteur n'est pas censé avoir quelques connaissances sur le porteur de ce nom. » O. Ducrot dans O. Ducrot et J.-M. Schaeffer (1995 : 367-368).

tel qu'il est construit dans et par le discours. Comme c'est à cette échelle que se pose la question du contenu d'une dénomination propre dans le cadre d'une collecte telle qu'on l'envisage, il convient maintenant de chercher à évaluer son périmètre de façon plus précise. Pour cela, on cherchera à rester dans le cadre d'une sémantique linguistique¹ où ce sont les relations entre unités qui sont prises en compte de manière prioritaire, par opposition à une approche componentielle, centrée sur la signification de l'unité prise individuellement².

II. FORMES DU CONTENU D'UNE DÉNOMINATION PROPRE

A. Contenu ouvert et cadre de la collecte d'informations

La précédente section a permis de caractériser la collecte d'informations que l'on souhaite réaliser à partir des dénominations propres présentes dans le corpus. On a pu voir qu'il ne s'agit pas de rechercher d'hypothétiques thèmes de discours, mais les éléments appartenant à des classes-objets dans lesquelles une dénomination propre est présente, soit comme objet initial, soit comme ingrédient. Si, comme on l'a proposé, on accepte d'étendre la portée de la notion de contenu du nom propre aux éléments présents dans les classes-objets incluant une dénomination propre, la collecte revient alors à établir le contenu des dénominations propres du corpus. C'est sur cette assise discursive que l'on propose de faire ensuite reposer l'insertion des dénominations propres dans un référentiel terminologique³.

¹ « La sémantique la plus productive semble donc être celle qui consiste à interpréter les relations non pas dans le mot simple mais entre mots (simples et complexes), entre mots et groupes de mots, entre propositions et entre énoncés : cette sémantique linguistique a l'avantage de se fonder sur la grammaire. En matière de langues spécialisées, elle permet de valider les intuitions d'hyponymie, de synonymie et d'antonymie, ainsi que les dépendances syntaxiques de terme à terme, donc d'élaborer des réseaux notionnels. » P. Lerat (1995 : 187).

² On s'écarte en cela de l'approche proposée par J.-L. Vaxelaire (2005 : 822-837), pour qui une lexicologie des noms propres passe, certes, par l'étude des contextes discursifs, mais dans la perspective de la sémantique interprétative de F. Rastier. En effet, cette approche ne nous semble pas rendre compte de la spécificité sémantique des noms propres, mieux restituée, semble-t-il, par la notion de contenu.

³ Cette approche soulève la question de la variabilité du contenu d'une dénomination propre en fonction des corpus dans lesquels il est établi. On note que cette variabilité est précisément ce que l'on cherche dans le cadre d'une collecte destinée à l'organisation de l'information. Les DP jouent le rôle de points fixes, mais aussi de réactifs, pour mettre en lumière ces phénomènes de variation qui passent inaperçus chez les locuteurs mais qui font obstacle à l'accès à l'information. Dans un

Cependant, cette approche ne va pas sans soulever de nouvelles questions relatives au choix des dénominations propres à étudier, ainsi qu'aux types d'informations à collecter. En effet, il est nécessaire de se demander maintenant si, parmi les trois mille dénominations propres recensées lors de la phase d'identification, toutes possèdent le même degré de pertinence au regard des informations qu'elles doivent permettre de collecter. Comme on l'a vu précédemment, celles-ci doivent présenter un intérêt en termes d'organisation de l'information. Il devient donc nécessaire d'identifier les dénominations propres dont le contenu dans le corpus est susceptible de répondre au but assigné à la collecte.

De façon plus précise, on peut chercher à pronostiquer la productivité de la collecte à partir de l'usage qui est fait des dénominations propres dans le corpus. Les propriétés textométriques qui leur ont été associées lors de la phase d'identification fournissent d'emblée un certain nombre d'indications. En effet, le classement textométrique indique une répartition inégale des fréquences et de la récurrence parmi les dénominations propres. Par ailleurs, lorsque celles-ci sont classées en fonction des composants de l'activité bancaire, on constate à nouveau des différences importantes dans les effectifs des classes. Enfin, on retrouve à l'intérieur de chaque classe une organisation hiérarchisée, tant du point de vue de l'ordre textométrique que du point de vue de la récurrence. Ces différentes hiérarchisations fournissent des indices sur la variabilité de la fréquence à laquelle les dénominations propres sont utilisées dans le corpus, mais elles ne permettent pas encore d'avoir une idée des formes que prend cet usage. Dans le chapitre 9, on a pu voir que cet usage dépend également de l'appartenance des éléments d'une classe au référentiel partagé par les interlocuteurs. Dans le cas des toponymes, on a pu constater que leur appartenance à un référentiel que le locuteur estime partagé a eu pour effet de limiter la présence des noms de lieu à une forme de citation, sans que rien ou presque soit révélé des caractéristiques de leurs référents. La fréquence ne constitue donc qu'un indice qui doit être interprété à partir de l'étude du fonctionnement discursif.

référentiel terminologique, la question reste ouverte de la notation de la provenance de telle ou telle propriété acquise dans tel ou tel corpus.

En effet, on peut supposer que la densité du réseau qui relie dans les discours les dénominations propres aux autres unités dépend à la fois du rôle qui leur est attribué et de leur appartenance à des pans de référentiels non partagés a priori entre les interlocuteurs. C'est donc la participation de la dénomination propre aux rhizomes du discours que l'on va chercher à évaluer maintenant. Autrement dit, cela revient à interroger les formes que peut revêtir le contenu des dénominations propres. Afin d'aborder cette question, on présentera d'abord deux types de formes possibles de contenu de dénomination propre, avant de présenter l'hypothèse de l'organisation de ce contenu comme réponse à une question topique.

Il est toujours possible d'imaginer un contexte abstrait dans lequel les conditions de productions d'un discours sont momentanément suspendues. Dans un tel contexte, les contraintes du genre ne viennent pas s'imposer aux locuteurs et leurs paroles peuvent être examinées en elles-mêmes. C'est une situation de ce type que le philosophe du langage P.-F. Strawson propose, lorsqu'il s'interroge sur le lien pouvant exister entre, d'une part, les faits distinctifs relatifs à Socrate, que des interlocuteurs pourraient être amenés à produire afin de s'assurer qu'ils parlent du même homme, et d'autre part, le nom « Socrate » :

« Supposons un groupe d'interlocuteurs qui utilisent, ou pensent qu'ils utilisent, le nom « Socrate » de la même manière, c'est-à-dire en référence à la même personne. Ensuite, supposons que nous demandions à chaque membre du groupe de formuler par écrit ce qu'il considère comme les faits les plus marquants concernant Socrate. Et puis, supposons qu'à partir de ces listes de faits nous formions une description composite comprenant les faits les plus souvent mentionnés. On pourrait croire qu'introduire valablement le terme dans ce groupe d'interlocuteurs, par le moyen du nom, implique l'existence d'une seule personne au sujet de laquelle toutes les propositions de cette description composite seraient vraies : mais ce serait trop exiger. Par contre, ce ne serait pas trop exiger que de dire qu'il doit exister une seule personne au sujet de qui une proportion raisonnable de ces propositions est vraie¹. »

¹ P.F. Strawson (1973 : 214-215).

On a vu précédemment que la conception du sens du nom propre adoptée dans ce travail est différente de celle dont Strawson se fait le promoteur, en proposant d'assimiler le sens de cette unité à un ensemble de propositions. Il reste que cette situation particulière, dans laquelle les interlocuteurs sont invités à rendre explicites les éléments du contenu du nom Socrate présumés dans la conversation, est éclairante à plusieurs titres.

Lorsqu'elles sont énoncées en dehors de tout genre de discours, les propriétés associées au référent d'une dénomination propre se réduisent à une simple liste de propositions inorganisées, à la manière d'un inventaire. Cela démontre a contrario la vertu organisatrice du discours sur les matériaux dont se compose le contenu d'un nom propre. Cependant, cette inorganisation ne semble pas être l'inconvénient majeur d'une telle liste : *« Les propositions qui composent la description composite de Socrate formeraient un tel ensemble [un ensemble présuppositionnel de propositions]. Ni les limites d'un tel ensemble, ni la question de savoir combien de membres constitueraient une proportion raisonnable ou suffisante, ne peuvent être fixées de manière précise pour un nom propre quelconque censé introduire un terme. Il ne s'agit pas là d'un défaut dans la notion d'ensemble présuppositionnel ; cela fait partie de l'efficace des noms propres¹. »*

Ce que le cas d'école présenté par P.F. Strawson permet de souligner, c'est le caractère ouvert de l'ensemble des propositions susceptibles de former le contenu du nom propre. Cette ouverture est inhérente à la nature des noms propres. C'est ce que J.R. Searle, dont la position est relativement proche de celle de P.F. Strawson², a également mis en lumière en s'interrogeant sur les conditions nécessaires et suffisantes qui pourraient être définies pour utiliser un nom propre à bon escient : *« Mais le caractère particulier et l'immense utilité pragmatique des noms propres dans notre langue résident précisément dans le fait qu'ils nous permettent de désigner ouvertement des objets sans être obligé d'établir après discussion quels caractères descriptifs constituent exactement l'identité de l'objet. Ils ne fonctionnent*

¹ *Ibid.* (1973: 215).

² G. Kleiber présente J.R. Searle et P.F. Strawson comme les tenants d'une conception modifiée de la « version forte » du sens individuel des noms propres soutenue par B. Russell et G. Frege, et selon laquelle ce sens est constitué par l'unique description définie à laquelle ils se substituent. G. Kleiber (1981 : 370 et suiv.).

*pas comme des descriptions, mais comme des crochets auxquels on peut accrocher des descriptions*¹. »

L'étude du contenu des dénominations propres dans le cadre d'un corpus précis constitue donc une limite « assumée » de nature à lever, au moins provisoirement, l'objection représentée par le caractère ouvert du contenu, qui pourrait le rendre impossible à cerner. Aussi, plutôt que d'entreprendre la recherche d'une vaine exhaustivité à partir de corpus hétérogènes où ont été regroupés des textes de diverses provenances, au nom d'une représentativité toujours sujette à caution, il a semblé préférable de chercher à conduire cette étude dans un cadre délimité avec précision. En effet, si la collecte d'informations doit se faire, non plus à l'échelle d'un seul texte, mais à celle d'un corpus de suivi appelé à s'étoffer de nouveaux textes, on présume que l'homogénéité de la série doit limiter l'éventail des sortes d'informations susceptibles d'être attachées à telle ou telle classe de dénominations propres. Il paraît donc non seulement utile, mais aussi prudent, de privilégier pour une telle collecte les séries textuelles chronologiques. En effet, on dispose ainsi du moyen de contrôler l'évolution des composantes du contenu. Par ailleurs, on a indiqué dans la première partie que l'exploitation d'un tel corpus prend tout son sens à partir du moment où elle s'insère dans un ensemble plus vaste destiné à rendre compte d'autres aspects d'un parler d'entreprise. Le contenu d'un nom propre enregistré dans un référentiel terminologique peut donc être enrichi par des collectes successives, réalisées à partir de différentes sources, elles-mêmes également organisées en séries textuelles chronologiques. Encore une fois, le but visé est toujours le contrôle de la dépendance dans laquelle se trouve le contenu vis-à-vis du discours dans lequel apparaît la dénomination propre.

Lorsqu'on assigne un cadre discursif précis à l'étude du contenu des noms propres, texte ou série textuelle homogène, il devient possible de chercher des régularités dans les associations d'informations caractérisant un groupe de noms propres. Un tel groupe peut consister en une classe pour laquelle les sortes d'informations présentées sont relativement récurrentes et organisées selon un

¹ J. R. Searle (1958), traduit de l'anglais par A. Rey (1976 : 108).

schéma similaire. Les dictionnaires de noms propres ont recours à ce type d'organisation.

B. Un exemple de contenu formaté

Si l'on admet que les dictionnaires constituent une forme de discours¹, et qu'à ce titre ils sont analysables comme d'autres genres plus « évidents », l'étude de leurs articles permet de faire état d'une forme particulière de structure de contenu des noms propres. On se servira ici, à titre d'illustration, du seul *Grand Robert des noms propres*², dans la mesure où certains de ses auteurs ont également produit une réflexion théorique sur leur œuvre, ainsi que sur les ouvrages appartenant au même genre³. Par rapport aux autres genres de discours, les dictionnaires de noms propres offrent une perspective singulière. Alors que dans un texte composé d'énoncés suivis, les dénominations propres et les propriétés associées à leurs référents apparaissent au fil du discours, le discours du dictionnaire juxtapose et concentre les contenus de noms propres livrés sous une forme explicite et organisée.

Les visées didactiques et informatives du dictionnaire qui ont guidé l'organisation des articles sont rappelées dans la présentation de l'ouvrage. Celle-ci fait état d'un canevas général qui a été appliqué à chaque notice. Les articles débutent par des données qui assument « *une fonction de repérage, réalisée d'une façon assez homogène et constante*⁴. » Les noms sont, en effet, suivis d'un classificateur, complété par des données spatio-temporelles succinctes. Le corps de l'article est ensuite organisé différemment selon les principales classes de dénominations propres. Ainsi, pour les personnes, les éléments biographiques, plus ou moins développés selon qu'il s'agit de personnages historiques ou de penseurs, de savants ou de créateurs, sont complétés par des éléments descriptifs ou critiques concernant leurs activités ou leurs œuvres. Pour les lieux, une description sommaire de l'ensemble est étoffée par la présentation de certains aspects culturels et

¹ A. Collinot et F. Mazière (1997 : 1). « *Nous avons délimité un parti pris de lecture qui traite le dictionnaire comme une forme de discours. D'où les deux grands axes qui ont orienté notre travail, historicité et systématisme des discours lexicographiques (...).* »

² A. Rey (1974).

³ A. Rey (1982).

⁴ A. Rey (1974 : xxiv).

historiques. Les notices consacrées aux pays apportent en outre des développements relatifs à leur géographie physique, humaine et économique. Enfin, les régions sont l'occasion d'autres développements encore : « *L'information est répartie hiérarchiquement : ainsi, les grandes unités politiques ou administratives (États des Etats-Unis, Républiques et territoires de l'U.R.S.S., régions de France, comtés de Grande-Bretagne, etc.) donnent lieu à des notices relativement importantes, évitant ainsi une surcharge quand il s'agit de traiter les pays auxquels elles appartiennent¹.* »

L'examen des notices révèle que deux sortes de motivations président à leur structuration. En effet, une première forme d'organisation résulte de choix éditoriaux opérés dans la répartition de l'information pour une série d'entrées désignant des réalités pouvant être emboîtées les unes dans les autres, comme dans l'exemple qui vient d'être cité. Une seconde forme, plus générale, est relative à l'ampleur des développements, qui peut varier de façon importante entre entités appartenant à la même classe ou sous classe. On observe qu'il y a un rapport entre la longueur de la notice et l'utilisation plus ou moins fine des rubriques du canevas dont il a été question plus haut, les auteurs n'ayant pas cherché à appliquer celui-ci de façon systématique, quelle que soit l'importance de l'entité décrite. Ainsi, la notice de « Parigné-l'Évêque » tient-elle sur une ligne « *PARIGNÉ-L'ÉVÊQUE [72250]. Comm. de la Sarthe, arr. du Mans. 4324 hab.* », alors que celle de « Paris » est développée sur plusieurs centaines de lignes et structurée à l'aide des rubriques suivantes : urbanisme, la rive droite, la rive gauche ; économie, industrie, activités tertiaires, communications, structure de l'agglomération ; histoire.

Ces variations dans le traitement des articles ne semblent pas devoir être imputées au caprice des rédacteurs. C'est même la thèse opposée que défend A. Rey : « *L'exigence de concision et de neutralité est fondamentale : elle vise entièrement à l'économie, à la densité de l'information, et donc à accroître ce qui est fourni au lecteur. Le texte des notices – souvent élaboré par des personnes fort capables de tenir un autre discours, et ayant l'expérience du style littéraire – est neutralisé intentionnellement par l'homogénéité d'un code, choisi et pratiqué en*

¹ Rey (1974 : xxv).

*commun par l'équipe rédactionnelle, et contrôlé avec souplesse*¹. » Le fait que la taille des développements varie en fonction de l'importance accordée à telle ou telle entité fait donc partie d'une politique concertée en matière de traitement de l'information. C'est cette même politique qui, en amont, est responsable du choix des entités devant faire partie de la nomenclature du dictionnaire. Ces choix, que l'auteur précité tente d'objectiver dans la mesure du possible, sont en prise avec un référentiel culturellement situé : « *On verra plus loin que les critères de choix, dans un dictionnaire comme celui-ci, sont fonction à la fois de ce que l'on sait de la notoriété du nom et de l'importance des informations qu'il suscite. Il faut dire dès l'abord que le caractère incertain et relatif du premier critère – ce qui est notoire en France ne l'est pas forcément en Chine, ce qui l'est au XX^e siècle ne le sera peut-être pas au XXI^e, etc. – et le caractère subjectif du jugement d'importance rendent le choix toujours contestable, autant qu'ils le rendent difficile*². »

Les notices de dictionnaire fournissent un exemple de structure « formatée » du contenu des noms propres. Cependant, on peut douter que ce qui fait la spécificité de ce genre – la concentration des informations et la visée didactique qui poussent à une organisation homogène de l'information –, soit très répandu. Il est même plutôt probable que la situation inverse soit de règle dans la plupart des genres de discours. Cela signifie que la structure du contenu des noms propres résultera en général d'une construction a posteriori. Celle-ci devra identifier des types d'informations qui ne sont pas étiquetés comme tels (« histoire », « urbanisme », etc.), mais livrés de manière indirecte au fil du texte. Par ailleurs, le repérage des types d'informations devra être envisagé non seulement à l'échelle d'un texte, mais aussi à celle de l'ensemble de la série textuelle chronologique rassemblée en corpus. Car, si la notice de dictionnaire concentre les informations en un point, quitte à proposer des renvois vers d'autres articles, les informations recherchées n'apparaissent dans les textes qu'au fil des occurrences de la dénomination ou de la classe de dénominations propres étudiées. Enfin, comme cela a déjà été mentionné précédemment à propos

¹ *Ibid.*

² A. Rey (1974 : xii).

du contenu du nom propre, cette construction opère un décentrement de l'analyse, qui revient à ne plus étudier le texte pour lui-même, mais pour les informations qu'il livre sur les dénominations propres qu'il contient, et cela, quel que soit par ailleurs le rôle attribué à cette sorte d'unités par l'énonciateur.

Doublement construite par l'observateur (décentrement de la lecture et construction explicite), cette structure de contenu n'en sera pas moins issue des matériaux du corpus étudié. Elle sera construite à partir des informations recensées dans les textes et, comme dans l'exemple fourni par les notices de dictionnaires, on peut supposer que sa richesse entretiendra un rapport de corrélation avec la manière dont telle entité, ou telle classe d'entités, sont valorisées dans un référentiel donné.

Dans la mesure où cette construction est susceptible de reposer sur un nombre important de données, certains types d'analyse détaillée, comme celle proposée dans le cadre de la logique naturelle pour l'analyse des schématisations, ne peuvent guère être utilisés de façon systématique. Il faudra donc trouver le moyen d'élaborer cette structure avec une méthode adaptée, sachant que deux paliers devront être distingués. Le premier se situe au niveau de la classe de dénominations propres et possède un niveau de maille qui s'arrête aux types d'informations véhiculées dans le corpus, relativement à une classe ou à un groupe de dénominations à l'intérieur d'une classe. Le second palier prend en compte, pour une dénomination propre donnée, les informations qui sont associées à celle-ci et qui doivent être recensées avec la forme linguistique qu'elles possèdent dans le corpus.

C. Une structure topique ?

Les deux sections précédentes ont permis d'identifier trois caractéristiques importantes du contenu d'un nom propre. En effet, celui-ci est un ensemble ouvert de propriétés associées au référent du nom propre, lorsque l'on projette d'étudier celui-ci à l'échelle d'un corpus de suivi, ou dans plusieurs corpus constitués en parallèle, afin de rendre compte des différentes strates d'un parler d'entreprise. Dans tous les cas, la constitution de séries textuelles chronologiques semble présenter l'avantage de ménager la saisie de certaines évolutions affectant le contenu, tout en garantissant un relatif contrôle grâce à l'homogénéité des textes sélectionnés. En second lieu, et contrairement à ce que l'on a pu observer pour les notices de

dictionnaires, il est probable que les informations relatives au référent d'une dénomination propre se trouvent réparties au gré des occurrences de la dénomination. Ces informations peuvent donc se présenter en ordre dispersé et être parfois assez éloignées les unes des autres dans les textes. Enfin, les effets de cette dilution sont renforcés par le fait que ces informations ne font ordinairement pas l'objet d'un signalement spécial qui faciliterait leur repérage. On pourrait donc conclure de ces trois observations que la reconstitution du contenu du nom propre à l'échelle de corpus importants relève de la gageure.

Cependant, un tel constat aurait pour inconvénient de contredire de nombreuses observations réalisées au cours de l'identification des dénominations propres du corpus. Lors de cette phase, on a en effet pu constater la récurrence des associations entre une dénomination propre et certaines informations. C'est le cas, par exemple, pour une dénomination propre qui conserve les relations établies avec d'autres unités linguistiques d'un rapport d'activité à l'autre. Mais ce phénomène est également observé pour des groupes de dénominations qui partagent les mêmes sortes d'informations associées, soit dans le même texte, soit d'une année sur l'autre, puisque le corpus utilisé pour cette étude regroupe des rapports d'activité dont la publication est annuelle. On a alors considéré que dans une perspective de collecte d'informations, mais aussi de suivi des évolutions affectant un référentiel particulier, il serait utile de pouvoir tirer parti de ces relatives régularités.

C'est pourquoi, dans un premier temps, on a cherché à caractériser et préciser ce phénomène. Les matériaux étudiés sont produits dans et par le discours. En revanche, leur classement en types d'informations composant le contenu d'un nom propre relève d'un processus d'analyse reposant sur une lecture et une interprétation. La lecture correspond à une phase de repérage des unités présentes dans le voisinage d'une dénomination propre, et susceptibles d'enrichir son contenu. L'interprétation comprend la sélection des unités pertinentes et leur classement en catégories d'informations composant une structure type de contenu. Mais là encore, cette structure a à voir directement avec les textes à partir desquels elle a été élaborée et ne provient en aucun cas d'une problématique que l'on aurait cherché à plaquer sur les textes. Le résultat de ces deux étapes aboutit à l'identification d'un schéma des

sortes d'informations qui sont généralement associées dans le corpus à telle dénomination ou à telle classe de dénominations.

Il a paru utile d'éclairer le schéma ainsi conçu à l'aide de la notion de *topos*. Il convient tout d'abord d'examiner cette notion et de préciser l'acception qui a été retenue dans le cadre de ce travail. À la suite de C. Plantin¹ qui propose une synthèse de la topique antique, on distingue deux plans qui ne correspondent pas au même degré d'abstraction par rapport au texte. Le *topos* correspond, d'une part, à un contenu substantiel apparaissant dans tel ou tel discours, et d'autre part, à une question topique, d'une portée plus générale. Celle-ci est « (...) *un système empirique de collecte, de production et de traitement de l'information à finalités multiples (narrative, descriptive, argumentative), essentiellement pratiques, fonctionnant dans une communauté relativement homogène dans ses représentations et ses normes. Les topiques expriment une ontologie populaire oscillant entre le cognitif et le linguistique. Elles connaissent différents degrés de généralité, la plus générale ayant la forme « qui a fait quoi, quand, où, comment, pourquoi...² ».* Considérer que le l'organisation du contenu d'un nom propre, ou d'un groupe de noms propres, s'apparente à une question topique présente plusieurs avantages.

En effet, la question topique permet de rappeler que la structure de contenu d'un nom propre est de nature discursive, et qu'elle doit être rattachée à un genre de discours précis, dont la série textuelle utilisée comme corpus constitue un échantillon. De ce point de vue, elle permet de présenter la structure du contenu comme la réponse apportée à une question générale adaptée au genre, et qui pourrait avoir la forme suivante : que convient-il de dire de telle ou telle entité nommée dans les rapports d'activité, compte tenu des visées communicationnelles associées, compte tenu également de l'importance accordée à telle entité, ou bien encore, des circonstances dans lesquelles est produite la restitution d'une activité économique révolue ? La structure apparaît dès lors comme indissociable des enjeux et des conditions de production des discours dont elle est tirée.

¹ C. Plantin, article « Topos » dans le *Dictionnaire d'analyse de discours*, sous la dir. de P. Charaudeau et D. Maingueneau (2002 : 576-580).

² *Ibid.* (2002 : 576).

Par ailleurs, la question topique fournit un modèle d'organisation. En effet, non seulement elle peut être organisée elle-même en plusieurs membres, comme dans l'exemple cité plus haut, mais chacun des membres de la question générale est susceptible d'être subdivisé en questions secondaires qui doivent se traduire dans les textes par des réponses adaptées. Si l'on prend l'exemple d'un produit bancaire dont il est fait mention dans le rapport d'activité de telle année, plusieurs informations relatives à ce produit peuvent être présentées : la campagne publicitaire qui a accompagné son lancement, les caractéristiques du produit, la clientèle visée, ce qu'il a rapporté, ce que fait la concurrence dans le domaine, la démarche marketing qui a conduit à son élaboration, le personnel ayant travaillé sur le sujet, etc. On remarque alors que, si ces questions sont toutes légitimes, le genre des rapports d'activité semble n'en retenir qu'une partie. C'est du moins ce que l'on peut constater à partir des « réponses » associées à la dénomination propre dans les textes du corpus, qui n'apprennent rien sur la concurrence ou le processus d'élaboration interne du produit.

La notion de question topique permet également d'aborder la structure de contenu dans ses rapports avec une *doxa* « (...) *c'est-à-dire un ensemble de représentations socialement prédominantes, dont la vérité est incertaine, prises le plus souvent dans leur formulation linguistique courante*¹. » Ainsi, « ce qu'il convient de dire de telle entité dans tel genre de discours », se rapporte non seulement à un référentiel, mais aussi à la manière communément admise de le verbaliser dans tel cadre communicationnel précis, ici les rapports d'activité. La structure de contenu, selon qu'elle est plus ou moins étoffée, donne ainsi une indication sur le degré de valorisation dont bénéficie telle entité, par exemple, dans le cadre de la restitution de l'activité d'un groupe bancaire². La différence de richesse de relations entre structures de contenu de dénominations propres devrait donc permettre de repérer les groupes de dénominations les plus productifs pour la collecte d'informations. Il faut cependant ne pas perdre de vue que cette richesse est

¹ C. Plantin, in P. Charaudeau et D. Maingueneau (2002 : 197).

² On a indiqué précédemment que l'appartenance à des pans de référentiel présumés partagés entre les interlocuteurs est également de nature à faire varier la quantité d'informations relatives au référent de la dénomination propre présentées dans les textes.

relative. En effet, on voit que lorsqu'elle est extraite d'un corpus précis, cette structure a de grandes chances d'être partielle au regard de ce qu'elle pourrait être à l'échelle d'un parler d'entreprise. Ce « défaut » peut être compensé au niveau de la structure en complétant son organisation à partir des éléments de la *doxa* présents dans d'autres séries textuelles.

Enfin, le fait d'assimiler la structure du contenu du nom propre à une question topique permet d'insister sur le caractère récurrent des relations que l'on constate dans les textes. La structure permet de mettre au jour une sorte de noyau de propriétés du référent que l'on trouve reconduites d'une année sur l'autre ou d'une dénomination à l'autre, de telle sorte que des groupes semblent se dégager du seul fait que leurs membres ont en commun ce lot de propriétés typiques dans un corpus donné. En attribuant des propriétés topiques à la structure de contenu de la dénomination propre, on se dispense de la démonstration du fait qu'une partie des discours constitue une réponse délibérée à telle ou telle question topique. Comme on l'a vu précédemment, la notion de contenu du nom propre permettait un décentrement de l'analyse qui se déplace alors du texte englobant à certaines de ses unités que l'on a investies du rôle de catalyseurs d'informations. Considérer que ce contenu est structuré comme une topique revient à faire l'hypothèse que les propriétés associées au référent d'une dénomination propre dans une série textuelle sont motivées par une topique sous-jacente, qui prend place à l'intérieur d'un genre discursif. Il reste à présenter les moyens utilisés pour faire apparaître de telles structures dans le corpus de rapports d'activité.

III. IDENTIFICATION ET CARACTÉRISTIQUES DES STRUCTURES DE CONTENU

A. Structure de contenu et rôle de la dénomination propre

1. Sélection des principaux représentants des classes de dénominations identifiées

On propose de faire reposer l'identification des structures de contenu sur plusieurs étapes, dont le point de départ est constitué par une dénomination, ou un groupe de dénominations propres, et le point d'arrivée par une structure-type du contenu de cette dénomination ou de ce groupe de dénominations propres. Les

premières étapes visent principalement à caractériser la manière dont les dénominations propres étudiées sont utilisées dans le corpus. Elles s'appuient sur les informations qui ont été collectées lors de l'identification des dénominations dans le chapitre précédent.

Dans la deuxième section de ce chapitre on a pu observer le fonctionnement d'un exemple de structure de contenu particulier fourni par les notices d'un dictionnaire de noms propres. On a pu remarquer à cette occasion que ce canevas n'est pas appliqué de façon systématique à tous les noms propres appartenant à telle ou telle classe, comme les noms de lieux ou de personnes. Seules les notices correspondant à des entités jugées particulièrement remarquables par les auteurs du dictionnaire bénéficient d'une structuration hiérarchisée de l'information selon un canevas pré-établi. Transposée dans le contexte de la collecte d'informations à partir des dénominations propres, cette observation incite à rechercher les dénominations propres les plus valorisées dans le corpus. On fait à ce sujet l'hypothèse que ce sont elles qui sont porteuses de la plus grande richesse de liens tissés avec les autres unités présentes dans le corpus. Cependant, on ne dispose pas de l'indice fourni par l'ampleur des notices dans les dictionnaires. En revanche, il est possible de s'appuyer sur un autre indice, fourni par les données textométriques associées à chaque dénomination propre. On va, en effet, considérer que la fréquence de la dénomination constitue une valeur approchée de l'importance qu'on lui accorde dans le corpus.

Une telle correspondance ne va pas de soi. On sait, par exemple, qu'elle n'est pas vérifiée en ce qui concerne les mots grammaticaux : quel enseignement thématique tirer du fait que le mot « de » est généralement le plus fréquent quel que soit le corpus de textes étudié ? On a vu dans la deuxième partie que la terminologie bancaire, que l'on s'attend à trouver dans les rapports d'activité d'une banque, n'occupe pas les premiers rangs du vocabulaire thématique commun du corpus. En revanche, il semble que la question se pose en des termes différents pour les noms propres, du fait de leur statut référentiel particulier. Les noms propres non modifiés (pour reprendre la terminologie de G. Kleiber), signalent à chaque occurrence, qu'il est question dans le texte de l'individu ou de certaines propriétés de l'individu qu'ils servent à désigner. C'est cette relation dite biunivoque qui autorise à établir une

correspondance, même relative, entre la fréquence du nom propre et l'importance qui peut lui être accordée dans un texte ou dans un corpus homogène. Il reste que cette correspondance doit néanmoins être établie avec prudence, car les usages d'une dénomination peuvent être variés, y compris dans une série textuelle homogène. On commencera donc par observer l'usage qui est fait dans le corpus des principaux représentants des classes de dénominations dont les effectifs sont les plus nombreux.

Dans le chapitre 8, on a vu que les classes de dénominations propres les mieux représentées dans le corpus sont les agents économiques, la catégorie « divers » et les objets. On détaillera ici les deux premières catégories afin d'obtenir des sous-classes plus homogènes. En effet, la catégorie des agents économiques doit être divisée en personnes morales (1247 représentants) et en personnes physiques (380 unités). Dans la catégorie « divers » on isole la sous-classe la mieux représentée, les noms de lieux (398 dénominations identifiées). Enfin, la catégorie « objet », qui est quatrième par l'effectif (251 unités), est également retenue. Pour chacune de ces classes, on retient les cinq représentants les plus fréquents ; ces items formeront le premier palier de chaque classe. On donne ci-dessous les principaux représentants de chaque catégorie. Lorsqu'au moins une variante existe, elle figure dans la deuxième colonne de chaque tableau. Dans ce cas, la fréquence affichée correspond au cumul des fréquences observées pour la forme de référence et ses variantes. Enfin, toutes les dénominations présentées ci-dessous correspondent à un segment répété.

*Tableau 72.
Dénominations de personnes morales (entreprises, organisations) les plus fréquentes*

Dénominations	Variantes	Fréquence
Crédit Agricole	- Crédit agricole - CREDIT AGRICOLE	2761
Caisses régionales	- Caisses régionales de Crédit Agricole - Caisses régionales de Crédit agricole - Caisses Régionales de Crédit Agricole - CAISSES REGIONALES DE CREDIT AGRICOLE - Caisses Régionales CAISSES REGIONALES	664
Groupe	GROUPE	603
Crédit Agricole Indosuez	- CREDIT AGRICOLE INDOSUEZ - CAI	593
Crédit Agricole S A		410

Ce tableau ne recense pas seulement des dénominations propres d'entreprises au sens strict. En effet, « Groupe », « Caisses régionales » sont des mots ou des termes pour lesquels l'usage récurrent de la majuscule constitue une forme de mise en valeur dans les textes du corpus. Ils pourraient tout aussi bien être rattachés à l'ensemble des termes. On a préféré ici tenir compte du critère formel de la majuscule pour les rattacher à l'ensemble des dénominations propres. On reviendra sur ce point dans le dernier chapitre où l'on s'intéressera à l'usage qui est fait de ces unités dans une perspective de collecte.

*Tableau 73.
Dénominations géographiques les plus fréquentes*

Dénominations	Variantes	Fréquence
France	FRANCE	566
Europe	EUROPE	236
Paris		120
Asie	ASIE	103
Italie		96

*Tableau 74.
Dénominations de personnes physiques les plus fréquentes*

Dénominations	Variante	Fréquence
Jean Laurent		32
Marc Bué	Marc BUE	27
René Carron	René CARRON	22
Bernard Michel		21
Thierry Coste		21

*Tableau 75.
Dénominations d'objets les plus fréquentes*

Dénominations	Variante	Fréquence
Compte Service	- Comptes Services - Compte Services - Compte Service Crédit Agricole - Compte Service Crédit agricole	43
Mozaïc	MOZAiC	39
Open	carte Open	20
CAC 40	CAC40	19
TLJ	Tous Les Jours des avantages	10

Ces quatre ensembles ne sont pas à reprendre tels quels si l'on veut approcher finement l'usage qui est fait de ces dénominations propres dans les textes du corpus. En effet, dans les précédents chapitres, on a montré que les dénominations propres sont, par exemple, réutilisées dans des dénominations plus longues, ou insérées dans des séries homonymiques. Les fréquences qui ont été reportées lors de la collecte sont celles des segments répétés correspondants. Pour cerner l'usage réel de ces dénominations dans le corpus, il est d'abord nécessaire de les distinguer des autres dénominations qui pourraient les inclure.

2. Repérage des dénominations propres exactes

Afin de préciser le fonctionnement d'une dénomination propre dans le corpus, il est nécessaire de l'isoler des dénominations dans lesquelles elle pourrait se trouver incluse. Pour cela, on réutilise la procédure de la phase d'identification, qui consiste à contrôler la dénomination à l'aide des segments répétés incluants. Mais la

position adoptée ici sera différente. En effet, lors de l'identification des dénominations, les segments incluants apparaissaient comme chargés d'inutiles scories. Ces bribes de phrases captées par les segments répétés vont être utilisées maintenant comme marques des limites de la dénomination propre étudiée. Par exemple, la dénomination propre « Crédit Agricole », qui correspond à un segment de 1700 occurrences, doit être distinguée de « Crédit Agricole Indosuez » (593 occurrences) ou « Crédit Agricole S A » (410 occurrences), pour ne citer que les exemples ayant les fréquences les plus élevées. On donne ci-dessous le résultat de cette tâche de délimitation des dénominations propres.

*Tableau 76.
Dénominations et syntagmes « délimiteurs »*

Dénominations	Syntagmes délimiteurs	Frq
Crédit Agricole	le Crédit agricole (399 occ.) du Crédit agricole (315 occ.) le Crédit Agricole (151 occ.) Le Crédit agricole (137 occ.) du Crédit Agricole (128 occ.) Le Crédit Agricole (65 occ.) au Crédit agricole (30 occ.) au Crédit Agricole (15 occ.)	1240
Crédit Agricole Indosuez	Crédit Agricole Indosuez a (50 occ.) Crédit Agricole Indosuez Crédit Agricole (37 occ.) Crédit Agricole Indosuez Indosuez (37 occ.) Crédit Agricole Indosuez et (21 occ.) Crédit Agricole Indosuez est (16 occ.) de CAI (19 occ.) CAI a (16 occ.) Indosuez CAI (13) CAI est (8 occ.) Groupe CAI (5 occ.) CREDIT AGRICOLE INDOSUEZ (22 occ.)	245
Groupe	Groupe a (70 occ.) Groupe et (28 occ.) Groupe en (25 occ.) Groupe de (20 occ.) Groupe s (18 occ.) Groupe les (16 occ.) Groupe est (16 occ.) Groupe ont (12 occ.) Groupe d (11 occ.) Groupe dans (10 occ.)	226
France	La graphie « FRANCE » n'est pas reprise car elle est majoritairement utilisée que dans des dénominations propres du type « VAL-DE-FRANCE » correspondant à des noms de Caisses régionales.	543
Paris	La graphie « PARIS » n'est pas reprise, car elle entre exclusivement dans la dénomination propre de la caisse régionale « PARIS ET ÎLE-DE-FRANCE »	120

Les syntagmes sont repérés à partir de la liste des segments répétés contenant la dénomination ; les segments sont sélectionnés en fonction de deux critères : le caractère optimal de la délimitation qu'ils contiennent et leur fréquence élevée. Le caractère optimal de la délimitation tient à sa valeur discriminante par rapport aux dénominations incluant. Par exemple, le segment « le Crédit Agricole » permet d'exclure les dénominations propres « Crédit Agricole Indosuez » et « Crédit

Agricole S.A. », car celles-ci s'écrivent toujours sans article ; « *le Crédit Agricole Indosuez » et « *le (ou la) Crédit Agricole S.A. » ne se rencontrant jamais dans le corpus. La fréquence élevée permet de garantir la représentativité de l'échantillon des occurrences de la dénomination propre étudiée. Les résultats de ce tri sont ensuite rassemblés dans un groupe de formes dont on contrôle la pertinence à partir de concordances. On obtient ainsi une valeur approchée de la fréquence de la dénomination, sachant que la grandeur réelle se situe entre deux artefacts : l'un repose sur l'assimilation de la dénomination propre au segment répété, l'autre est créé par la recherche de délimiteurs et parie sur des rencontres récurrentes susceptibles de varier, entre autres, en fonction du rôle de l'entité nommée.

On remarquera que les segments permettant de délimiter la dénomination propre peuvent appartenir aussi bien aux contextes syntactiques que non syntactiques. Ceux-ci fournissent, par exemple, le segment « Crédit Agricole Indosuez Indosuez », où la répétition d'Indosuez, due à l'existence de listes de dénominations, marque la fin de la dénomination propre « Crédit Agricole Indosuez » de façon aussi pertinente que « Crédit Agricole Indosuez a » pour un contexte syntactique. Par ailleurs, le contrôle par les concordances ou l'inventaire distributionnel est nécessaire dans certains cas. C'est ainsi qu'appliqués à « Groupe », ils ont permis de montrer que l'ambiguïté provenant du fait que ce mot puisse désigner aussi bien le groupe Crédit Agricole que le groupe Crédit Agricole S.A. ne résiste pas vraiment à l'examen. C'est en effet très majoritairement le Crédit Agricole qui est ainsi pointé dans les rapports d'activité de la période¹. Ce caractère très majoritaire assure une représentativité suffisante pour poursuivre l'étude de l'unité « Groupe ». Enfin, cette vérification ne s'est avérée nécessaire que pour 5 dénominations sur 20, ce qui ne représente pas une charge de vérifications exorbitante. D'une manière générale, celle-ci n'a pas besoin d'être systématique.

¹ On a ainsi une confirmation de la remarque faite dans la deuxième partie à propos de cette unité et de la différence d'usage constaté entre la forme avec et sans majuscule (Chapitre 7).

3. Le « rôle » des dénominations propres : entre structures et activités

a. La démarche suivie

Comme cela a été montré dans le chapitre 9, le « rôle » des dénominations propres atteste, à l'intérieur d'un genre de discours, l'existence d'une pragmatique propre à l'énonciateur. Celle-ci détermine l'usage qui est fait de ces unités dans les textes d'un corpus, organisés sous la forme d'une série textuelle homogène et chronologique. L'étude du fonctionnement discursif des toponymes a permis d'identifier deux rôles principaux, celui de mention simple, réservé aux contextes non syntactiques, et celui d'insertion dans le discours relatif à la restitution de l'activité bancaire. Dans ce deuxième cadre, on a constaté que le rôle pouvait prendre plusieurs aspects. Certains toponymes servent ainsi à désigner par métonymie une caractéristique saillante de leur référent en rapport avec le domaine de référence économique. Les noms de « pays-marchés » servent également à organiser la restitution de l'activité bancaire en permettant l'ouverture de domaines de discours spécifiques.

L'identification du ou des rôles attachés à une dénomination ou à une classe de dénominations propres permet ainsi de préciser leur usage dans les textes d'un corpus. Celui-ci est déterminé par le genre auquel se rattachent les textes du corpus – d'où l'avantage qu'il y a à utiliser une série homogène de textes qui permet de contrôler la variable générique. Au genre semblent attachés un certain nombre d'usages habituels relativement à une classe de dénominations propres. Par exemple, un lecteur ordinaire pourra s'attendre à ce que des traitements différents soient appliqués au nom de pays « France », selon qu'il sera question du pays dans une brochure touristique, un manuel de géographie, un roman ou un corpus de rapports d'activité. Dans les deux premiers cas, le contenu du nom est susceptible d'être amplifié par des descriptions plus ou moins détaillées de son référent « la France ». Ce type d'amplification peut naturellement se retrouver dans un roman – elle correspond dans un roman du XIX^e siècle aux propriétés associées aux différents personnages¹ –, mais ce dernier genre littéraire offre également un terrain propice à

¹ On renvoie à titre d'exemple à Gollut et Zufferey (2000).

l'apparition de connotations associées par le narrateur à la dénomination. On sait l'usage qu'en a fait Marcel Proust dans *À la recherche du temps perdu*. Enfin, on a vu que les rapports d'activité utilisent la dénomination « France » non seulement comme simple repère spatial, mais aussi comme pays et comme nom de marché économique.

En effet, on sait qu'à l'intérieur d'un même texte le nom propre peut être utilisé pour désigner le référent envisagé dans sa globalité ou seulement pour l'une de ses facettes :

« Admettons que je veuille, par exemple, parler à mon interlocuteur d'un certain aspect de la ville de Paris. Autrement dit, je veux faire de cet aspect précis du particulier nommé Paris l'individu sur lequel portera mon assertion. Une des solutions possibles consiste à combiner en une description définie le prédicat de dénomination être appelé /Paris/ avec un ou des prédicats descriptifs restricteurs :

a) Le Paris de mon enfance n'est plus le Paris d'aujourd'hui

b) J'aime le Paris nocturne / le Paris des banquiers, etc.

Une telle procédure est possible avec tout nom propre :

c) Le Victor Hugo de la jeunesse ne vaut pas le Victor Hugo de la vieillesse¹. »

On ajoutera que ces restrictions opérées sur les aspects du référent dont il est question peuvent emprunter une forme moins explicite que les descriptions données en exemple. En effet, le contexte discursif dans lequel s'insère la dénomination propre peut également jouer un rôle restrictif.

On voit que pour ces deux raisons au moins – les variations d'usage induites par chaque genre de discours et le fait que seuls certains aspects du référent peuvent être mis en évidence –, la manière dont les dénominations propres sont utilisées dans un texte ou une série textuelle homogène n'est pas prédictible avec précision. Dans le cadre d'une collecte d'informations réalisée à partir des dénominations propres, l'examen de ce point doit permettre d'affiner la connaissance des sortes d'informations que l'on peut escompter collecter à partir de telle ou telle classe de dénominations propres. On doit pouvoir identifier également si plusieurs rôles sont

¹ G. Kleiber (1981 : 336). On a vu dans le précédent chapitre 9 que G. Kleiber a finalement renoncé assimiler le sens des noms propres à l'abréviation d'un prédicat de dénomination *être appelé /N/ (x)* au profit d'un sens instructionnel.

assignés à la même dénomination ou classe de dénominations propres, un traitement homogène à l'échelle d'un texte ou d'un corpus entier n'étant pas prévisible.

Dans une première approche, on peut chercher à préciser l'usage qui est fait des principaux groupes de dénominations propres en observant la manière dont les occurrences de leurs représentants les plus fréquents sont réparties entre les deux sortes de contextes, à dominante syntactique ou non syntactique. Comme on l'a vu précédemment, ces deux contextes sont caractéristiques de traitements discursifs suffisamment dissemblables pour correspondre à rôles différents. Afin de conduire cette étude, on rassemble autour de chaque dénomination propre de référence ses éventuelles variantes, afin de constituer un groupe de formes et de polyformes selon les cas. Ce groupe est ensuite projeté sur un graphique dans lequel tous les paragraphes des textes du corpus ont été rassemblés en fonction de leur appartenance à l'un ou l'autre des deux contextes. Les occurrences des groupes de formes sont donc réparties entre deux pôles, CDN et CDS. On présente ci-dessous les ventilations des quatre classes de dénominations propres présentées dans la première section.

b. Les dénominations propres de personnes morales

À la lumière des observations réalisées dans le chapitre 9, la répartition des occurrences des groupes de formes des principales dénominations propres de personnes morales indique que la mention simple, suffisante pour les contextes à dominante non syntactique, est généralement minorée par rapport à l'insertion dans le discours suivi, ce qui constitue une indication favorable pour l'implication de ces unités dans la restitution de l'activité bancaire.

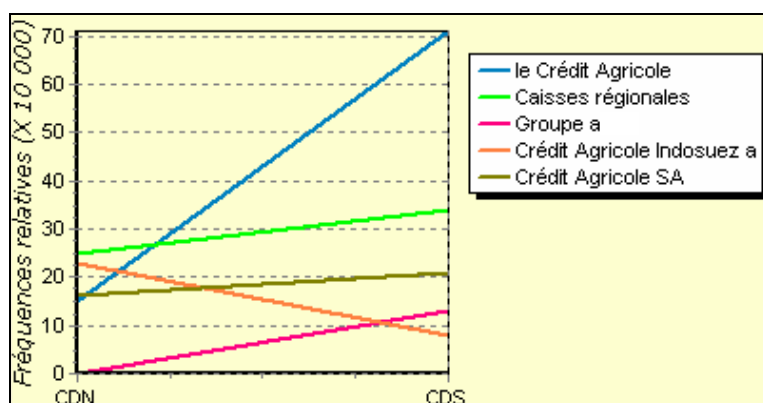


Figure 78.
Répartition entre CDN et CDS des occurrences des groupes de formes des dénominations propres de personnes morales les plus fréquentes

En effet, on note que l'orientation dominante de la répartition entre CDN et CDS va ici dans un sens opposé à celui qui caractérise le groupe des dénominations propres prises dans leur ensemble. En effet, on constate que les représentants les plus fréquents des personnes morales sont proportionnellement plus nombreux dans les CDS que dans les CDN. C'est particulièrement flagrant pour le groupe de formes « le Crédit Agricole », mais on constate le même mouvement, quoique plus atténué, pour « Caisses régionales¹ » et « Crédit Agricole SA ».

La répartition différente que l'on constate pour le groupe « Crédit Agricole Indosuez a » est due au fait que cette dénomination est réutilisée de nombreuses fois dans les contextes à dominante non syntactique (voir dans la section précédente la présence de segments répétés typiques de ces contextes). Cela s'explique en grande partie par la présence d'implantations de cette banque dans de nombreux pays : lorsque les implantations internationales du groupe bancaire sont présentées dans les CDN, « Crédit Agricole Indosuez » revient beaucoup plus fréquemment que

¹ « Caisses régionales » ne constitue pas une dénomination propre. Elle fait néanmoins partie de ces unités qui se situent sur un *continuum* entre termes et dénominations propres. En effet, l'utilisation du pluriel combiné à l'usage de la majuscule sert toujours dans le corpus à désigner le groupe des caisses régionales. Lorsqu'une caisse est désignée en particulier, c'est la dénomination complète qui est utilisée. Par ailleurs, on ne trouve pas d'occurrence de « caisse régionale » utilisée dans l'absolu pour désigner le concept de caisse régionale dans le corpus.

« Crédit Agricole », « Caisses Régionales » ou « Crédit Agricole SA », qui correspondent à des dénominations d'entreprises localisées en France.

On note cependant que quatre dénominations sur cinq sont présentes dans des proportions assez voisines dans les CDN. C'est qu'elles sont également insérées dans des dispositifs typiques des CDN, tels que les organigrammes financiers ou des schémas censés clarifier l'organisation du groupe bancaire. De ce point de vue, la répartition des occurrences du groupe de formes « Groupe a » fait office d'exception confirmant la règle. Contrairement aux dénominations propres et aux termes, qui fonctionnent sans difficulté en dehors du contexte phrastique, l'unité « Groupe » est handicapée par l'ambiguïté qu'elle pourrait acquérir dans de tels contextes. En effet, son utilisation oblige alors à préciser de quel groupe il s'agit. C'est pourquoi, on ne la rencontre dans les CDN que sous la forme complexe « Groupe Crédit Agricole », « Groupe Crédit Agricole Indosuez », « Groupe Sofinco », etc., où les dénominations propres jouent le rôle de restricteurs¹.

Le groupe des dénominations propres de personnes morales se caractérise donc par une présence dominante dans les CDS, ce qui suggère une participation étroite à la restitution de l'activité bancaire. Une présence moindre, mais non négligeable, dans les CDN indique que ces dénominations sont également utilisées pour la présentation de la structure du groupe bancaire.

c. Noms de lieux

La répartition des fréquences des dénominations de toponymes entre contextes à dominante syntactique et non syntactique présentée dans le chapitre 9 a mis en évidence des situations contrastées. Le rôle de simple repère géographique est partagé par la plupart des dénominations appartenant à cette classe. Lorsqu'il est presque exclusif, comme pour la sous-classe des odonymes, on a noté une présence majoritaire dans les contextes non syntactiques. En revanche, cette étude a

¹ La forme « Groupe » est en quelque sorte « condamnée » aux CDS dans la mesure où, en dehors des cas où il entre dans une unité complexe, il ne peut fonctionner que comme reprise anaphorique de « Crédit Agricole » ou « Crédit Agricole SA ». Le groupe de polyformes qui reprend des segments dans lesquels la forme est incluse avec des mots grammaticaux et des particules verbales a pour fonction d'isoler les cas ambigus où « Groupe » se combine avec des dénominations propres. Un contrôle manuel montre, par ailleurs, qu'il est bien absent des CDN sous sa forme simple « Groupe ».

également permis de montrer que les dénominations propres de villes et de pays sont, quant à elles, susceptibles de revêtir d'autres rôles. La désignation d'une place financière d'importance mondiale pour certains noms de villes, ou celle d'un marché économique national pour tel ou tel nom de pays, sont des usages courants reposant sur la sélection, par un processus métonymique, d'une caractéristique saillante de tel ou tel espace géographique. Enfin, les noms de pays fonctionnent également dans les discours comme des introducteurs d'univers de discours utilisés pour organiser la restitution des activités du groupe bancaire. Ces deux rôles se manifestant pour l'essentiel dans les contextes à dominante syntactique. Dès lors, il n'est pas surprenant de constater que les répartitions des fréquences des toponymes les plus récurrents n'aient pas toutes le même profil.

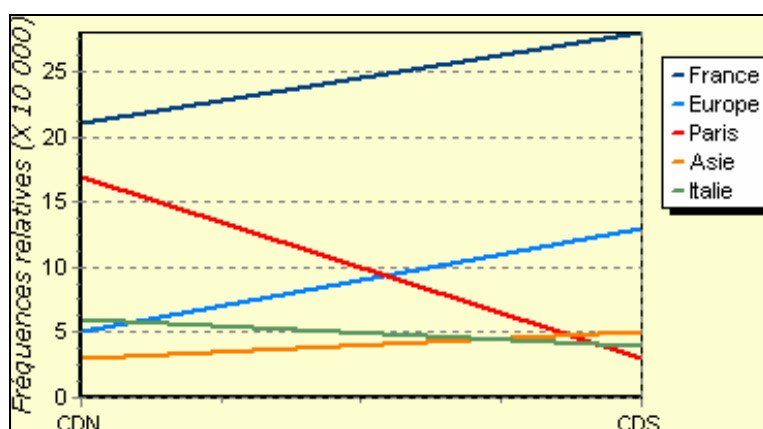


Figure 79.

Répartition entre CDN et CDS des occurrences des noms de lieux les plus fréquents

La ventilation de ces cinq toponymes entre contextes à dominante syntactique et non syntactique peut se lire, en fonction des remarques qui viennent d'être faites, comme la manifestation plus ou moins marquée d'un ou de plusieurs rôles tenus par ces dénominations dans les textes du corpus. Ainsi, les noms géographiques « France » et « Europe » présentent une répartition de leurs occurrences similaire entre les deux contextes, au profit toutefois des contextes à dominante syntactique, ce qui atteste leur usage pour au moins deux rôles sur trois parmi ceux qui ont été identifiés. Compte tenu de la dominante CDS qui caractérise les deux courbes, on peut même avancer que ces deux toponymes endossent tous les rôles recensés dans le corpus pour leur classe. Dans une moindre mesure, c'est

également le cas pour « Asie ». Quant à la dénomination « Italie », elle réside une répartition plutôt équilibrée entre les deux sortes de contextes, mais cette fois en faveur des CDN. Seule la courbe de Paris présente une répartition alignée sur celle des odonymes, ce qui montre un rôle important de repère géographique combiné, mais dans des proportions moindres, avec un rôle de place financière.

Compte tenu de ces observations, il conviendra de se demander de quelle manière les noms de lieux participent à la restitution de l'activité bancaire quand ce n'est pas seulement pour la situer dans l'espace ou lui ouvrir des cadres discursifs. De ce point de vue, la situation des odonymes est emblématique. En effet, lorsque « Wireless Road », rue de Bangkok en Thaïlande, est mentionnée parmi les coordonnées d'un bureau de représentation du Crédit agricole, c'est sans plus d'explications. La mention d'un odonyme présuppose l'existence d'une voie, *a fortiori* lorsque celle-ci apparaît dans un dispositif de coordonnées postales. Ce qui est vrai du très spécifique l'est aussi du plus général : East Monroe Street, Chicago, Illinois, Etats-Unis, sont utilisées comme des données encyclopédiques censément partagées par tous. En dehors des informations collectées à partir des toponymes pris pour eux-mêmes (on a vu qu'ils permettent alors d'accéder aux évolutions affectant le référentiel géographique de l'organisation étudiée), on peut d'ores et déjà supposer que cette classe d'entités n'offrira pas un rendement très élevé pour la collecte de vocabulaire¹.

d. Noms de personnes physiques

Le graphique ci-dessous met immédiatement en lumière le fait que les dénominations propres de personnes physiques sont presque exclusivement cantonnées à un rôle de mention simple dans le corpus de rapports d'activité. En effet, leur présence prépondérante dans les contextes à dominante non syntactique ne permet pas une intégration complète à la restitution de l'activité du groupe bancaire.

¹ Voir *infra*.

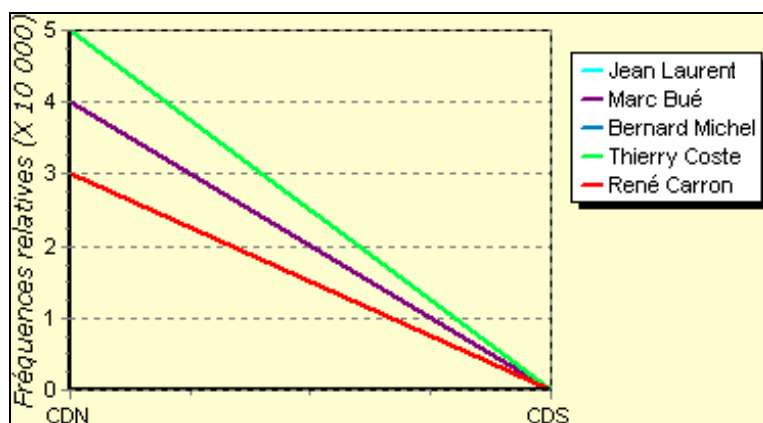


Figure 80.
Répartition entre CDN et CDS des occurrences des dénominations propres de personnes physiques

En effet, lorsque l'on a cherché à déterminer quelles étaient les formes typiques des CDN¹, sur dix formes, neuf prénoms ont été sélectionnés. Il n'est donc pas surprenant de constater, maintenant que l'on examine la ventilation des occurrences des dénominations propres de personnes physiques entre CDN et CDS, que celles-ci sont majoritairement présentes dans les CDN. On remarquera également que certains profils sont superposés dans le graphique, certaines dénominations partageant le même profil.

Cependant, il peut paraître surprenant de voir que ces dénominations de personnes, dont la notoriété sort peu du contexte propre au Crédit Agricole, figurent dans des contextes plutôt réservés aux structures et aux éléments partagés par les interlocuteurs. On peut alors émettre l'hypothèse selon laquelle ces noms de personnes physiques ont pour rôle d'incarner les structures abstraites des organigrammes dans lesquels ils apparaissent². En effet, on sait que les organigrammes des grandes entreprises fonctionnent à la manière d'un leurre, qui donne l'impression de parler de leur organisation interne, alors qu'en réalité ils cèlent toute information révélatrice à un public externe. Cependant, dans la mesure

¹ Voir chapitre 3.

² Il existe néanmoins des exceptions qui concernent les directeurs généraux et les présidents de la Caisse Nationale de Crédit agricole puis de Crédit Agricole S.A. En effet, lors d'un changement de dirigeant, les rapports d'activité présentent le curriculum vitae du nouvel arrivant dans les contextes à dominante syntactique.

où les rapports d'activité fonctionnent aussi comme un document de communication destiné à un large public, on peut supposer que la présence de ces noms constitue un réel enjeu de pouvoir : être ou ne pas être dans l'organigramme du rapport d'activité. On retiendra néanmoins que les noms de personnes physiques semblent n'entretenir qu'un rapport indirect¹ avec la restitution de l'activité d'un groupe bancaire.

e. Dénominations propres d'objets

Une rapide comparaison avec le graphique des noms de personnes physiques fait ressortir que la répartition des principaux représentants des dénominations propres d'objets est orientée dans le sens inverse : leur présence semble en effet réservée aux CDS.

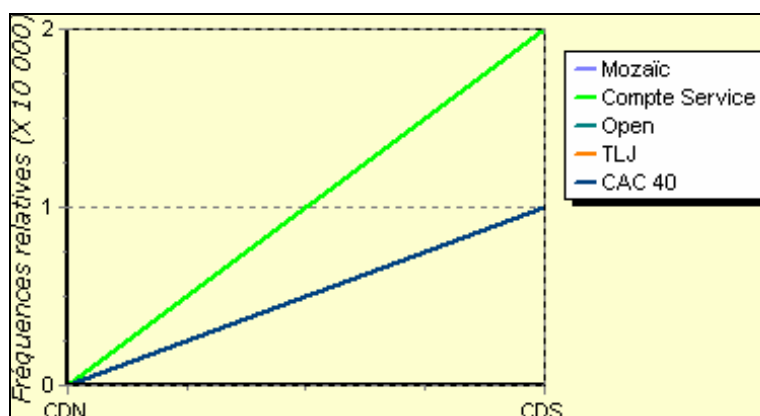


Figure 81.

Répartition entre CDN et CDS des occurrences des dénominations propres d'objets

Cependant, comme les noms de personnes physiques, ces unités sont nettement moins fréquentes que les dénominations de personnes morales et de lieux, ce qui explique la faiblesse des fréquences relatives. Autrement dit, il s'agit d'unités plutôt rares dans le corpus. En revanche, elles semblent impliquées de façon non négligeable dans la restitution de l'activité bancaire. Leur absence dans les CDN, où

¹ On voit que la perspective adoptée ici s'écarte des approches de type veille économique dans lesquelles le jeu d'apparition et de disparition des dirigeants d'entreprises est suivi de près, comme un indice possible de changement de stratégie économique. Dans le contexte des rapports d'activité, il ne peut s'agir que d'un lien indirect avec la restitution de l'activité bancaire, telle que l'on souhaite l'étudier ici. Mais ces apparitions et disparitions de dénominations ne livrent en général rien qui puisse éclairer les non-initiés sur la portée de tels changements.

elles auraient pu figurer, par exemple dans des listes, à la manière de ce que l'on peut trouver dans des catalogues, confirme en retour le statut non commercial des rapports d'activité.

Les deux sortes de contextes CDN et CDS proposent chacun une version différente de la restitution de l'activité économique. Les premiers livrent l'organisation de l'entreprise à un moment donné, celui qui correspond à l'année du rapport d'activité ; il s'agit d'une représentation statique, où les aspects structuraux dominent. On a vu dans la première partie des exemples de fiches signalétiques ou d'organigrammes. Dans ces contextes, les dénominations propres peuvent être mises en relation les unes avec les autres sans l'intermédiaire d'aucune autre unité linguistique, par simple juxtaposition. À l'inverse, les contextes à dominante syntactique offrent une restitution de l'activité comme processus ou série d'événements, dans laquelle les dénominations propres sont en contact avec toutes les composantes du vocabulaire des rapports d'activité.

L'exploration des rôles susceptibles d'être tenus par les dénominations propres à l'aide de la répartition de leurs occurrences entre contextes CDN et CDS a permis de distinguer trois sortes de profils. C'est ainsi que les dénominations de personnes morales et les noms de lieux apparaissent comme des unités « amphibies » pouvant être utilisées dans les deux sortes de contextes. En revanche, les noms de personnes physiques semblent plutôt cantonnés aux CDN, alors que c'est l'inverse pour les noms d'objets, qui sont presque exclusivement présents dans les CDS. Compte tenu de ce qui a été rappelé plus haut au sujet des deux sortes de contextes, la collecte d'informations sera probablement plus productive pour les noms de personnes morales et d'objets. Les noms de personnes physiques présentent les caractéristiques inverses. Enfin, les noms de lieux posent la question de leur lien avec la restitution de l'activité bancaire, au-delà d'une simple fonction de localisation qui aurait pu être limitée aux seuls CDN. On a pu voir dans le précédent chapitre qu'une telle liaison est établie lorsqu'un aspect caractéristique d'une ville ou d'un pays est sélectionné par rapport au domaine de référence économique qui sous-tend le contenu des rapports d'activité. Par ailleurs, les noms de pays sont utilisés pour ouvrir des domaines de discours servant à encadrer la restitution de

l'activité bancaire. Cette position singulière devra être précisée par rapport aux buts poursuivis par la collecte.

Cet examen des rôles principaux pouvant être tenus par les dénominations propres dans les textes du corpus, mention simple dans des dispositifs non syntactiques, insertion plus ou moins étendue dans les discours, montre le caractère hétérogène du fonctionnement discursif de cette classe d'unités. Au-delà de ce simple constat, un tel examen fournit aussi, semble-t-il, une indication utile dans la perspective de la collecte d'informations. L'étude des structures du contenu de ces classes de dénominations propres va permettre d'approfondir la question du traitement discursif qui leur est réservé dans les textes du corpus.

B. Structure type du contenu d'une dénomination propre

1. Identification des structures types

Comme on l'a vu précédemment, la notion de structure type vise à formaliser l'usage qui est fait d'une dénomination propre ou d'un groupe de dénominations dans un corpus, à partir des relations typiques établies dans les discours avec des unités pouvant être rattachées à la même classe-objet. Plus ces relations typiques sont nombreuses, plus on peut considérer que la structure du contenu d'une dénomination ou d'un groupe de dénominations est riche. Cette notion relative de richesse vient ainsi affiner les indications fournies par les données textométriques. On fait l'hypothèse que plus cette structure est riche, plus les unités de la catégorie étudiée sont susceptibles de remplir correctement la fonction d'accès aux classes-objets que l'on attend d'elles. On dispose alors d'un moyen pour évaluer la « productivité » de la collecte.

Par ailleurs, la complétude de la structure et la richesse des unités qu'elle permet de collecter paraissent à même de fournir des indications utiles sur l'organisation interne de la classe de dénominations étudiée. La reconstitution d'une telle structure doit en effet permettre d'évaluer l'homogénéité d'une classe ou d'une

sous-classe et d'écarter des unités qui, bien qu'ayant un profil statistique proche des dénominations propres rassemblées, appartiennent en fait à une catégorie différente¹.

Enfin, cette structure fournit un repère utile afin d'identifier une éventuelle évolution dans la manière dont une catégorie de dénominations propres est utilisée dans une série textuelle chronologique. En effet, si la structure-type peut être établie à partir des données fournies par l'ensemble du corpus, rien n'empêche, dans un deuxième temps, de suivre année après année la manière dont la structure est réutilisée. Celle-ci peut également être amenée à évoluer, non seulement du fait de la relative plasticité des dispositifs textuels rassemblés dans le corpus (chapitre 2), mais aussi parce que le corpus est ouvert et que de futures parties viendront le compléter. La question qui se pose à l'échelle de la structure-type du contenu d'une catégorie de dénominations propres doit également être soulevée pour chaque dénomination. En effet, on peut se demander si, par exemple, la manière de catégoriser une entité nommée n'est pas susceptible de varier en fonction du développement de son activité et de la manière dont celle-ci est restituée. On dispose ainsi d'une meilleure prise sur l'évolution d'un référentiel que celle qui se limite à la comptabilisation des apparitions ou des disparitions de noms d'entités, comme dans certains systèmes de veille².

Il est nécessaire d'établir une distinction dans l'utilisation d'une structure-type selon que l'on est en phase d'identification ou en phase de collecte des informations constituant le contenu d'une dénomination propre dans un corpus donné. Pour la première phase, qui sera abordée dans cette section, il s'agit de déterminer les types d'informations associés à une classe de dénominations propres. Autrement dit, identifier la structure-type d'une dénomination ou d'une classe de dénominations propres, c'est rechercher les membres d'une question topique sous-jacente à partir des données attestées dans le corpus. Les unités sont repérées dans le voisinage des dénominations propres ; les informations ainsi collectées sont résumées en vue d'obtenir une structure simple et suffisamment générique pour embrasser tous les cas d'associations que l'on peut rencontrer dans le corpus.

¹ F. Erlos (2006).

² T. Poibeau (1999).

Lorsque la structure-type a été identifiée, elle est utilisée comme fil conducteur pour la collecte systématique des unités composant le contenu de chaque dénomination propre. Ces unités sont alors organisées en fonction de la structure type, qui fonctionne aussi comme un moyen d'évaluer l'écart du contenu en cours de constitution par rapport au type. Cette deuxième phase sera utilisée dans le dernier chapitre pour les dénominations qui auront été retenues pour la collecte d'informations.

a. Champ textuel pour l'identification d'une structure type de contenu

La notion de structure de contenu permet de franchir un pas supplémentaire dans la compréhension du fonctionnement discursif d'une catégorie de dénominations propres. Ce faisant, on est également amené à s'interroger sur les contextes discursifs pertinents qui peuvent être explorés à partir d'une dénomination propre. Pour la logique naturelle, les classes-objets se déploient dans le cadre d'énoncés formant des configurations : « *Ainsi les énoncés sont organisés en configurations de dimensions variables, lesquelles à leur tour se composent pour constituer un tout. Le tout peut être de la taille de À la recherche du temps perdu ou de « Bains interdits. Eau polluée*¹. » On constate cependant que les exemples utilisés par cet auteur sont fréquemment des extraits de textes limités à quelques paragraphes, ou bien des articles de presse relativement brefs. Ces contextes suffisent d'ailleurs en général à faire émerger plusieurs objets de discours. Dans la première section de ce chapitre, on a pu le vérifier avec l'exemple proposé par J.-M. Marandin.

Une autre approche, celle proposée par F. Rastier dans le cadre d'une sémantique interprétative, fournit un ordre de grandeur similaire, dès lors qu'il s'agit de caractériser l'espace de déploiement textuel d'un thème : « *Le thème, par ses récurrences, intéresse la macrosémantique. Mais en tant qu'unité, il relève de la mésosémantique. D'après nos relevés, ses lexicalisations diverses apparaissent généralement dans un espace inférieur à trois cents mots. Un espace de cinquante*

¹ J.-B. Grize (1990 : 73).

mots environ suffit pour identifier quatre occurrences d'un thème sur cinq¹. » F. Rastier établit ainsi une certaine correspondance entre l'espace au sein duquel le thème se déploie et le découpage typographique du texte en paragraphes : « Retenons cependant que le paragraphe est l'homologue (non l'analogue) sur le plan de l'expression de la période sur le plan du contenu : c'est donc dans son empan que l'on trouve l'essentiel des co-occurents correspondant au thème recherché². » Ces constats doivent être rapprochés des conclusions précédentes au sujet de la fenêtre de calcul des co-occurents des dénominations propres. En effet, on a pu voir dans le premier chapitre 8 que la notion de chaîne de référence, telle que l'a précisée C. Schnedecker³, permet d'établir un parallèle entre le champ d'empaquetage thématique d'une dénomination propre et le découpage du texte en paragraphes. Ces éléments concordants militent donc en faveur de l'adoption du paragraphe comme unité d'exploration textuelle des informations associées à une dénomination propre dans les textes.

b. Méthode utilisée pour l'identification d'une structure type de contenu

Pour cette phase exploratoire, la méthode proposée consiste à rassembler les dénominations propres « isolées » ou non à l'aide de segments incluant dans des groupes de formes et de polyformes. On réutilise donc les groupes déjà constitués pour la phase précédente de contrôle des contextes. Les formes et polyformes de chaque groupe sont successivement projetées sur la carte des sections de Lexico. Cette carte reprend les neuf grandes parties du corpus découpées à leur tour selon que leurs rubriques se rattachent à l'un ou l'autre des deux types de contextes, à dominante syntactique ou non syntactique. L'unité la plus petite dans laquelle la présence d'occurrences est signalée correspond aux paragraphes originaux des textes des rapports d'activité. On dispose ainsi pour chaque dénomination propre, avec ses variantes ou ses éventuels syntagmes incluant, de la cartographie de l'ensemble de ses occurrences dans le corpus.

¹ F. Rastier (1995 : 229), repris dans F. Rastier (2001 : 202-203).

² F. Rastier (2001 : 212).

³ C. Schnedecker (1997).

Les paragraphes dans lesquels une occurrence au moins des formes ou polyformes est présente sont coloriés automatiquement, ce qui permet de les repérer rapidement. Chaque paragraphe correspond à un contexte qu'il s'agit d'analyser dans le but de mettre au jour un type de relation établie entre la dénomination propre et une sorte d'unité. Si l'on projette le premier groupe de formes, c'est-à-dire le groupe correspondant à la dénomination propre « Crédit agricole » comprise dans des segments délimiteurs, il devient possible d'examiner tous les paragraphes dans lesquels les occurrences du groupe apparaissent. On a ainsi, pour le premier contexte à dominante non syntactique, un extrait de sommaire :

« *LE GROUPE*

Un Groupe uni et décentralisé

Chiffres clés

Le Crédit agricole en 1995 »

De même, on a pour le premier contexte à dominante syntactique, un titre accompagné d'une phrase :

« *PROFIL*

Le Crédit agricole est la première banque française par ses parts de marché, le niveau de ses résultats et le montant de ses fonds propres. »

Pour la dénomination propre « Crédit agricole », les paragraphes dans lesquels les segments incluants apparaissent constituent un ensemble de 1027 sections qu'il peut être fastidieux d'examiner une à une. Cependant, on ne recourra pas à ce stade au « raccourci » que pourrait fournir le calcul des co-occurents. En effet, l'identification des relations établies entre l'unité étudiée et les autres unités présentes dans son voisinage repose sur l'observation de l'organisation syntaxique des phrases et des relations sémantiques tissées par le discours. Outre que l'examen des co-occurents ne dispenserait pas en définitive de l'étude des contextes dans lesquels ils co-occurrent avec la dénomination propre étudiée, cette approche aurait pour inconvénient d'introduire une rupture prématurée avec les relations syntaxiques et sémantiques qu'il convient d'analyser à ce stade.

Il existe cependant une autre manière d'épargner à l'analyste un examen exhaustif des paragraphes lorsqu'ils sont très nombreux. Celle-ci tient compte du fait

qu'à ce stade on cherche avant tout à établir des relations et des unités types. Cela autorise donc à travailler à l'aide d'échantillons du corpus. Comme on l'a vu plus haut, la structure type de contenu est élaborée à partir des données présentes dans l'ensemble du corpus. Lorsque celui-ci correspond à une série textuelle chronologique, cela revient à adopter un point de vue synchronique. Afin de tenir compte de l'ensemble de la période couverte par le corpus, on étudiera principalement trois parties sur neuf, situées au début, au milieu et à la fin de la série textuelle (il s'agit des parties 1995, 1999 et 2003). D'autre part, le but de cette analyse est de mettre au jour une structure-type correspondant à une question topique sous-jacente. L'examen systématique des paragraphes comportant des occurrences est utile tant qu'il permet de découvrir de nouvelles relations. Lorsqu'il ne permet plus de faire que des constats redondants avec ceux qui ont déjà pu être faits, autrement dit, lorsqu'à un moment de l'analyse la structure-type semble saturée, la systématique n'est plus requise et l'on peut procéder à nouveau par sondages dans l'échantillon retenu pour le corpus. Enfin, on notera que la difficulté liée au grand nombre de contextes à explorer ne se présente pas pour toutes les classes de dénominations. On peut trouver dans la section suivante les résultats qui ont été obtenus pour les quatre groupes de dénominations qui ont été étudiés.

2. Structures types des classes de dénominations propres étudiées

a. Structure type du contenu des dénominations propres de personnes morales

Pour chacune des classes de dénominations propres étudiées, dénominations de personnes morales, noms de lieux, noms de personnes physiques et noms d'objets, on a collecté les relations-types dans la perspective de constituer une structure type pour la classe. Pour cela, on a cumulé les indications fournies par l'observation du fonctionnement discursif des cinq premiers représentants par ordre textométrique de chaque classe. À ce stade, ont également été cumulées les observations portant sur des parties différentes (les trois parties de l'échantillon) et sur les deux sortes de contextes discursifs (CDN et CDS). On donne ci-dessous pour chaque classe la structure type que l'on a obtenue.

La figure ci-dessous représente les principales sortes d'informations associées dans les discours du corpus aux dénominations propres de personnes

morales. Les informations sont présentées de façon schématique à l'aide d'intitulés génériques et explicites qui ne se trouvent pas sous cette forme dans les textes :

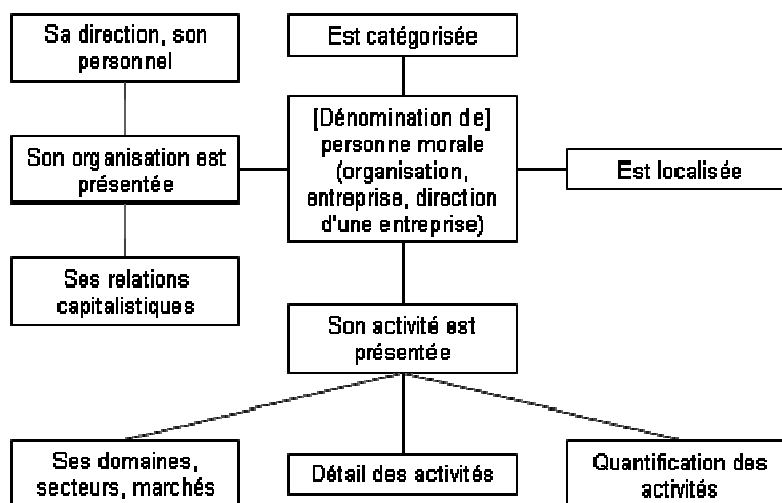


Figure 82.
Structure-type du contenu de la classe des dénominations propres de personnes morales

On donne ci-dessous, à titre d'exemple, certains paragraphes contenant des informations ayant permis d'élaborer la structure du contenu de l'une des cinq dénominations, « Crédit Agricole Indosuez ». Ces paragraphes sont tirés de l'année 1999, Crédit Agricole Indosuez n'existant pas en 1995. Les paragraphes extraits sont présentés dans l'ordre de mention des types d'informations formant la structure de contenu.

La catégorisation, mais aussi les relations capitalistiques (création de filiales) ainsi que le rattachement à des domaines d'activité peuvent faire l'objet de mentions plus ou moins explicites, comme dans l'exemple donné ci-dessous :

(1)« 1996

A la suite de l'acquisition de la Banque Indosuez, naissance de Crédit Agricole Indosuez, banque des grandes clientèles, de l'international et des marchés, et d'Indocam, spécialisée dans la gestion d'actifs. » Dans cet extrait, Crédit Agricole Indosuez est présentée comme la banque des grandes clientèles, de l'international et des marchés du Crédit Agricole ; autrement dit, on apprend simultanément qu'il

s'agit d'une filiale du groupe bancaire, et dans quels secteurs d'activité elle intervient.

Les implantations de cette banque sont nombreuses et sont regroupées dans les contextes à dominante non syntactique, comme dans cet exemple :

(2) « *Allemagne : Beneficial Bank AG (BBAG), Crédit Agricole Indosuez, Crédit Agricole Indosuez Cheuvreux Deutschland GmbH, Carr Futures SNC, Indocam Frankfurt.* »

Cette localisation peut être combinée avec d'autres sortes d'informations, comme les noms des dirigeants, le domaine d'activité, regroupées dans des sortes de descriptions signalétiques:

(3) « *GRANDES CLIENTELES ET INTERNATIONAL, CREDIT AGRICOLE INDOSUEZ*

9, quai du Président Paul Doumer 92920 Paris La Défense Cedex

Tél : 01 41 89 20 00

Président du Conseil de surveillance : Lucien Douroux Président du Directoire : Marc-Antoine Autheman. »

La quantification des activités peut être générale ou détaillée par métiers. Dans ce paragraphe, ce sont les résultats globaux de Crédit Agricole Indosuez qui sont communiqués :

(4) « *Dans la banque des grandes clientèles et de l'international, l'année 1999 a été marquée par le redressement rapide et profond de Crédit Agricole Indosuez (CAI). Son PNB a augmenté de 22,4 % et son RBE de 79 %. CAI a ainsi dégagé un résultat net, part du groupe, de 179 millions d'euros (1,174 milliard de francs).* »

Dans cet autre exemple, c'est l'activité d'un secteur particulier de Crédit Agricole Indosuez qui fait l'objet d'une présentation succincte et d'un chiffrage :

(5) « *Dans le domaine des financements structurés, où un partenariat a été conclu avec le Crédit Lyonnais aux Etats-Unis, les opérations de titrisation en tant qu'arrangeur ou chef de file ont représenté 3 milliards d'euros (19,7 milliards de francs). Elles ont concerné des sociétés françaises, italiennes, espagnoles, suédoises et américaines. Par ailleurs, Crédit Agricole Indosuez a obtenu quinze mandats d'arrangeur en leasing sur des actifs maritimes, aéronautiques, des équipements lourds et des installations industrielles.* » Enfin, l'entreprise peut être rattachée de

façon elliptique à un domaine d'activité général par simple juxtaposition, comme cela se trouve fréquemment dans les contextes à dominante non syntactique :

(6) « *CREDIT AGRICOLE INDOSUEZ Grandes clientèles et international* ».

La présentation de l'organisation de l'entreprise et de son personnel, ainsi que celle des domaines d'activité concernés, se font de façon plus ou moins précise selon les faits remarquables du moment :

(7) « *L'équipe de financements aéronautiques a obtenu vingt-deux mandats de direction portant sur quarante-huit avions d'une valeur totale supérieure à 2 milliards de dollars et onze mandats de conseil/arrangeur. Quant à l'équipe des financements maritimes, elle a réalisé soixante-deux transactions permettant de financer cent quatorze navires dont trente-deux unités neuves. Crédit Agricole Indosuez a ainsi continué de figurer parmi les premiers arrangeurs mondiaux dans ce secteur.* »

Enfin, la direction de l'entreprise fait fréquemment l'objet d'une présentation elliptique mais nominative :

(8) « *CREDIT AGRICOLE INDOSUEZ*

Marc-Antoine Autheman

Président du Directoire, Directeur Général Adjoint de la CNCA. »

Cette collecte a été contrôlée à l'aide des occurrences présentes dans les paragraphes de la partie 2003. D'une manière générale, on constate que lorsque l'on recherche un certain niveau de généralité, les informations sont relativement redondantes et permettent assez vite de saturer la structure de contenu. Ce repérage manuel, qui se fait à grosses mailles et au fil du texte, permet de surmonter un grand nombre de difficultés liées à la diversité avec laquelle toutes les sortes d'informations relatives à l'entité nommée sont communiquées dans les textes du corpus. Ainsi, dans le premier paragraphe cité, l'information selon laquelle « Crédit Agricole Indosuez » est une filiale du Crédit Agricole n'est donnée, ou confirmée, que de façon indirecte par les deux substantifs « acquisition » et « naissance ». Dans les citations 2), 3), 6) et 8), la localisation, le rattachement à un domaine d'activité ou la mention de la direction se font par simple juxtaposition. Les paragraphes 4), 5) et 7), quant à eux, mettent en évidence la sinuosité du parcours qu'il convient d'emprunter lorsque les variantes d'une dénomination se succèdent, ou lorsqu'une catégorisation chemine par cataphore et métonymie jusqu'au référent de la

dénomination propre. La richesse de la structure-type du contenu des dénominations propres peut laisser penser à ce stade que cette classe constitue un accès privilégié aux autres composantes du vocabulaire du corpus. C'est une situation toute différente que l'on trouve lorsqu'on examine le traitement discursif des noms de lieux.

b. Structure-type du contenu des noms de lieux

La deuxième classe de dénominations propres par l'importance, les toponymes, présente un fonctionnement discursif singulier. D'une manière générale, on peut dire qu'ils sont utilisés comme points de référence spatiale partagés par les interlocuteurs. Ils jouent donc leur rôle de « particuliers de base » pour reprendre la terminologie de P.-F. Strawson : *« Nous pouvons, les uns à l'égard des autres, préciser sur quelles choses particulières porte notre discours parce que nous pouvons intégrer les comptes rendus et les histoires de chacun dans un unique tableau du monde ; le cadre de ce tableau est un cadre unitaire spatio-temporel doté d'une dimension temporelle et de trois dimensions spatiales. Par conséquent, les choses étant ce qu'elles sont, l'identification des particuliers repose finalement sur la possibilité de localiser les choses particulières dont nous parlons à l'intérieur d'un unique système spatio-temporel unifié¹. »*

Néanmoins le précédent chapitre a permis de mettre en évidence deux autres aspects du fonctionnement discursif des toponymes dans les contextes à dominante syntactique des rapports d'activité. Le premier concerne leur catégorisation. Ni très fréquente, ni explicite de manière récurrente, celle-ci est utilisée par l'énonciateur afin de sélectionner un aspect caractéristique des référents des noms de villes et de pays : places financières pour les premières, et marchés pour les seconds. Par ailleurs, les noms de pays sont également utilisés afin d'ouvrir des univers de discours dans lesquels la restitution de l'activité prend place. Le travail d'élaboration de la structure-type, qui ne s'appuie pas sur un repérage général tel que celui qui est fourni par le calcul des co-occurents, a permis de préciser le cas qu'il convient de faire de ce dernier rôle des noms de pays. Contrairement aux odonymes et aux noms

¹ P.F. Strawson (1973 : 41-42).

de villes, la question de la sélection des noms de pays comme unités-pilotes destinées à réaliser une collecte de deuxième niveau s'est en effet posée. L'élaboration de la structure de contenu type de cette sous-classe, par l'examen détaillé du fonctionnement discursif sur lequel elle repose, a permis de reconsidérer les observations du chapitre 9.

En effet, celle-ci a montré que le rôle d'organisateur discursif assigné à certains noms de pays repose sur leur caractérisation comme marchés nationaux. Or, c'est parce qu'ils sont implicitement utilisés pour désigner des territoires économiques sur lesquels l'activité du groupe bancaire peut se déployer, qu'ils permettent une organisation des discours des rapports d'activité. De ce point de vue, ils fonctionnent plus comme des contenants que des contenus ou, pour reprendre la terminologie de l'ontologie qui a déjà été utilisée, ils ont un statut d'ensemble complexe et non de composant de base. Ces derniers doivent être préférés pour la collecte, dans la mesure où leur structure de contenu comporte, comme on l'a déjà vu à propos des dénominations propres de personnes morales, un rattachement à un ensemble complexe.

En raison même de ce statut, une collecte réalisée à partir des noms de pays se heurte à une difficulté majeure. Dans la mesure où ils jouent le rôle de cadre discursif, ils embrassent un contenu disparate, qui se perd en relations multiples sans qu'il soit possible d'en retenir valablement une plutôt qu'une autre. Comme l'a montré le commentaire des principaux co-occurrents d'une vingtaine de noms de pays et de « France », donné dans le précédent chapitre, les noms de pays permettent un découpage de l'activité où peuvent se rencontrer aussi bien diverses dénominations propres appartenant à des classes différentes que des thématiques variées allant du développement international du groupe bancaire aux ressources humaines, en passant par le développement durable. Pour l'élaboration de la structure de contenu on a donc retenu seulement les relations récurrentes, même si, comme cela a été rappelé, la relation de catégorisation est loin d'être toujours explicite. Le schéma présenté ci-dessus livre en quelque sorte un résumé de ces observations.

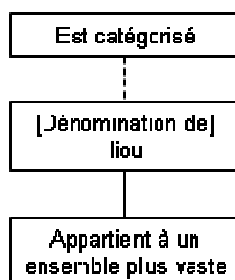


Figure 83.
Structure type du contenu de la classe des toponymes

Le fait que la catégorisation soit rarement explicite a été signalé à l'aide des pointillés qui relient le nom du lieu à la catégorisation de ce dernier. Par ailleurs, certains toponymes sont également présentés dans le discours et, plus rarement dans les contextes non syntactiques, comme étant rattachés à des ensembles géographiques plus vastes. Dans tous les cas, la pauvreté de la structure type du contenu des toponymes suggère que les éléments de cette classe ne constituent pas le meilleur accès au vocabulaire du corpus.

c. Structure type du contenu des dénominations de personnes physiques

Comme on l'a vu précédemment, les dénominations propres de personnes physiques sont majoritairement présentes dans les contextes à dominante non syntactique. Elles apparaissent ainsi dans des dispositifs de type fiche signalétique où elles sont mises en relation avec d'autres dénominations propres, par l'intermédiaire d'un ou de plusieurs titres. Comme cela est fréquent dans de tels contextes, les informations sont communiquées essentiellement par juxtaposition d'unités. Les dénominations de personnes physiques sont donc globalement coupées des autres composantes du vocabulaire des rapports d'activité. Dans l'exemple ci-dessous, tiré du rapport d'activité de 1995, on voit que la mention de titres permet de rattacher une dénomination propre « Marc Bué » à deux entités nommées : « *Marc Bué, Président de la Fédération nationale du Crédit agricole, Président de la Caisse régionale de Crédit agricole du Pas-de-Calais.* » Dans la dernière partie du corpus, on retrouve un type de relation identique, même si les entités nommées sont différentes : « *NORD DE FRANCE 10, avenue Foch 59000 Lille Tél. : 03 20 63 70 00. Président : Marc Bué, Directeur général : Alain Diéval.* » Ce type de relations a par conséquent été reporté dans la structure de contenu type de la classe :



Figure 84.

Structure type du contenu de la classe des dénominations de personnes physiques

En plus de ces mentions de base, l'année 1999 insère l'une des occurrences de la dénomination dans une sorte de *curriculum vitae*. Il est à noter que cet ajout exceptionnel se fait dans un contexte de type CDS, on passe en effet d'une présentation structurale, réservée aux CDN, à la mention d'un événement :

« Le Conseil d'Administration de la Caisse nationale de Crédit agricole a procédé, le 25 mai 2000, à la désignation de son nouveau Président, Marc Bué, qui succède ainsi à Yves Barsalou. Marc Bué, 61 ans, est agriculteur à Fillièvre, dans le Pas-de-Calais. Administrateur de la Caisse régionale du Pas-de-Calais en 1968, il en est devenu Président en 1978. Il a ensuite été désigné membre du Conseil fédéral (1986-1989), Vice-Président (1989-1992), puis Président de la Fédération nationale du Crédit agricole (1992-2000). Dans le même temps, il est devenu administrateur (1990-1999) puis Vice-Président (1999-2000) du Conseil d'Administration de la Caisse nationale de Crédit agricole. Membre du Conseil de surveillance de Crédit Agricole Indosuez depuis 1997, Marc Bué est également Vice-Président de la Confédération nationale de la mutualité, de la coopération et du Crédit agricole et Vice-Président du Groupe Alliance. Il est enfin membre de la section des finances du Conseil économique et social. Marc Bué [photographie]. »

Dans cet exemple, on voit que même lorsque la mention d'une personne physique a lieu dans un contexte syntactique, celle-ci reste relativement coupée des autres strates du vocabulaire du corpus relatives à l'activité, hormis peut-être en ce qui concerne le fonctionnement de l'institution. En effet, le contenu du paragraphe est comme clos sur lui-même. En raison de son statut de *curriculum vitae*, il fonctionne comme une parenthèse dans le reste du texte. Cette coupure est redoublée par l'apparition d'une photographie dont la dernière mention du nom dans l'extrait cité constitue la légende. L'élaboration de la structure de contenu confirme ainsi la spécialisation de la mention des dénominations de personnes physiques, qui en fait une classe relativement coupée du vocabulaire qu'il s'agit de collecter.

d. Structure-type du contenu des dénominations propres d'objets

Contrairement aux dénominations de personnes physiques, les dénominations propres d'objets sont majoritairement présentes dans les contextes syntactiques. Leur intégration dans la restitution de l'activité bancaire se traduit par la richesse de la structure de contenu qui a été identifiée pour leur classe. On retrouve certaines propriétés associées, caractéristiques de la structure type du contenu des dénominations propres de personnes morales, telles que le rattachement à un segment de clientèle, un marché ou un secteur économique, la catégorisation, la description et la quantification. Cela s'explique en partie par la présence majoritaire de dénominations de produits et de marques (Mozaïc, Compte Service, Open, TLJ) qui sont directement liés à l'activité économique des personnes morales évoquées dans le corpus. Cependant, les dénominations propres d'objets sont moins fréquentes et leur structure de contenu est moins riche que celle qui caractérise les dénominations propres de personnes morales.

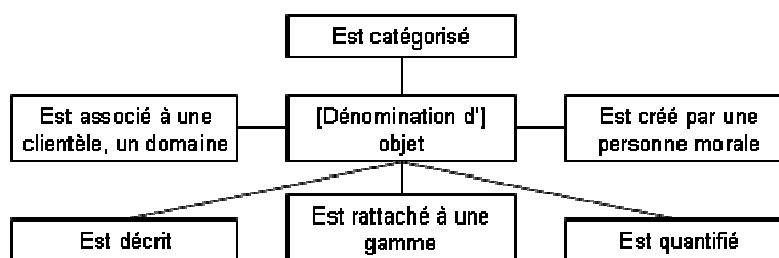


Figure 85.
Structure-type de la classe des dénominations propres d'objets

On donne ci-dessous un exemple prototypique dans lequel trois paragraphes suivis sont mobilisés dans le rapport d'activité de 1995 afin de présenter les différentes propriétés d'un nouvel l'objet, en l'occurrence, un contrat portant sur des services associés au compte bancaire :

« §1L'évolution des outils, notamment informatiques, du Crédit agricole lui permet aussi d'adapter son offre à la diversité des attentes des particuliers. Le lancement, à la fin de 1994, du "Compte Service Crédit agricole" a ainsi marqué une étape importante dans la volonté du Groupe d'établir une relation simple et claire avec ses clients.

§2 Conçu sous une forme évolutive, ce contrat repose sur la personnalisation et la construction d'une relation à long terme prenant en compte les événements de la vie d'un client et l'évolution de ses capacités financières. Reposant sur un socle de services de base, le Compte Service Crédit agricole propose le choix entre deux formules : le découvert contractualisé et le placement automatique des excédents de trésorerie.

§3 Chacune d'elles peut être enrichie par différentes options articulées autour de cinq pôles : la gamme des cartes, la gestion contractuelle de la trésorerie et de l'épargne, le crédit avec la carte Open et l'assurance-assistance. La montée en puissance, en 1995, du Compte Service a été rapide : à la fin de l'année, 550 000 contrats avaient été souscrits. (...) » Ces informations se répartissent ainsi dans les différentes branches de la structure type :

Tableau 77.
Structure-type des dénominations propres d'objets et informations collectées dans la partie 1995 pour « Compte service »

Propriétés de la structure type de contenu	Contextes d'apparition de la dénomination
Est catégorisé	§2 : ce contrat
Est créé par une personne morale	§1 : Crédit agricole
Est quantifié	§3 : à la fin de l'année, 550 000 contrats avaient été souscrits.
Est rattaché à une gamme	§1 : Compte Service Crédit agricole
Est décrit	§2 : Reposant sur un socle de services de base, le Compte Service Crédit agricole propose le choix entre deux formules : le découvert contractualisé et le placement automatique des excédents de trésorerie.
Est associé à une clientèle, un domaine	§1 : attentes des particuliers

Dans la mesure où la première année du corpus suit de peu la période de lancement du produit (décembre 1994), ce dernier fait l'objet d'une description particulièrement fouillée. Un phénomène similaire a déjà été observé au sujet des personnes physiques. Dans l'exemple présenté plus haut, la survenue d'un événement extraordinaire, en l'occurrence la désignation d'un nouveau président, a

entraîné l'apparition d'un développement particulier adapté aux circonstances dans un contexte inhabituel pour une dénomination de personne physique¹. Ici, l'apparition d'un nouveau produit fait l'objet de développements qui multiplient les occurrences de la dénomination propre (3 occurrences pour « Compte Service Crédit agricole »). Autre phénomène qui sera étudié plus en détail dans le dernier chapitre, la modification des informations associées est frappante avec cet exemple. Le contrôle effectué sur les occurrences rassemblées dans un paragraphe de la partie 2003 modifie en effet les données présentées dans le précédent tableau :

« La politique visant à adjoindre à tout compte de dépôt à vue une convention de Compte Service Crédit Agricole a été poursuivie avec succès en 2003 : plus de deux ouvertures de compte sur trois se concrétisent par l'équipement Compte Service. Englobant les services indispensables au quotidien (découvert, carte de paiement ou banque en ligne), ce contrat permet de fidéliser la clientèle, et l'incite à la domiciliation de ses revenus. Fin 2003, le nombre de contrats Compte Service Crédit Agricole dépasse 7,7 millions : 72 % des comptes courants actifs ont été équipés (soit + 450 000 sur un an). » On donne ci-dessous un tableau reprenant la structure type dont les sortes d'informations sont illustrées à l'aide des données collectées dans l'unique paragraphe où apparaissent des occurrences de « Compte Service ».

¹ Ces cas de figure, rencontrés au gré des exemples, peuvent être recherchés de façon plus systématique en observant la ventilation des fréquences d'une dénomination ainsi que sa répartition en fonction des deux sortes de contextes. En effet, compte tenu de la nature des rapports d'activité, ces variations sont généralement associées à un surcroît d'informations relatives à l'entité nommée étudiée. On verra dans le dernier chapitre le parti que l'on peut tirer de l'observation de ces variations pour cibler des zones de texte correspondant à des gisements d'informations.

*Tableau 78.
Structure-type des dénominations propres d'objets et informations collectées dans la
partie 2003 pour « Compte service »*

Propriétés de la structure type de contenu	Contextes d'apparition de la dénomination
Est catégorisé	<i>une convention l'équipement Compte Service ce contrat contrats Compte Service Crédit Agricole</i>
Est créé par une personne morale	<i>Crédit agricole (donné dans le paragraphe précédent)</i>
Est quantifié	<i>plus de deux ouvertures de compte sur trois se concrétisent par l'équipement Compte Service Fin 2003, le nombre de contrats Compte Service Crédit Agricole dépasse 7,7 millions : 72 % des comptes courants actifs ont été équipés (soit + 450 000 sur un an)</i>
Est rattaché à une gamme	<i>Par déduction : gamme « Compte Service ».</i>
Est décrit	<i>Englobant les services indispensables au quotidien (découvert, carte de paiement ou banque en ligne),</i>
Est associé à une clientèle, un domaine	<i>la clientèle [en général]</i>

Tout se passe comme si la richesse de la catégorisation de la dernière partie du corpus reflétait l'accumulation des discours produits en neuf ans sur le produit cité en exemple. On remarque néanmoins que seul « contrat » est commun aux deux tableaux, ce qui souligne la nécessité de « tenir le fil » de ces catégorisations qui dans un même contexte ont tendance à se cumuler plutôt qu'à s'exclure. Dans tous les cas, la richesse de la structure-type des dénominations d'objets indique qu'il est pertinent de retenir cette classe pour la collecte de vocabulaire. On notera cependant que la relative régularité de la structure type que l'on peut constater à l'échelle d'une dénomination propre est loin de s'appliquer à tous les éléments d'une même classe.

3. Structure-type et hiérarchisation des classes de dénominations propres

La structure-type élaborée à partir du fonctionnement discursif des principaux représentants d'une classe englobe bien les relations possibles pour les

dénominations moins fréquentes, mais son schéma ne s'applique pas à elles de façon exhaustive. On observe donc que pour les membres d'une même classe de dénominations propres, l'utilisation inégalement répartie de la structure-type du contenu vient compléter la hiérarchisation établie par la fréquence textométrique. Ce que ce constat laisse entrevoir, c'est non seulement que certaines classes semblent plus productives que d'autres pour la collecte du vocabulaire, mais aussi que cette inégalité se retrouve à l'échelle des unités appartenant à une même classe. Dans le tableau ci-dessous on illustre ce constat à l'aide d'exemples appartenant à la classe des dénominations propres de personnes morales.

*Tableau 79.
Richesse de la structure type de contenu et membres de la classe des dénominations propres de personnes morales*

Types de relations	CAI¹	Ucabail	Finistère	Air France	Areva
Est catégorisée	X	X	X		
Est localisée	X	X	X		
Son activité est présentée	X	X			
- Quantification des activités	X	X	X		
- Détail des activités	X	X			
- Ses domaines, secteurs, marchés	X	X			
Son organisation est présentée	X	X			
- Sa direction, son personnel	X	X	X		
- Ses relations capitalistiques	X	X		X	X

Afin d'illustrer l'application inégale du schéma de la structure type de contenu aux membres de la classe des dénominations propres de personnes morales, on a choisi cinq exemples d'entités ayant des profils variés. Il est assez normal de constater que pour « CAI » le schéma s'applique de façon exhaustive. Cela tient à ce

¹ « CAI » est le sigle de Crédit Agricole Indosuez (+ 500 occ.), « Ucabail » est la dénomination d'une filiale de crédit-bail (101 occ.), « Finistère » est la version courte de la dénomination complète « Caisse régionale de Crédit agricole du Finistère » ou « CRCAM du Finistère » (19 occ.), « Air France » (5 occ.) est le nom d'une compagnie aérienne française, « Areva » (1 occ.) est le nom d'une entreprise française opérant dans le domaine de l'énergie nucléaire.

que, d'une part, la dénomination est très fréquente et a été retenue à ce titre pour l'élaboration de la structure-type, et d'autre part, la complétude des sortes d'informations qui lui sont associées dans le corpus a déjà été vérifiée à travers l'exemple de collecte donné plus haut. Bien que sa fréquence soit cinq fois inférieure à celle de la dénomination précédente, « Ucabail », nom d'une autre filiale du groupe Crédit agricole, possède également l'ensemble des propriétés répertoriées dans la structure-type. Le champ d'application de la structure s'étend donc au moins dans certains cas au-delà des cinq unités utilisées pour l'élaborer.

Cependant, il semble qu'avec « Finistère », dénomination abrégée de « Caisse régionale de Crédit agricole du Finistère », une limite ait été franchie. Cette dénomination propre possède 19 occurrences et ne bénéficie que d'une partie des sortes d'informations recensées dans la structure-type. En ce qui la concerne, la présentation de l'activité est partielle et tient seulement dans une quantification très globale. Par ailleurs, on note que cette information disparaît à partir de 2001, comme la localisation, qui s'éclipse en 2002 avec la présentation des principaux membres de la direction. Cette irrégularité dans l'association des informations relatives à cette entité est à rapporter au fonctionnement plus général des contextes à dominante non syntactique. En effet, on a vu dans les précédentes parties que le type de contexte non syntactique est sujet à de brusques évolutions qui ont pour effet de faire disparaître certains éléments présents dans les dispositifs textuels qu'il recèle. Enfin, on notera que dans ces contextes la catégorisation est attribuée de façon indirecte, en quelque sorte par héritage : c'est le titre de la rubrique « Les Caisses régionales » qui s'applique de façon distributive à chacune des entités présentées. Il faut néanmoins remarquer que le régissant de la dénomination propre sert de classificateur dans les noms de caisses régionales : « Caisse régionale de Crédit agricole de...¹ ».

La situation des deux dernières dénominations propres de personnes morales est toute différente. La faible différence de fréquence n'a pas d'incidence sur l'application de la structure type de contenu. On n'apprend rien sur Air France ou Areva dans les rapports d'activité, si ce n'est que ce sont des entreprises pour

¹ Cette question a été traitée dans le chapitre précédent au sujet de l'application de certaines signalétiques aux dénominations propres comportant « Crédit agricole ».

lesquelles une entité du groupe Crédit Agricole a réalisé certaines prestations relatives à leur capital. On notera que la catégorisation est absente du contexte de ces dénominations, et que la nature de leur référent ne peut être déduite qu'à partir des prédications entrant dans leurs voisinages.

L'élaboration des structures types des principales classes de dénominations propres du corpus de rapports d'activité permet maintenant d'opérer certains choix afin de guider la collecte d'unités caractéristiques du vocabulaire du corpus.

C. Structure type et collecte d'informations

1. Choix parmi les relations composant les structures-types de contenu

Les étapes de la démarche qui ont été présentées dans les sections précédentes ont permis de préciser les caractéristiques des structures types de contenu de quatre classes de dénominations propres. On dispose maintenant de la possibilité de choisir parmi ces classes celles qui paraissent les mieux à même de satisfaire aux besoins de la collecte de vocabulaire. On rappelle que celle-ci doit permettre de faire émerger des éléments appartenant au vocabulaire caractéristique d'un parler d'entreprise et susceptibles de jouer un rôle dans l'organisation de l'information sur les sites d'un intranet. Cet apport peut être direct, lorsque les unités extraites sont utilisées pour opérer des classements de l'information, mais il peut également être indirect, lorsque ces mêmes unités fournissent des indications sur les strates du parler d'entreprise à utiliser dans une situation de déséquilibre cognitif entre émetteur et destinataire de l'information.

L'établissement des structures-types de contenu permet de distinguer quatre sortes de relations si l'on essaie de regrouper celles-ci en fonction de leur productivité présumée pour la collecte d'informations.

Parmi les relations les plus fréquentes, il y a celles qui mettent en relation des dénominations propres entre elles, à l'intérieur d'une même classe ou entre unités appartenant à des classes différentes. Ces relations présentent un intérêt encyclopédique certain, mais leur enseignement est relativement pauvre du point de vue de la collecte que l'on projette. Elles comprennent la localisation des personnes morales, la présentation de leurs états-majors ou la mention de liens capitalistiques ; elles s'étendent également au rattachement d'un toponyme à un toponyme englobant

(typiquement, « Europe » > « France »). Il en va de même pour les dénominations propres d'objets en ce qui concerne les relations « création par une personne morale » et « rattachement à une gamme ». Ces relations ne seront donc pas retenues à ce stade. Il en va de même pour une autre sorte de relations, celle qui est établie entre une dénomination propre et des données chiffrées, ce qui est le cas pour les structures types de contenu des dénominations propres de personnes morales et d'objets (ces chiffres sont utilisés pour des quantifications monétaires ou par unités).

Cependant, trois sortes de relations paraissent plus prometteuses dès lors qu'il s'agit de capter des strates du vocabulaire des rapports d'activité. Il s'agit de la catégorisation, du rattachement à un domaine (secteur ou marché) ou de l'association à une clientèle, et de la description de l'activité de la personne morale ou des caractéristiques d'un objet.

Cette dernière relation ne sera néanmoins pas retenue dans la mesure où les descriptions qu'elle regroupe sont dirigées vers un référent unique, ce qui présente deux inconvénients. Le premier tient à ce que chaque description risque d'être trop spécifiquement liée à un objet particulier et de ne pas présenter un niveau de généralité suffisant lui permettant d'être utilisée pour d'autres objets. D'autre part, comme cela apparaît déjà dans les paragraphes 5 à 7 et 1 à 4 donnés plus haut à titre d'exemples pour les dénominations propres de personnes morales et d'objets, ces descriptions comportent des éléments susceptibles d'être recensés par les deux autres relations de catégorisation et de rattachement à un domaine d'activité. On s'en tiendra donc à ces dernières pour la collecte du vocabulaire.

La relation de catégorisation présente un intérêt particulier pour cette collecte dans la mesure où elle marque l'intérêt porté par l'énonciateur, qui cherche par ce moyen à combler par anticipation les éventuelles lacunes de son destinataire. Mais aussi, elle donne accès à ce que ce même émetteur considère comme la catégorie de base « naturelle » qu'il convient d'utiliser dans cette situation de communication. Depuis les travaux, entre autres, de la psychologue Eleanor Rosch, on sait que pour les locuteurs d'une langue donnée il existe une hiérarchie lexicale « naturelle » à laquelle ces derniers recourent dans leurs échanges les plus quotidiens. Dans cette organisation tacite du lexique, le niveau de base, par exemple, « chien », intermédiaire entre le niveau superordonné, par exemple « animal » et le niveau

subordonné, par exemple « épagneul », occupe une position d'équilibre due à son caractère informatif suffisamment homogène pour lui permettre de catégoriser facilement un objet. En effet, le niveau superordonné ne possède pas en général suffisamment de traits distinctifs pour cela, et le niveau subordonné, le plus riche en traits distinctifs, présente l'inconvénient de présupposer la connaissance de la catégorie de niveau intermédiaire¹.

Enfin, le rattachement à un secteur, ou à un domaine d'activité, ou encore l'association à une forme de clientèle, constituent une deuxième source d'informations pertinente dans la perspective que l'on a adoptée. En effet, dans la mesure où ces unités sont utilisées pour organiser la restitution de l'activité du groupe bancaire, il est probable qu'elles puissent également être mises à contribution afin de procéder à des regroupements d'informations diffusées sur les sites intranets. Là encore, ces unités pourront être utilisées pour elles-mêmes et/ou pour fournir des indications sur le niveau de regroupement des données et les strates de parler d'entreprise perçus comme étant relativement adéquats à une situation de déséquilibre cognitif entre interlocuteurs.

2. Choix des classes de dénominations propres

A l'issue de cette analyse, seules deux classes de dénominations propres seront retenues pour la collecte de vocabulaire ; pour chacune d'elles, deux relations types présentes dans leur structure type de contenu ont été sélectionnées. On donne ci-dessous un aperçu de la sélection opérée à partir des schémas des structures type de contenu :

¹ On renvoie pour une présentation détaillée à G. Kleiber (1991 : 82 et suiv.). Par ailleurs, G. Kleiber a formulé l'hypothèse selon laquelle « *les particuliers porteurs d'un nom propre sont saisis prioritairement (et non nécessairement !) dans une catégorie basique.* » G. Kleiber (1996 : 582).

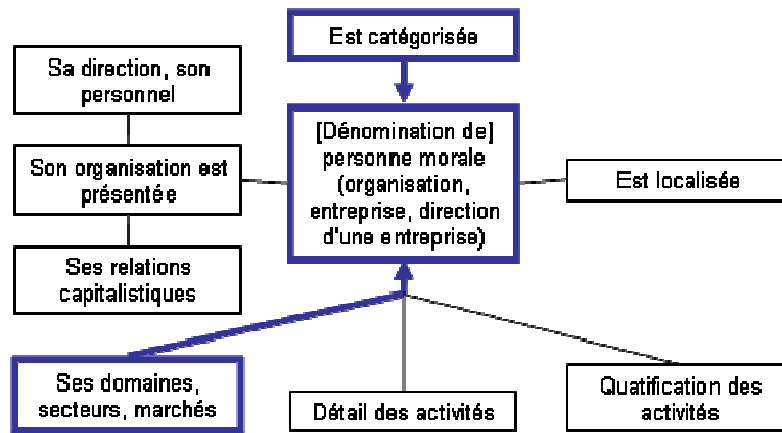


Figure 86.
Restrictions apportées à l'étude de la structure-type des dénominations propres des personnes morales

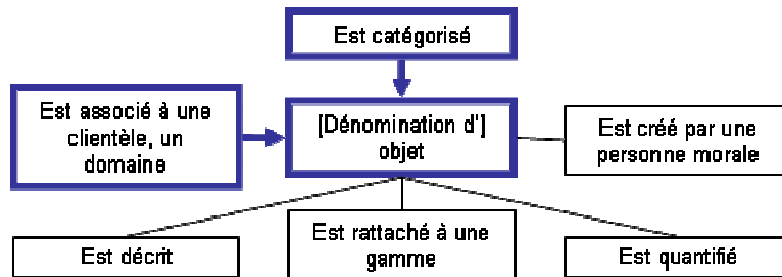


Figure 87.
Restrictions apportées à l'étude de la structure-type des dénominations propres d'objets

Le fait que certaines dénominations propres soient retenues comme moyen afin d'explorer le vocabulaire du corpus ne doit pas faire perdre de vue que les informations qui leur sont associées par les discours correspondent à des propriétés de leurs référents. Ce point est rappelé par le sens des flèches figurant dans les deux schémas donnés ci-dessus. De ce point de vue, la relation de catégorisation doit être lue ainsi : une organisation ou un objet dénommés « x » sont catégorisés de telle et telle manière dans le corpus. Il en va de même pour le rattachement à un domaine d'activité qui doit être compris de cette manière : une organisation « x » fait partie des ingrédients de tel domaine ou secteur d'activité dans le corpus. Enfin, l'association à une clientèle ou à un domaine quelconque résulte de l'application d'un schéma discursif du type « une entreprise /crée/ un produit /pour/ une clientèle ». On voit que dans les deux cas les flèches sont orientées vers la personne

morale ou l'objet, parce que c'est dans et par le discours que se fait l'apport d'informations nouvelles au contenu de la dénomination propre. Il conviendra dans le dernier chapitre de préciser les caractéristiques de ces relations de deux points de vue. Auparavant, il reste à préciser les formes linguistiques que peut prendre l'établissement de telles relations dans le discours. Par ailleurs, dans la mesure où l'on souhaite également intégrer les dénominations propres à un référentiel terminologique, il sera nécessaire de préciser par quelles relations sémantiques et à quelles conditions ce rattachement pourra être établi.

On a cherché dans ce chapitre à asseoir sur des bases théoriques explicites la collecte d'unités caractéristiques du vocabulaire du corpus à partir des dénominations propres. On rappelle que l'on souhaite utiliser ces unités afin d'accéder aux éléments constitutifs des classes-objets construites dans et par les discours. Ces informations, captées à travers les unités linguistiques qui les manifestent dans les textes, peuvent entrer dans la constitution du contenu des dénominations propres. Le périmètre de ce dernier est donc restreint à celui du corpus utilisé, même si celui-ci est un corpus de suivi organisé sous la forme d'une série textuelle chronologique. Par ailleurs, on a fait l'hypothèse qu'un tel contenu pouvait recevoir une organisation particulière conférée par la manière dont certaines informations sont associées aux dénominations propres dans les discours. En effet, le contenu des dénominations propres pourrait constituer une réponse à une forme de question topique relative aux informations qu'il convient de diffuser à propos d'un type d'entités dans un genre de discours donné. De la vérification de cette hypothèse dépendent plusieurs points. Tout d'abord, le fait qu'un tel contenu soit doté d'une certaine organisation est de nature à garantir l'établissement de relations sémantiques formalisées dans un référentiel terminologique. En second lieu, l'existence d'une telle structure rend possible l'adoption de critères de collecte explicites. On a donc proposé une méthode pour l'établissement des structures types du contenu des principales classes de dénominations propres recensées dans le corpus : noms de personnes morales, de lieux, de personnes physiques et d'objets.

La vérification de l'hypothèse de l'existence d'une structure type de contenu à conduit à distinguer des structures propres à chaque classe de dénominations propres. On a constaté par ailleurs que le « programme » de la structure type n'est pas appliqué de la même manière à tous les éléments d'une même classe. La classe des dénominations propres de personnes morales offre une forme d'articulation où la fréquence des dénominations entretient une relation assez étroite avec la richesse de la structure. En bref, plus la dénomination est fréquente, plus elle a de chances de réaliser dans le discours les différentes formes d'associations prédites par la structure. Cela se vérifie aussi pour la relation inverse : moins la dénomination propre a d'occurrences, plus il y a de risques pour que ce ne soit qu'une partie du

programme de la structure qui se trouve vérifié dans les discours. Ce que l'on fait ainsi apparaître c'est que la classe des dénominations propres de personnes morales est la classe par excellence qui est utilisée pour la restitution de l'activité bancaire dans les rapports d'activité. Cependant, son allure de prototype ne doit pas masquer d'autres possibilités attestées dans le corpus.

En effet, les observations relatives à la classe des toponymes semblent aller à l'encontre de ces remarques. Un point commun avec la classe précédente réside dans les écarts très importants entre les fréquences des différentes dénominations pour lesquelles on constate un rapport allant de 1 à 100 et au-delà. Cependant, cette hiérarchie textométrique n'a pas pour pendant une variation dans la richesse de la structure. Celle-ci est, si l'on peut dire, aussi pauvre pour les principaux représentants de la classe que pour ses plus humbles unités correspondant à des hapax.

La situation est encore différente chez les « classes moyennes », c'est-à-dire celles qui, quoique assez bien représentées dans le corpus, possèdent néanmoins des fréquences nettement moins élevées que celles des principales unités des deux classes précédentes. De fait, les écarts de fréquence y sont plus tassés. Dans la classe des objets, on retrouve néanmoins le fonctionnement observé pour les dénominations propres de personnes morales, à savoir que l'importance de la fréquence est globalement en relation avec la richesse de la structure. Il convient cependant de nuancer cette remarque, parce que l'on constate malgré tout une différence entre le traitement discursif des produits bancaires et celui des autres objets appartenant au domaine¹ : les premiers font en général l'objet d'une présentation plus complète que les seconds. Enfin, dernier cas de figure rencontré, le traitement discursif appliqué aux personnalités est le plus égalitariste de tous : les différences de fréquences sont relativement faibles au sein de cette classe et la structure s'y retrouve appliquée de façon assez uniforme. Seule la présentation de certains *curriculum vitae* vient étoffer occasionnellement la structure de contenu de certains noms de dirigeants.

¹ Pour les dénominations d'objets que l'on a retenues, il s'agit de « CAC 40 », indice boursier.

Si l'on rapporte ces observations à la notion de référentiel spécifique dont il a été question dans la première partie, on s'aperçoit que les relations entre fréquence, rôle et structure type du contenu de la dénomination propre peuvent s'éclairer différemment. La combinaison des indications fournies par ces trois caractéristiques des dénominations aide à mieux saisir ce qui est en jeu dans les rapports d'activité : à l'intérieur d'un référentiel spécifique, et étant donné l'importance que cela revêt dans telle situation de communication, quels objets convient-il de valoriser ? Ordinairement latente, la fonction de la structure type de contenu se révèle alors pleinement, comme pense-bête argumentatif, ou comme question topique : « s'il est question de telle entité lors de la restitution de l'activité du groupe bancaire, il ne faut pas oublier de parler de telle ou telle de ses propriétés et de taire telles autres ». En définitive, les rapports d'activité peuvent se lire à travers leurs dénominations propres comme la manifestation renouvelée chaque année d'une ontologie en devenir. Celle-ci est donnée comme une organisation hiérarchisée du monde, dans laquelle les entités sont ordonnées par ordre d'importance : les personnes morales agents économiques, les produits de leur activité, les marchés, les débouchés et le reste du monde. Si les forces productives sont quelque peu occultées, qu'à cela ne tienne, elles sont « représentées » par les états-majors : les rapports d'activité sont des lieux de pouvoir où la mention du nom fonctionne aussi comme signe de distinction.

On voit donc que la combinaison de ces trois sortes d'indications, fréquence, rôle, structure type du contenu, permet d'identifier parmi l'ensemble des dénominations propres, les classes, puis les unités à l'intérieur de ces classes, qui constituent les points d'accès potentiellement les plus productifs en termes de collecte d'information. En effet, c'est parce qu'une classe de dénominations et, au moins certains de ces membres, sont suffisamment immergés dans les différentes manifestations des dispositifs textuels, qu'ils posséderont une meilleure capacité à capter le vocabulaire du corpus. La méthode proposée permet de partir immédiatement des dénominations propres du corpus pour collecter le vocabulaire, mais elle gagne à ce que soient repérées les sortes de dénominations propres qui sont plus particulièrement utilisées dans le genre de discours dont le corpus constitue un échantillon. Ce premier repérage fournit alors l'indication selon laquelle telle classe

de dénominations propres, ou tout au moins certains de ses membres, devraient être privilégiés pour procéder à la collecte du vocabulaire. Pour mener à bien la collecte envisagée on a retenu deux sortes de relations : celle qui opère une catégorisation du référent de la dénomination propre et celle qui le rattache à un secteur d'activité, à un marché ou à une forme de clientèle. Les deux classes retenues sont également celles pour lesquelles ces relations sont avérées. Il s'agit de la classe des dénominations de personnes morales et de celle des noms d'objets. Il convient d'aborder maintenant la réalisation de la collecte de deuxième niveau et de discuter les résultats qu'elle a permis d'obtenir.

CHAPITRE 11.
UN RÉFÉRENTIEL EN MOUVEMENT

Dans les chapitres précédents ont été abordées les questions soulevées par l'identification des dénominations propres présentes dans le corpus. Leur collecte systématique a permis d'avoir une idée précise sur la place occupée par cette catégorie d'unités dans le vocabulaire du corpus. Il a alors également été possible d'associer les unités et leurs variantes, ce lien étant formalisé par une relation d'équivalence, telle que l'on en rencontre dans les référentiels terminologiques. Ce travail préparatoire réalisé, plusieurs classes de dénominations propres pouvaient être utilisées afin de réaliser la collecte d'unités terminologiques représentatives du parler d'entreprise du Crédit Agricole, tel qu'il se manifeste dans les discours des rapports d'activité. La sélection de certaines classes de dénominations propres a nécessité que l'on précise la manière dont leurs éléments fonctionnent dans les textes du corpus. L'étude des toponymes a permis de montrer que le fonctionnement discursif des dénominations propres peut s'exprimer sous la forme de rôles différents endossés par ces unités au fil des textes. Ces rôles manifestent l'actualisation de certaines propriétés du référent de la dénomination propre, une même dénomination étant susceptible de jouer plusieurs rôles dans le même texte.

On s'est alors demandé si la présentation de ces propriétés obéit à une certaine structuration dans le genre de discours étudié. On a alors fait l'hypothèse que le contenu des dénominations propres en discours est organisé à la manière d'un topos pour un genre donné. Cette structuration a permis d'opposer les classes de dénominations propres en fonction de la richesse des relations qu'elles établissent avec les autres unités présentes dans les textes du corpus. Deux classes ont alors été retenues afin de collecter deux sortes d'informations : celles qui visent à catégoriser les référents des dénominations propres, et celles qui les rattachent à des formes de clientèles ou à des secteurs d'activité. Ce chapitre traite des différentes questions soulevées par la réalisation de la collecte au sens large, et propose une évaluation des résultats obtenus au regard des objectifs fixés au début de ce travail.

On présentera tout d'abord la manière dont cette collecte a fait l'objet d'un travail préparatoire, destiné à calibrer au mieux l'usage des techniques textométriques utilisées. La collecte devant concerner plusieurs centaines d'unités, il s'est avéré nécessaire d'identifier des procédures fiables et reproductibles, adaptées

à toute la gamme des fréquences associées aux dénominations propres du corpus. Il s'agira ensuite de présenter les questions soulevées par la réalisation de la collecte elle-même. En effet, la maîtrise de l'approche exploratoire qui a été utilisée repose sur la collecte d'indices syntaxiques, devant permettre de repérer et de valider les relations sémantiques étudiées. Il conviendra alors d'aborder les difficultés qui se présentent, dès lors que l'on cherche à formaliser ces relations afin d'intégrer les dénominations propres dans un référentiel terminologique. Enfin, les résultats de cette collecte seront évalués sur plusieurs plans. La caractérisation des unités collectées permettra de préciser dans quelle mesure cette approche a permis, d'une part, de capter une image fidèle du vocabulaire du corpus, et d'autre part, de mettre au jour une organisation de la restitution de l'activité bancaire qui soit plus proche des usages réels que d'un modèle théorique. On se demandera alors dans quelle mesure cette approche permet également de saisir les évolutions affectant le référentiel d'une institution en permanente transformation.

I. ORGANISATION DE LA COLLECTE

A. Unités-pilotes

1. Présentation de la démarche

Deux classes de dénominations propres ont été retenues pour la collecte : les dénominations propres d'objets et de personnes morales. En effet, on a vu plus haut que les structures de contenu de ces deux classes comportent deux relations dignes d'intérêt pour leur exploitation dans une perspective d'organisation de l'information : la relation de catégorisation du référent de ces dénominations ainsi que son rattachement à une ou plusieurs clientèles ou secteurs d'activité.

Les buts assignés à cette collecte, donner une image fidèle des manifestations d'un parler d'entreprise dans les textes du corpus, éclairer la manière dont s'effectue le découpage de l'activité bancaire dans ce contexte et permettre de suivre les évolutions affectant le référentiel d'une institution, ne nécessitent pas qu'elle soit exhaustive. On vise plutôt un résultat qualitatif qui induit une approche du même ordre : il s'agit de choisir, parmi les catégories les plus sollicitées dans les rapports d'activité, les représentants à même de fournir la collecte la plus riche en termes d'informations collectées. C'est pourquoi on s'attachera tout d'abord à fixer la liste des unités pilotes qui ont été utilisées pour guider la collecte. Cependant, celle-ci devra pouvoir en cas de besoin être conduite à partir de n'importe quel type de dénominations propres, aussi bien les plus fréquentes que celles qui ont un statut de hapax dans le corpus. Il s'agira donc dans un deuxième temps de présenter les procédures textométriques qui ont été utilisées pour répondre aux différents cas de figure qui ont été rencontrés. Une deuxième finalité de cette présentation de l'organisation de la collecte se fera jour : s'agit en effet de démontrer la possibilité d'utiliser des procédures explicites et dans la mesure du possible, outillées, pour

réaliser une collecte de vocabulaire semi-automatisée¹. C'est pourquoi on s'attachera dans cette section, comme dans les suivantes, à décrire précisément comment certains résultats ont été obtenus. Cela apparaît, semble-t-il, comme un préalable nécessaire aux évaluations qui seront présentées à la fin du chapitre.

Dans l'approche que l'on propose, la collecte de vocabulaire prend appui sur un certain nombre de dénominations propres. Celles-ci ont été sélectionnées en fonction de propriétés liées à leur utilisation dans le corpus, elle-même cernée grâce à l'identification de rôles et de structures types de contenu. Cette sélection a un double but : d'une part, intégrer les dénominations propres à un référentiel terminologique, et d'autre part, enrichir ce dernier avec les unités collectées à partir des dénominations propres. C'est la conjugaison de ces orientations qui a conduit à privilégier deux classes de dénominations propres : les dénominations propres de personnes morales et d'objets. À l'intérieur de ces deux classes, on a choisi de retenir les cinquante premières unités classées en tête lorsque les listes auxquelles elles appartiennent sont triées par ordre inverse de fréquence. Le fait de retenir les unités ayant le plus grand nombre d'occurrences présente l'intérêt d'aborder les cas les plus complexes. Néanmoins, on s'est astreint à vérifier par des sondages que les observations conduites à partir de ces échantillons valaient au moins pour l'ensemble de la classe. En effet, dans le cadre de ce travail on vise moins un recensement exhaustif que la démonstration, d'une part, qu'une telle collecte est viable, et d'autre part, que les résultats qu'elle permet d'obtenir peuvent être évalués. On donne ci-dessous les deux listes d'unités utilisées pour la collecte.

2. Dénominations propres d'objets utilisées pour la collecte

La liste des dénominations propres d'objets utilisées pour la collecte recense les cinquante-deux premières dénominations par ordre de fréquence décroissante. Celles-ci s'échelonnent de trente-neuf occurrences pour « Mozaïc », à trois

¹ On considère en effet que l'explicitation des procédures utilisées est nécessaire, sans impliquer obligatoirement un codage sous la forme de programmes informatiques ; on sait que ces derniers ont tendance à rigidifier la collecte en fixant la procédure pour un corpus donné ou en fonction d'une certaine conception linguistique (c'est le cas pour les traitements d'extraction automatique de terminologie, mais aussi en matière de textométrie, comme on a pu le voir au début de cette partie à propos des noms propres).

occurrences pour les vingt dernières dénominations retenues. En fait d'objets, la liste se compose très majoritairement de noms de produits et de services vendus par le groupe bancaire. En cela, l'extrait ci-dessous est représentatif de la composition de la liste complète, où l'on compte environ 10% d'objets divers pour 90% des dénominations propres de marques, de produits et de services¹. Pour le détail des fréquences, on renvoie à la liste complète donnée en annexe et signalée dans le premier chapitre de cette partie.

*Tableau 80.
Liste des cinquante-deux dénominations propres d'objets utilisées pour la collecte triées par ordre alphabétique*

Agilor	Dynatop	Master Sélection	Santé
Ascencia	Détente Jardin	Monéo	TLJ
Avantage	Edrane	Mozaïc	Talisman
CA Certificat	Export	Open	Tiwi
CA TV	Filfax	Optim	Trialto
CAC 40	Filvert	Orchestral	Télécommerce
Carissime	Floriane	PREDIprofessionnels	Unicash
Carte des Caisses régionales	GAV	Pacte Solidarité Logement	carte Mozaïc
Compte Service	Green Way	Pep s	carte Mozaïc de paiement
Compte Service Mozaïc	Immedians	Pleinchamp	ratio Cooke
Compte Service Professionnel	Lionvie	Practis	
Confluence	Livrets Jeunes	Prediagri	
Créances Clients Service	MEDD	SBF 120	
DECRET DU 6 AVRIL 1988	Maestro	Santeffi	

Les dénominations d'objets, comme la plupart de dénominations propres, possèdent un certain nombre de variantes qui n'ont pas été listées dans ce tableau où seules les formes de référence ont été retenues. On rappelle que les variantes ont été abordées dans le premier chapitre de cette partie ; elles seront à nouveau évoquées dans la section suivante au sujet de la description des différentes phases de la collecte.

3. Dénominations propres de personnes morales utilisées pour la collecte

La liste des dénominations de personnes morales utilisées pour la collecte recense cinquante-cinq unités retenues principalement en fonction du nombre de

¹ On dénombre dans la liste des dénominations d'objets 252 entrées dont 31 correspondent à des objets divers et 221 à des dénominations propres de marques, de produits et de services; dans l'extrait donné ci-dessous, on compte pour 52 dénominations propres, 5 objets divers et 47 dénominations de produits et de services. Dans les deux cas, on constate un rapport de un à dix environ.

leurs occurrences qui vont de plus d'un millier pour « Crédit Agricole » à une vingtaine pour « Unicredit ». Comme pour les dénominations propres d'objets, le fait de retenir les fréquences les plus élevées permet d'aborder les principales questions soulevées par ce type de collecte. En effet, les unités de fréquence moins élevée sont beaucoup plus nombreuses, ce qui entraîne deux conséquences. La première est d'allonger démesurément le temps de traitement, ce qui serait justifié dans un contexte purement terminographique, mais n'apporterait pas de conclusions supplémentaires dans le cadre de ce travail. Par ailleurs, et c'est la deuxième conséquence, cela multiplierait les cas de figures pour lesquels au moins l'une des deux relations de la structure-type ne serait pas attestée, comme cela a déjà été présenté dans le chapitre 10. En effet, on a pu alors observer que les classes de dénominations propres sont hiérarchisées en fonction des fréquences mais aussi en fonction de la réalisation du programme de la structure type de contenu. Moins la fréquence est élevée, moins il y a de chances pour que la structure type de contenu soit réalisée, y compris en ce qui concerne les deux relations retenues pour la collecte du vocabulaire. On donne ci-dessous la liste des unités-pilotes utilisées dans le cadre de ce travail.

Tableau 81.

Liste des cinquante-cinq dénominations propres de personnes morales utilisées pour la collecte

BES	CNCA	Crédit Lyonnais	Intesa
BFT	Caisse régionale	EFL	IntesaBci
BGP	Caisses locales	Etat	Lukas
Banco Acac	Caisses régionales	FNCA	Pacific
Banco Ambrosiano Veneto	Calyon	Fastnet	Predica
Banco del Desarrollo	Carr Futures	Finaref	Segespar
Banque Commerciale de Grèce	Cedicam	Fondation du Crédit Agricole Pays de France	Sofinco
BGPI	Conseil	Groupe	Sofipar
Bisel	Crédit Agricole	groupe Crédit Agricole	Transfact
Bureaux permanents	CA AM	groupe Crédit Agricole S A	UI
CAC	Crédit Agricole Indosuez	IDIA	Ucabail
CAES	CAI Cheuvreux	Indocam	Unicredit
CAL FP	Crédit Agricole S A	Indosuez	Unifica
CLAM	Crédit Foncier de Monaco	Indosuez W I Carr Securities	

La liste de dénominations propres rassemblées dans ce tableau appelle un certain nombre de commentaires. Sa constitution s'appuie sur la phase d'identification des dénominations propres et, comme on l'a indiqué plus haut, le critère de sélection a été principalement la fréquence (les unités retenues ont une

fréquence supérieure ou égale à 20). Lors de cette étape préalable à la collecte, on a toléré l'entrée de dénominations qui n'ont parfois de « propre » que la présence d'une majuscule dans l'un au moins de leurs éléments ; il en va de même pour certains pluriels dont la présence dans une liste de dénominations propres peut paraître incongrue. On rappelle que pour la première phase d'identification et de recensement des dénominations propres, on a pris le parti de collecter de manière assez large afin de prendre en compte la forme d'usage qui consiste en quelque sorte à hausser certaines unités au-dessus de leur classe en leur attribuant une majuscule non prédictible en langue ou dans l'usage grammatical courant. On se réserve ainsi la possibilité de réaliser un examen attentif du fonctionnement discursif de ces unités¹ qui, dans le cadre de la collecte, pourront éventuellement servir de contre-exemples lorsque des comportements caractéristiques seront identifiés.

B. Une approche textométrique de la collecte

1. Traitement textométrique appliqué aux unités de faible et de moyenne fréquence

Dans l'approche textométrique de la collecte de vocabulaire que l'on va exposer, la fréquence des unités étudiées va déterminer les sortes de traitement que l'on va leur appliquer. D'expérience, la collecte, parce qu'elle repose sur une extraction d'informations manuelle réalisée à partir de nombreux contextes, nécessite une assistance renforcée dès lors que les occurrences de l'unité étudiée dépassent la vingtaine. La procédure proposée vise donc à simplifier la masse d'informations à prendre en compte, tout en préservant la qualité d'une extraction réalisée à même les textes. Le traitement des unités relativement peu fréquentes sera illustré à l'aide des dénominations propres d'objets, et les dénominations propres de personnes morales permettront d'illustrer le traitement appliqué aux unités de fréquence moyenne à élevée.

Le traitement des unités de faible fréquence a pour point de départ la sélection de l'une des dénominations figurant dans la liste présentée plus haut. Un

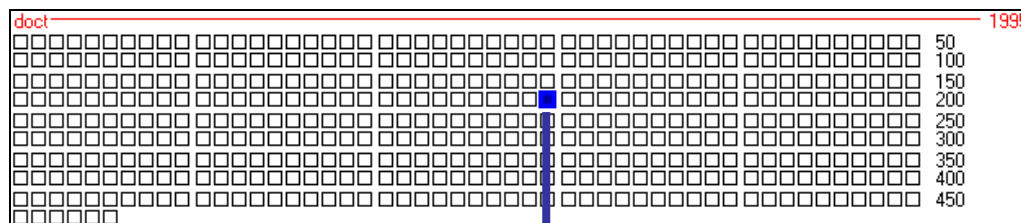
¹ « Bureaux permanents », « Caisse régionale », « Caisses régionales », « Caisses locales », « Conseil », « Groupe », « groupe Crédit agricole » et « groupe Crédit Agricole SA ».

contrôle est alors opéré afin de vérifier si la liste complète utilisée pour l'extraction ne comporte pas de variantes qui n'auraient pas été rapprochées lors de la phase d'identification et de recensement. On constate généralement que le nombre de variantes augmente avec la fréquence de l'unité étudiée. De ce point de vue, la liste constituée pour présenter la collecte n'est pas représentative de la liste complète des dénominations propres d'objets, dans la mesure où elle rassemble les dénominations ayant les fréquences les plus élevées. Ce sont seulement six unités qui possèdent des variantes sur une liste qui en comporte une cinquantaine, mais on observe une proportion bien inférieure à 10% d'unités concernées par des variantes pour le reste de la liste (dénominations ayant une fréquence inférieure à 3¹). Les variantes sont rassemblées dans un groupe de formes du logiciel Lexico. Un tel groupe permet de manipuler les variantes d'une dénomination comme un seul ensemble.

Parallèlement, les textes du corpus sont représentés sous la forme d'une carte de sections. Cette autre fonctionnalité du logiciel Lexico a déjà été présentée dans les précédents chapitres. On rappelle seulement ici que les sections correspondent à des segments du texte de taille inférieure, égale ou supérieure à la phrase. Pour des raisons qui ont été précisées antérieurement et qui ont trait au fonctionnement discursif des dénominations propres, c'est le paragraphe qui a été retenu comme unité pour le découpage en sections des textes du corpus. Les paragraphes sont à leur tour regroupés en ensembles plus vastes correspondant à l'une des partitions du corpus. Pour ce qui concerne les dénominations propres d'objets, dont on a vu plus haut qu'elles sont très majoritairement présentes dans les contextes à dominante syntactique, c'est la partition principale du corpus qui a été retenue et qui reprend, sans les scinder en contextes syntactiques et non syntactiques, les textes des différents rapports d'activité dont le corpus est composé. Le groupe de formes est alors projeté sur la carte des sections. Les paragraphes contenant au moins l'une des unités recensées dans le groupe sont distingués des autres par une coloration

¹ Cela s'explique en partie par le fait que la fréquence trois joue un rôle charnière pour la répartition d'une unité dans les différentes parties du corpus. En effet, s'il arrive fréquemment que pour un même sujet la dénomination de l'objet soit sollicitée à deux ou trois reprises, quatre occurrences correspondent plutôt à une répartition sur deux parties, voire trois. Le fait d'apparaître dans plus d'une partie, autrement dit, dans deux rapports d'activité ou plus, génère fréquemment l'apparition de variantes, surtout lorsque la dénomination propre correspond à une unité complexe.

spéciale. On illustre ce point à l'aide de la première unité de la liste « Agilor », qui comporte six occurrences dans le corpus.



*Dans le même temps, **Agilor**, crédit prescripteur développé avec les constructeurs et les distributeurs de matériel agricole, a connu un succès important. Les demandes ont porté sur 3,6 milliards de francs, en augmentation de 58% par rapport à 1994.*

Figure 88.

Section sélectionnée par la présence d'une occurrence de l'unité « Agilor » dans la première partie du corpus, et texte du paragraphe correspondant

La lecture du paragraphe dans lequel au moins une dénomination propre d'objet est présente fournit ensuite les indications nécessaires à la collecte des unités associées à la dénomination propre par les deux types de relations qui ont été présentées dans le chapitre 10. Ce point sera développé dans la prochaine section. Il reste à présenter la démarche adoptée lorsque les occurrences sont beaucoup plus nombreuses et se comptent par dizaines ou centaines. On devine que dans ce cas, le dépouillement systématique des paragraphes est non seulement fastidieuse, mais surtout source d'erreurs et d'oublis. En effet, il est facile de comprendre que plus la quantité d'information est élevée, plus elle est délicate à traiter sans assistance logicielle.

2. Traitement textométrique appliqué aux unités de fréquence moyenne à élevée

a. Méthode adoptée

Le traitement textométrique appliqué aux unités de fréquence moyenne (à partir de vingt ou trente occurrences) ou élevée (plusieurs centaines d'occurrences), vise avant tout à simplifier la lecture de la masse d'informations qui se présentent lors de la phase de collecte. Les traitements lexicométriques, et plus particulièrement le recours aux calculs statistiques, n'ont pas d'autre finalité que de permettre ce genre de simplification. Les premières étapes du traitement appliqué à ce type

l'inconvénient qu'il y aurait à appliquer la même méthode sans tenir compte de l'importante différence de fréquences. On a donc cherché à simplifier leur lecture.

b. Constitution d'un corpus de deuxième niveau

Pour cela, on s'appuie sur le calcul des co-occurents. Dans Lexico, les co-occurents d'une forme ou d'une polyforme sont obtenus par la méthode des spécificités. Schématiquement, celle-ci va comparer le contenu des paragraphes sélectionnés par les occurrences des unités appartenant au groupe de formes, avec les autres paragraphes du corpus. Les formes et polyformes apparaissant plutôt dans les paragraphes sélectionnés sont affectées par ce calcul d'un coefficient de spécificité positif et, à l'inverse, celles qui sont peu présentes ou absentes des paragraphes en question sont affectées d'un coefficient de spécificité négatif. Les co-occurents appartiennent aux formes et polyformes dotées d'un coefficient de spécificité positif. Plus celui-ci est élevé, plus la forme ou polyforme peut être considérée comme représentative des paragraphes où les unités du groupe de formes sont également présentes. Il reste que ce calcul, lorsqu'il est appliqué à la recherche des co-occurents des dénominations propres, mais également dans d'autres configurations, produit un résultat très volumineux de formes et polyformes co-occurentes. En effet, toutes les formes simples et les segments répétés de fréquence différente sont rapportés par ce calcul des co-occurents. Si l'on tient compte des seules formes dotées d'un coefficient de spécificité positif, on obtient pour l'exemple précédent 1380 co-occurents ; si l'on restreint cette liste aux formes dotées d'un coefficient supérieur ou égal à trois, on réduit la liste à 849 co-occurents. Outre la longueur de la liste, la lecture du résultat est rendue difficile par le caractère redondant des informations rapportées. On donne ici, à titre d'illustration, les vingt premières formes et polyformes co-occurentes du groupe « BES », présentées par coefficient de spécificité décroissant.

Tableau 82.
Principaux co-occurents du groupe de formes « Banco Espirito Santo »

Forme	Frq. Tot.	Fréquence	Coeff.
Santo	34	34	***
Banco	167	62	***
Portugal	54	31	46
BES	21	21	41
Espirito	19	19	37
Banco Espirito Santo	19	19	37
Espírito	15	15	30
Espírito Santo	15	15	30
Banco Espírito Santo	13	13	26
Grèce	55	20	25
Tranquilidade	12	10	18
Téléphone	34	13	17
Credibom	10	9	17
Portugal Banco	8	8	16
Portugal le	8	8	16
Fax	38	13	16
Au Portugal	8	8	16
le BES	8	8	16
de Grèce	26	10	14
Santo BES	7	7	14

Dans ce tableau présentant les vingt premières formes et polyformes co-occurentes du groupe de formes « Banco Espirito Santo », on retrouve en bonne position les formes et polyformes composant ce groupe. Cela vient du fait que le calcul s'applique également à elles, puisqu'elles font naturellement partie des paragraphes qu'elles ont permis de sélectionner ! Elles occupent à elles seules la moitié des formes les plus spécifiques. Mais ce caractère redondant concerne également d'autres unités, comme « Portugal », présent quatre fois. Enfin, la liste propose des dénominations propres comme « Grèce », « Credibom » ou encore « Tranquilidade », dont on a vu plus haut qu'elles présentent un intérêt limité pour la collecte envisagée, ainsi que des unités moins attendues comme « Fax » et « Téléphone », caractéristiques des contextes non syntactiques¹.

C'est dans le but de pallier cet inconvénient pour la réalisation d'une collecte du type de celle que l'on envisage, que l'on a été amené à proposer une étape supplémentaire dans le traitement. Comme les lignes précédentes le laissent

¹ Voir le chapitre 3.

entendre, il s'est agi de trouver une manière de simplifier la lecture de la liste des co-occurents, tout en préservant la possibilité de contrôler en contexte les propositions du calcul des spécificités. Pour atteindre ce but, on a constitué un corpus de deuxième niveau, dont toutes les données proviennent du corpus servant à réaliser la collecte. Ce corpus de deuxième niveau est établi à partir des listes de co-occurents obtenues au cours de l'étape qui vient d'être décrite. Pour chaque dénomination propre de personne morale appartenant à l'une des deux listes utilisées pour la collecte, on constitue un groupe de formes comprenant ses variantes. Le groupe est projeté sur la carte des sections et le calcul des spécificités est lancé. Mais au lieu de chercher à exploiter directement le résultat de ce premier calcul, on sélectionne tous les co-occurents ainsi obtenus lorsqu'ils sont dotés d'un coefficient de spécificité égal ou supérieur à 3. Ce seuil présente l'avantage d'être suffisamment discriminant, tout en fournissant un extrait substantiel de la liste complète des co-occurents. Un corpus¹ est alors constitué avec les stocks de co-occurents de toutes les unités pilotes. Ce corpus est appelé C² par convention, pour « corpus de deuxième niveau ».

c. Contrôle de l'homogénéité des stocks de co-occurents du corpus C²

La procédure d'extraction se déroule alors en plusieurs étapes que l'on va décrire brièvement. Dans la mesure où le corpus C² est composé de parties correspondant aux stocks des co-occurents des unités pilotes, il est possible d'appliquer à ces ensembles regroupant des segments répétés et des formes simples une analyse factorielle des correspondances. On obtient ainsi une carte des différentes parties qui se rassemblent par affinités de vocabulaire. On reviendra ultérieurement sur ce point qui constitue ici une étape technique obligée. Il est alors possible, avec le logiciel Lexico, de lancer un calcul des spécificités s'appliquant à un ou plusieurs stocks de co-occurents appartenant au corpus C² auquel une AFC a été appliquée. Dans la procédure normale, on commence par sélectionner un stock particulier, mais il est possible, dans un deuxième temps, de chercher à établir des

¹ Caractéristiques textométriques du corpus de deuxième niveau (C²) : 55 parties, 162273 occurrences pour 4909 formes, dont 1049 hapax. La fréquence maximale y est de 11 399 pour la forme « de ».

regroupements et de connaître en quoi certains ensembles de stocks de vocabulaire s'opposent à d'autres ensembles, également visibles sur le graphique présentant les résultats de l'AFC. À ce stade, ce calcul va permettre de contrôler l'homogénéité de l'ensemble des unités pilotes utilisées pour la collecte. On donne ci-dessous le résultat d'une AFC pratiquée sur le corpus C² composé des stocks des co-occurents des cinquante-cinq unités pilotes listées dans la précédente section.

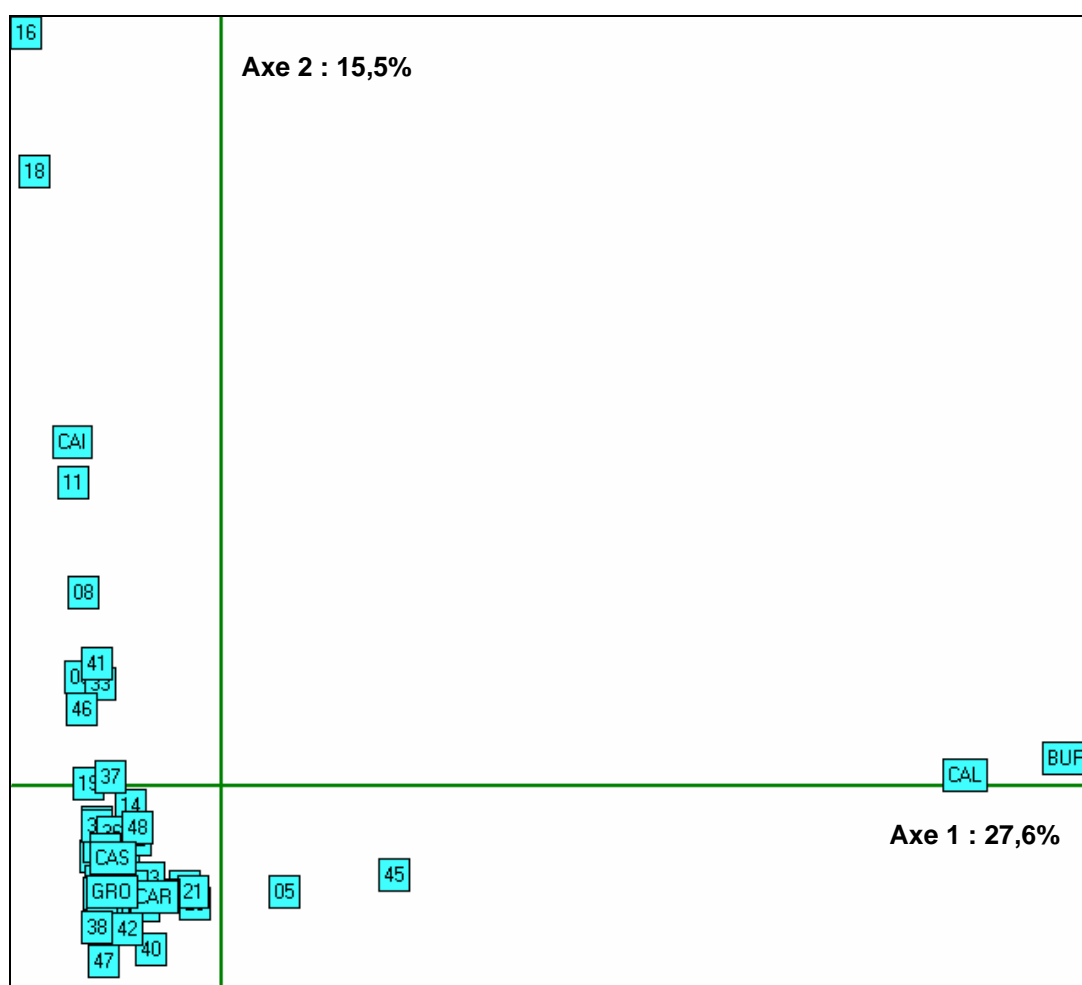


Figure 90.
AFC appliquée au corpus C² composé de 55 stocks de co-occurents

Lecture de la figure 90 : le graphique est organisé autour de deux axes représentant les deux principaux facteurs calculés par l'analyse factorielle des correspondances pour le corpus C². L'axe 1 horizontal représente 27,6% de l'information extraite du corpus, alors que l'axe 2 vertical en représente 15,5%. Le calcul a porté sur les formes dont la fréquence est égale ou supérieure à 10. Le premier axe oppose pratiquement l'ensemble des stocks de co-occurents à quatre d'entre eux situés à droite de l'axe 2. Le second axe rassemble, en bas dans le coin

gauche de la figure, l'essentiel des stocks de co-occurents. Chaque étiquette donnée entre parenthèses correspond au stock des co-occurents de l'une des formes-pilotes suivantes :

(01) Sofinco	(15) Indosuez	(29) Fond. Pays de France	(43) Sofipar
(02) Crédit Lyonnais	(16) IWICS	(30) CLAM	(44) Unicredit
(03) Indocam	(17) Transfact	(31) BAV	(45) Unifica
(04) Predica	(18) Carr Futures	(32) Etat	(46) CFM
(05) Caisse régionale	(19) Calyon	(33) CAL FP	(47) Banco Acac
(06) groupe Crédit Agricole	(20) FNCA	(34) IntesaBci	(48) CAES
(07) CNCA	(21) Conseil	(35) BCG	(BUR) Bureaux permanents
(08) CA AM	(22) BGP	(36) Segespar	(CAG) Crédit Agricole
(09) Ucabail	(23) CAC	(37) BGPI	(CAI) CAI
(10) Pacifica	(24) Cedecam	(38) EFL	(CAL) Caisses locales
(11) CAI Cheuvreux	(25) groupe CA S A	(39) IDIA	(CAR) Caisses régionales
(12) UI	(26) Lukas	(40) Banco del Desarrollo	(CAS) Crédit Agricole S A
(13) Intesa	(27) Bisel	(41) Fastnet	(GRO) Groupe
(14) BFT	(28) Finaref	(42) BES	

Le graphique présentant le résultat de l'AFC appliquée aux cinquante-cinq parties du corpus C² met en évidence, aux extrémités des axes, des stocks de co-occurents dont la présence a pour effet de minimiser les différences entre les autres parties du corpus. On peut alors se demander, compte tenu des conclusions de la première partie et du chapitre 10, si la cause de ce phénomène n'est pas liée aux propriétés générales du vocabulaire des stocks, selon que leurs occurrences sont concentrées dans l'une des deux sortes de contextes, à dominante syntactique et non syntactique. On a vu dans le précédent chapitre que les principaux représentants de la classe des dénominations propres de personnes morales sont majoritairement présents dans les contextes à dominante syntactique, sans pour autant être absents des contextes non syntactiques. On peut alors contrôler dans le corpus d'origine (RA), si les occurrences des unités correspondant aux étiquettes situées aux différentes extrémités des axes du graphique se conforment à cette tendance générale¹. Pour cela, on projette les unités et leurs variantes sur un graphique de ventilation des occurrences dans lequel le corpus de premier niveau (corpus d'origine des rapports d'activité, ou corpus RA), a été partitionné en fonction des deux sortes de contextes.

¹ Il s'agit des étiquettes 16 (Indosuez W.I. Carr Securities - IWICS), 18 (Carr Futures), situées en haut de l'axe 2, des étiquettes 40 (Banco del Desarrollo) et 47 (Banco Acac) situées en bas de l'axe 2, ainsi que des étiquettes 05 (Caisse régionale), 45 (Unifica), CAL (Caisses locales) et BUR (Bureaux permanents) situées sur la partie droite de l'axe 1.

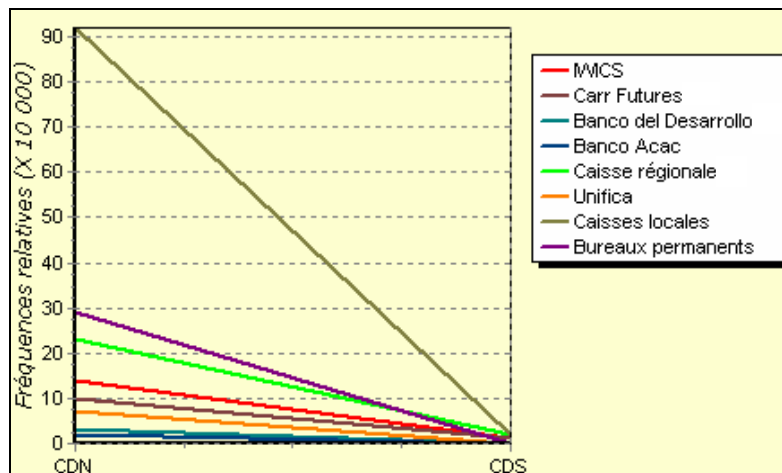


Figure 91.
Répartition des occurrences de huit formes-pilotes et de leurs variantes entre contextes à dominante syntactique et non syntactique dans le corpus RA

Les fréquences relatives des occurrences des huit stocks de formes projetés sur le corpus RA indiquent une ventilation à dominante CDN assez nette, qui s'écarte donc du profil identifié plus haut pour les principaux représentants des dénominations de personnes morales. Compte tenu du caractère atypique de ces huit unités, on peut décider de leur appliquer, de façon exhaustive ou sous la forme de sondages, la collecte réservée aux unités de faible et de moyenne fréquence, présentée dans la précédente section. En effet, on a vu qu'une présence majoritaire dans les CDN a de fortes chances de produire une collecte relativement pauvre, compte tenu des objectifs que l'on a assignés à cette dernière. Par ailleurs, on a intérêt à retirer ces groupes de co-occurents du traitement décrit dans cette section, car leur présence introduit une distinction entre les stocks qui se situe en deçà des aspects thématiques, en opposant d'abord des ensembles de formes caractéristiques des contextes à dominante syntactique et non syntactique. Enfin, la présence de stocks de co-occurents trop hétérogènes risque de fausser le calcul des spécificités pour les mêmes raisons. On donne ci-dessous le résultat de l'AFC lorsque l'on retire du corpus les huit stocks dont il vient d'être question¹.

¹ Caractéristiques textométriques du corpus C² modifié après contrôle de l'homogénéité des stocks de vocabulaire : il se compose de 47 parties représentant 128 527 occurrences pour 4 166 formes, dont 966 hapax. La fréquence maximale est de 10 243 et correspond à la forme « de ».

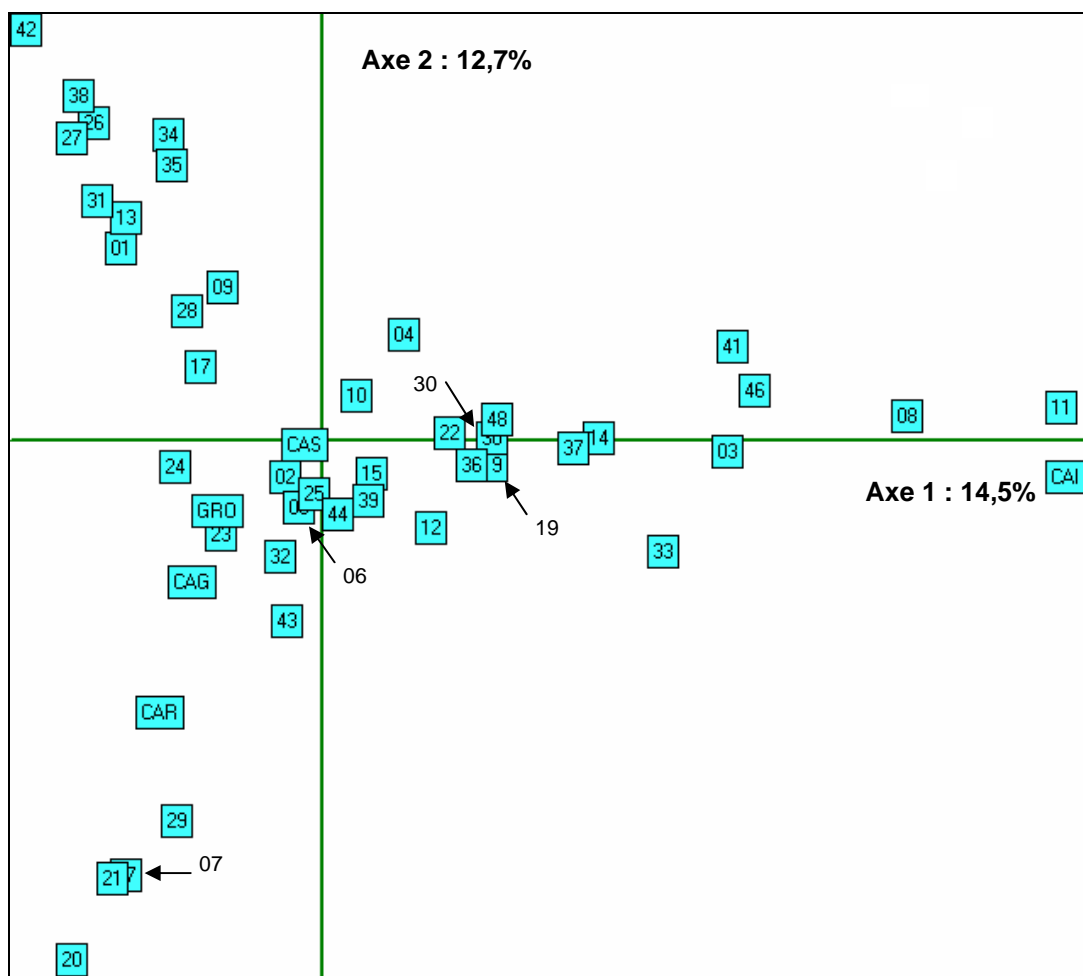


Figure 92.
AFC appliquée au corpus C² composé de 47 stocks de co-occurents

L'aménagement du corpus C² réalisé à la suite du contrôle de l'homogénéité des différents stocks de co-occurents dont il se compose permet d'obtenir une répartition des différentes parties sous la forme d'un *continuum* le long des deux axes dessinés par les principaux facteurs de l'AFC¹. On a ainsi assuré les bases

¹ L'examen de la forme la plus fréquente pour chaque partie permet un contrôle similaire. En effet, on constate que parmi les stocks retirés du corpus C², cinq d'entre eux n'ont pas pour fréquence la plus élevée la forme « de », mais une autre unité possédant une majuscule et correspondant dans tous les cas à une partie de dénomination propre. On dispose là d'un indice du fait que les co-occurents de la forme en question se situent majoritairement dans des contextes à dominante non syntactique. Il reste que si l'on suit cet indice, deux autres stocks auraient également dû être retirés (le 11 et le 41). Par ailleurs, pour trois des huit stocks éliminés du corpus C² (il s'agit des stocks 05, 40 et 47), la forme « de » est néanmoins la plus fréquente. Les deux types de contrôles se recoupent donc partiellement sans être tout à fait équivalents. Pour ce qui est de la collecte, l'incidence est limitée, puisque les

d'une exploration thématique des différentes parties du corpus C². Pour mettre au jour ce qui caractérise plus particulièrement chaque stock de co-occurents, on a recours au calcul des spécificités.

d. Sélection des formes les plus caractéristiques de chaque stock de co-occurents

En appliquant un calcul des spécificités¹ à l'un des stocks du corpus C², on cherche à savoir ce qui le distingue des autres stocks. Ce faisant, on est conduit à rechercher en quelque sorte ce qui est spécifique au sein du spécifique, puisque le stock est lui-même composé d'unités spécifiques appartenant aux contextes dans lesquels la dénomination étudiée apparaît dans le corpus d'origine (RA). Le calcul des spécificités appliqué à l'un des stocks de co-occurents produit une liste de formes simples affectées d'un coefficient de spécificité positif ou négatif. On décide alors de sélectionner parmi ces formes celles qui obéissent à trois critères. Le premier est relatif au coefficient de spécificité, qui doit être positif et posséder une valeur égale ou supérieure à 4. Ce seuil présente un compromis empirique intéressant entre le volume d'unités retenues et la pertinence de la sélection, au vu des résultats qu'elle permet d'obtenir pour la collecte de vocabulaire.

Le deuxième critère porte sur la catégorie grammaticale des formes ayant un coefficient de spécificité égal ou supérieur à 4. Dans la mesure où les unités que l'on recherche peuvent prendre la forme de substantifs ou de syntagmes dans lesquels un adjectif est susceptible d'entrer, seules les formes appartenant à ces deux catégories sont sélectionnées.

Enfin, parmi ces substantifs et adjectifs on ne retient pas les formes graphiques qui entrent dans la formation de la dénomination pilote ou de ses variantes. Dans l'exemple que l'on donne ci-dessous pour « Crédit Agricole Epargne Salariale », les formes graphiques répondant aux deux premiers critères mais correspondant à « Crédit », « Agricole », « Epargne » et « Salariale » sont exclues².

unités-pilotes auxquelles la deuxième méthode n'est pas appliquée peuvent être traitées avec la première méthode.

¹ Pour ce calcul, le seuil de probabilité est toujours fixé à 5% et la fréquence est abaissée à 2.

² Il s'agit d'une simple précaution destinée à éviter des explorations redondantes. En effet, on rappelle que le calcul des spécificités réalisé dans le corpus d'origine pour repérer les co-occurents apparaissant dans les mêmes paragraphes que « Crédit Agricole Epargne Salariale » et ses variantes, a

En revanche, les formes sans majuscule initiale correspondant à ces mots sont retenues. Les formes sélectionnées servent à constituer une nouvelle liste que l'on rassemble dans un groupe de formes créé et enregistré afin de garder une trace de cette sélection. On donne ci-dessous les formes sélectionnées parmi les formes spécifiques du stock de co-occurents « Crédit Agricole Epargne Salariale¹ » :

*Tableau 83.
Sélection de formes spécifiques du stock de co-occurents de CAES²*

épargne	capitaux	mandats	tenue
retraite	ELS	FCPE	appel
salariale	entreprises	dispositions	clientèle
Entreprise	longue	position	inflation
Longue	offres	groupes	cible
Salariés	PEE	marquée	Répartition
salariés	label	Plan	moyennes
mandat	capacité	petites	françaises
intermédiaire	lancement	Solidaire	rendement
segment	complément	garantie	forte

Le groupe de forme ainsi constitué est enregistré. Il pourra ainsi être sollicité autant de fois que nécessaire. Ce point est utile dans l'approche que l'on a retenue, car la même liste va être utilisée simultanément dans deux instances du logiciel Lexico, ouvertes pour les deux corpus que l'on utilise, le corpus de deuxième niveau C², regroupant les stocks de co-occurents des unités-pilotes, et le corpus d'origine RA. Les formes rassemblées dans cette liste constituent autant de suggestions destinées au repérage des unités fonctionnant comme catégorisations ou comme secteurs d'activité de rattachement pour le référent de la dénomination « Crédit Agricole Epargne Salariale ». L'utilité de disposer de deux instances de Lexico ouvertes simultanément pour les deux corpus C² et RA tient à la possibilité qui est ainsi offerte de réaliser une exploration portant sur deux sortes de contextes.

sélectionné, parmi les formes et les segments répétés présents dans ces paragraphes, la dénomination et ses variantes. Ces dernières se retrouvent donc dans le stock de co-occurents de « Crédit Agricole Epargne Salariale » présent dans le corpus C².

¹ Cette dénomination propre fait partie d'un groupe de formes rassemblant trois segments répétés correspondant aux variantes suivantes : « CAES » (14 occ.), « Crédit Agricole Epargne Salariale » (9 occ.) et « C A E S » (2 occ.). Le stock de co-occurents de ce groupe de forme porte le n° 48.

² Les formes sont rangées de haut en bas et de gauche à droite, en fonction de l'ordre dans lequel elles ont été sélectionnées dans la liste des formes spécifiques du groupe de co-occurents de CAES dans le corpus C².

Comme on s'en souvient, le corpus C² rassemble des formes et des segments répétés co-occurents de la dénomination et de ses variantes dans le corpus RA. Ces formes et ces segments, rassemblés dans le stock de co-occurents, sont séparés les uns des autres par un signe conventionnel, le signe dièse en l'occurrence. Pour explorer ce stock, on utilise l'une des formes sélectionnées dans la liste qui vient d'être constituée et on observe, à l'aide d'une concordance affichée dans Lexico, ses cotextes gauches et droits captés par les segments répétés (la forme graphique projetée est également présente dans ces concordances, mais son repérage n'apporte aucune information supplémentaire). Les concordances de Lexico affichent simultanément les segments incluant la forme et le stock de co-occurents auquel appartiennent les segments répétés, ce qui permet d'isoler celui qui correspond à la dénomination propre étudiée.

À ce stade, les segments répétés fournissent fréquemment, non seulement le syntagme probablement lexicalisé susceptible d'être collecté, mais aussi des morceaux de phrases qui permettent de le délimiter avec certitude, comme cela a déjà été constaté à propos de l'identification des dénominations propres. Par ailleurs, la présentation simultanée des autres stocks de co-occurents permet de commencer à documenter la récurrence de l'unité jugée intéressante dans le cadre de la collecte. Ce premier repérage peut suffire dans bon nombre de cas. Cependant, il peut être aussi nécessaire de vérifier l'usage qui est fait de l'unité repérée à l'échelle du corpus entier (il s'agit du corpus d'origine). On utilise alors la deuxième instance de Lexico.

La deuxième instance de Lexico est utilisée pour l'exploration du corpus d'origine RA. Celui-ci est affiché sous la forme d'une carte de sections correspondant aux paragraphes originaux, regroupés en fonction des parties principales du corpus, elles-mêmes découpées en contextes à dominante syntactique et non syntactique. La dénomination propre étudiée, ainsi que ses variantes, sont projetées sur la carte des sections. Il est alors possible de projeter également une ou plusieurs formes appartenant à la même liste que celle qui est utilisée pour l'exploration du corpus C². La carte des sections affiche alors simultanément, à

l'aide de codes couleur différents, les paragraphes contenant une occurrence de la dénomination propre étudiée et ceux où la forme projetée est présente. Lorsqu'un paragraphe contient les deux sortes d'occurrences, il est signalé par une double coloration. On opère de cette manière un retour au contexte d'origine qui a servi pour le calcul des co-occurents. Mais il est aussi possible d'aller plus loin, et de vérifier, par exemple, la fréquence et la récurrence de la forme ou de la polyforme repérée dans la première instance de Lexico sur le corpus C². On verra plus loin que ce point est important pour décider de l'intégration ou non d'une unité candidate.

La procédure appliquée consiste donc à vérifier le fonctionnement sur l'axe syntagmatique des formes de la liste destinée à explorer les stocks de co-occurents. En théorie, toutes les formes de la liste doivent être contrôlées de la même manière. La pratique montre que les informations mises en évidence deviennent rapidement redondantes au fur et à mesure que l'exploration se déroule. Cela tient à deux raisons principales. En effet, les formes sélectionnées étant ordonnées en fonction du coefficient de spécificité, les premières explorations livrent en général les informations les plus pertinentes, c'est-à-dire celles qui révèlent des unités jouant un rôle de catégoriseur ou de secteur d'activité de rattachement. Ces unités correspondant fréquemment à des syntagmes lexicalisés, les formes composant les syntagmes sont réparties dans la liste.

À côté de ce bruit inhérent aux unités collectées, la liste comporte aussi des formes inutiles pour la collecte. En effet, si elles participent bien dans les discours à la restitution de l'activité liée au référent de la dénomination étudiée, elles le font sans être rattachées à cette dernière par l'une des deux relations recherchées. Ces scories sont également inévitables, car elles font partie d'une démarche d'exploration qui, par définition, ne permet pas de connaître à l'avance le détail des résultats de la collecte. On donne ci-dessus un schéma récapitulatif des procédures qui viennent d'être présentées ; il donne, par ailleurs, un exemple de consultation simultanée des contextes d'apparition d'une même forme (« retraite ») dans les deux corpus C² et RA.

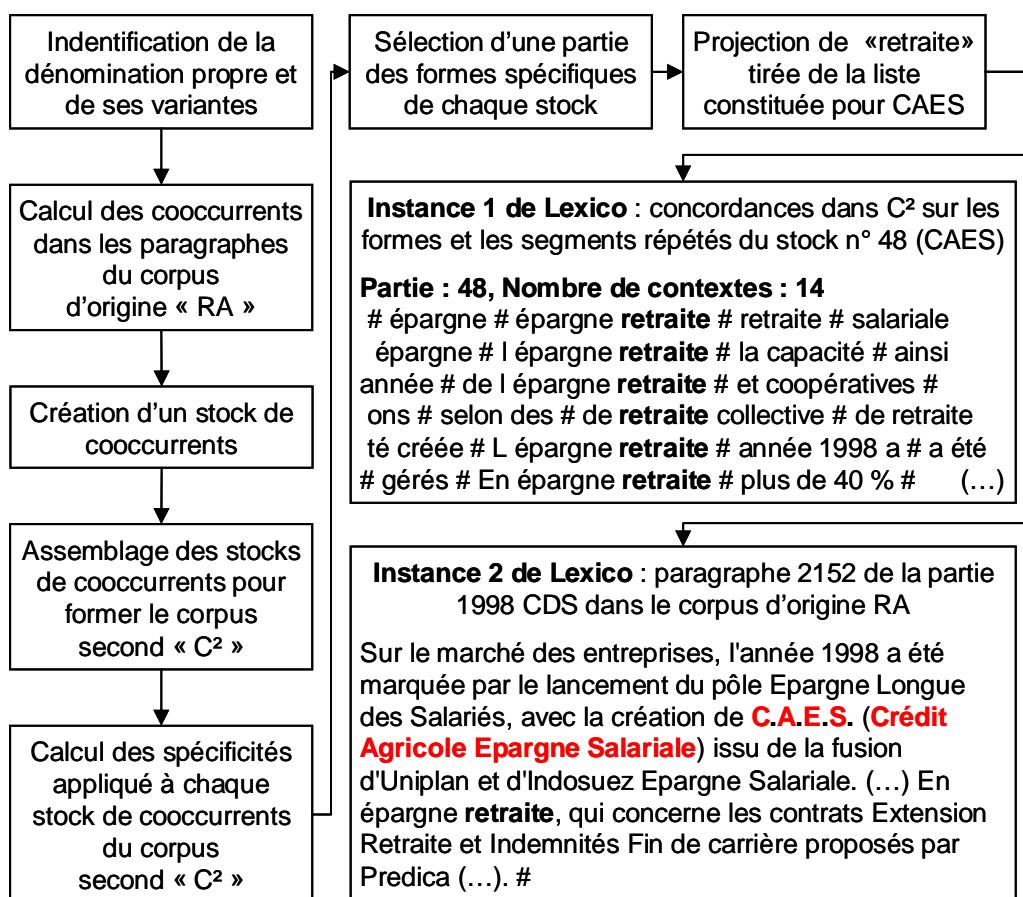


Figure 93.

Processus de repérage des unités à collecter pour une unité-pilote (ici, CAES), et projection d'une forme spécifique sélectionnée (« retraite ») sur les corpus C² et RA ouverts simultanément dans deux instances de Lexico

Dès lors qu'une vision globale des principales formes co-occurentes de la dénomination propre a été fournie par cette procédure, il n'est plus indispensable de procéder à une lecture exhaustive de tous les paragraphes du corpus d'origine où cette unité apparaît. Le fait d'utiliser une série textuelle chronologique permet de travailler sous la forme de sondages en observant plus particulièrement les contextes appartenant aux premières occurrences, aux occurrences médianes puis finales de la dénomination étudiée. Pour les cas, plutôt rares, où un examen plus approfondi doit être réalisée, il est possible de s'appuyer sur la carte des sections, afin de réaliser deux autres sortes de sondages. Le premier permet de contrôler l'exhaustivité d'une collecte qui repose, comme cela a été montré plus haut, sur deux calculs des spécificités appliqués successivement au corpus RA puis à son corpus de deuxième niveau C². Pour cela, on projette sur la carte des sections du corpus RA l'ensemble

des formes sélectionnées dans le groupe de formes co-occurentes obtenues à partir de C², ainsi que le contenu du groupe de formes contenant la dénomination et ses variantes. Les paragraphes signalant une occurrence de ces dernières en même temps que l'absence d'occurrences appartenant au groupe formes co-occurentes peuvent alors être examinés, afin de combler éventuellement les lacunes de la collecte. En fait, sur plusieurs dizaines de dénominations, ce contrôle s'avère négatif, à quelques exceptions près. Celles-ci résident dans des associations rares ou situées dans la dernière partie du corpus. La seconde possibilité de contrôle consiste à projeter l'une des formes co-occurentes sur la carte des sections où la dénomination et ses variantes ont déjà été projetées. Il est alors possible d'examiner les paragraphes où la forme co-occurrence est présente mais la dénomination absente. On dispose alors du moyen de mesurer ce que la dénomination permet de capter de l'usage d'un terme ou d'une expression candidats.

Comme on l'a déjà indiqué précédemment, ce dispositif a été mis en place afin de procéder au repérage et à la collecte d'unités candidates pour les deux relations de catégorisation et de rattachement tissées autour des dénominations propres. Il convient maintenant de présenter les critères de sélection de ces unités qui ont été mis en œuvre.

II. RÉALISATION DE LA COLLECTE

A. Matérialisation linguistique des relations-types retenues pour la collecte

1. Constructions discursives de la relation de catégorisation

a. Prise en charge énonciative et opérations discursives

Lorsque l'on a cherché à caractériser le vocabulaire des rapports d'activité, on a formulé l'hypothèse selon laquelle le locuteur entretiendrait l'illusion d'un vocabulaire dont le sens serait transparent pour ses interlocuteurs. Comme on a pu le montrer, cette conception a pour effet perceptible, entre autres, de bloquer la prise en charge cognitive nécessaire pour rendre plus accessibles certaines unités réputées difficiles d'accès pour un néophyte. Dans le corpus de rapports d'activité, seules les dénominations propres font exception à la règle. Elles seules semblent faire l'objet d'un apport cognitif destiné à permettre l'identification des réalités désignées au fil

du texte. Encore faut-il nuancer ce propos, car on a constaté par la suite que toutes les dénominations ne font pas l'objet des mêmes attentions. En effet, l'établissement des structures-types de contenus a permis de mettre en évidence une double hiérarchie, entre classes de dénominations propres – ou plus précisément, entre classes de référents désignés dans les textes à l'aide d'une dénomination propre – et, à l'intérieur d'une même classe, entre unités plus ou moins valorisées dans le genre de discours des rapports d'activité.

Ces enseignements ont permis d'explicitier les choix ayant guidé la collecte. Cependant, ils se révèlent trop généraux dès lors que l'on aborde la réalisation de la collecte elle-même. En effet, il s'agit maintenant de la sélection d'unités particulières entrant dans le champ textuel des dénominations propres. Afin de justifier le choix de telle ou telle unité, il convient de déterminer des critères opérationnels pour la collecte. La mise au jour de ces derniers suppose que l'on ait précisé au préalable la manière dont les relations sont établies dans le discours entre une dénomination propre et des unités candidates. Cet aspect de la question étant lié à la nature des relations étudiées – on rappelle qu'il s'agit de la catégorisation du référent de la dénomination propre, et de son rattachement à un secteur d'activité particulier ou à une clientèle –, les deux relations types seront examinées séparément.

Comme on l'a rappelé au début de la troisième partie, il est communément admis que l'utilisation d'un nom propre présuppose que les interlocuteurs connaissent, même de façon sommaire, certaines caractéristiques du référent désigné. Ce principe¹, nécessaire au bon déroulement de la communication, est respecté dans les domaines les plus divers. Dans le cas des rapports d'activité, l'enjeu est double : le lecteur qui glisserait trop rapidement sur certains mots inconnus passerait non seulement à côté des réalités désignées, mais surtout, ne percevrait pas ce qui fait la particularité de l'organisation que les rapports d'activité sont censés promouvoir. Comme cela a été montré dans la seconde partie, l'énonciateur semble avoir pris certaines précautions afin de prévenir les lacunes

¹ Celui-ci est à rattacher au plus général principe de coopération (*cooperative principle*) énoncé par H. Paul Grice, et en particulier à la catégorie de règles dite de modalité (« Soyez clair »). H. P. Grice, « Logic and Conversation » (1975) ; on renvoie à la traduction de F. Berthet et M. Bozon (1979).

prévisibles des destinataires de ces publications. En effet, on a constaté que, dans les textes du corpus, les dénominations propres font généralement l'objet d'apports cognitifs utiles pour rendre explicite la nature de leur référent.

Cette forme de prise en charge énonciative a déjà fait l'objet d'investigations, entre autres, dans un genre de discours très différent, puisqu'il s'agit de romans d'auteurs du XIX^e siècle. Il serait facile d'opposer les deux genres ; néanmoins, dans la mesure où cette question est de portée générale, il a paru utile de reprendre certains aspects de la démarche proposée par J.-D Gollut et J. Zufferey¹. En utilisant pour corpus les incipit des romans de la *Comédie humaine* ou bien, dans une étude plus récente, celui de *Salammbô* de Gustave Flaubert, ces auteurs proposent une grille d'analyse pour les opérations référentielles du type de celles que l'on cherche maintenant à préciser, et qui consistent à définir un objet en le rattachant à une catégorie.

Les options théoriques de ces auteurs s'inscrivent dans le cadre plus général de la construction des objets de discours qui a également été retenue dans ce travail : *« Par l'idée d'opération, il s'agit de décrire la référence comme une activité discursive appliquée à l'élaboration d'objets de connaissance en vue de leur communication. Cette acception, fortement inspirée par le concept de schématisation introduit en logique naturelle par J.-B. Grize, demande que soit avant tout pris en compte, au titre de la référence, le travail destiné à aménager l'accessibilité cognitive des objets présentés au destinataire². »*

On s'attachera ici à présenter succinctement les observations que ces auteurs font au sujet des noms propres présents dans les premières phrases des romans balzaciens. Pour ce type d'unités, J.-D. Gollut et J. Zufferey relèvent que l'opération de catégorisation est réalisée selon deux modes principaux. Le premier correspond à une prédication attributive réalisée à l'aide de l'opérateur copule + substantif, comme dans cet exemple *« M. de Manerville le père était un bon gentilhomme*

¹ J.-D Gollut et J. Zufferey, *Construire un monde – Les phrases initiales de la Comédie humaine*, (2000), et « La référence en début de texte : *Salammbô* », in *Textes et discours : catégories pour l'analyse*, (2005).

² J.-D Gollut et J. Zufferey (2005 : 43).

normand bien connu du maréchal de Richelieu [...]. (14. Le Contrat de mariage)¹. »

Le second réalise la catégorisation à l'aide d'une extension attributive prenant la forme d'une apposition, comme dans cet autre incipit : « *Au commencement du mois d'octobre 1829, M. Simon Babylas Latournelle, un notaire, montait du Havre à Ingouville [...]. (55. Modeste Mignon)². »*

Dans un premier temps, on reprendra ces distinctions, afin de présenter les formes de catégorisations des entités nommées relevées dans le corpus de rapports d'activité. On sera alors conduit à compléter la liste des procédés syntaxiques déjà recensés par les auteurs précédemment cités, entre autres, parce que ces derniers se sont volontairement limités au seul cadre de la phrase. Par ailleurs, la collecte étant ici réalisée à l'échelle du corpus entier, il conviendra de présenter les grands types de catégorisations observés et de caractériser les rapports qu'ils peuvent entretenir avec la nature de l'entité catégorisée.

b. Procédés de catégorisation : entre ellipse et périphrase

Dans le corpus de rapports d'activité, la catégorisation des référents désignés à l'aide d'une désignation propre emprunte les deux procédés qui viennent d'être présentés, avec néanmoins une nette préférence pour l'extension attributive. Celle-ci, en offrant la possibilité d'opérer la catégorisation de façon incidente, permet à l'énonciateur de gommer autant que possible les traces de son intervention, ce qui est conforme à la manière qui a déjà été observée dans les deux précédentes parties au sujet des rapports d'activité. La catégorisation peut être alors être perçue comme le résultat d'un arbitrage opéré par l'énonciateur, entre ce qu'il est nécessaire de dire et la nécessité de ne pas trop le faire sentir. La collecte des unités catégorisantes doit donc se laisser guider par les indices syntaxiques qui rattachent ces dernières à la dénomination propre. On en donne ci-dessous un nouvel exemple :

« *En 1999, Prediagri, produit de retraite complémentaire conçu par Predica, a enregistré une croissance soutenue des souscriptions.* » Dans cet extrait, l'extension attributive, placée en apposition entre deux virgules, vient rattacher le référent de la

¹ J.-D Gollut et J. Zufferey (2000 : 62).

² *Ibid.* (2000 : 69).

dénomination propre inconnue à la classe des produits distribués par les banques ou les compagnies d'assurance ; ce dernier point ne peut cependant être éclairci que si l'on sait rattacher le référent de « Predica » à l'une des deux sortes d'entreprises. C'est ce que permet de faire cet extrait, situé une trentaine de pages en amont dans le même rapport d'activité (1999) :

« 1986 : Diversification dans la "bancassurance" avec la création de Predica, compagnie d'assurance-vie, suivie, en 1990, par celle de Pacifica, compagnie d'assurance-dommages. »

L'apposition, cette forme classique de l'extension attributive, n'est pas le seul procédé auquel l'énonciateur des rapports d'activité a recours. En effet, d'autres moyens sont également utilisés, pour lesquels les liens établis entre dénomination propre et unités catégorisantes sont de force variable.

En effet, dans le cadre des extensions attributives, on relève, outre l'apposition, un usage fréquent de la dénomination propre en position d'épithète dans des syntagmes lexicalisés, qui fonctionnent parallèlement à l'utilisation de la seule dénomination propre. On a ainsi « gamme Dynatop », « offre Mozaïc », « carte Open », « site Pleinchamp.com », « prêts Agilor », « contrat d'épargne retraite Confluence » ou encore, « contrat d'assurance-vie multisupport Floriane ». Dans cette série d'exemples, le degré de précision des régissants va croissant, allant du mot passe-partout, comme « offre », au terme désignant un concept précis avec « contrat d'assurance-vie multisupport ». Ce sont donc des unités appartenant aux différentes strates du parler d'entreprise qui sont ainsi déterminées par des dénominations propres. Par ailleurs, l'apposition peut dans certains cas introduire plus qu'une simple catégorisation en proposant une véritable périphrase définitionnelle, comme dans ces deux exemples : « *Crédit permanent adossé à la carte bancaire Eurocard-Mastercard et géré en partenariat avec Sofinco, Open a enregistré, en trois ans, 300 000 ouvertures de comptes [...].* », ou encore, « *Dans le même temps, Agilor, crédit prescripteur développé avec les constructeurs et les distributeurs de matériel agricole, a connu un succès important.* »

On a déjà indiqué que les prédications attributives à fonction catégorisante sont assez peu utilisées dans le corpus de rapports d'activité. Il convient néanmoins

de mentionner deux sortes d'usages caractéristiques. Le premier utilise la prédication attributive afin de marquer le lien dénominatif unissant une entité à son nom. Cela vaut plutôt pour les produits, mais se trouve également pour les sociétés, comme dans les deux exemples donnés ci-dessous : « (...) *le Crédit Agricole a lancé un programme de fidélisation baptisé "TLJ" (Tous Les Jours des avantages).* », et « *Désormais regroupée sous le nom de Crédit Agricole Alternative Investment Products Group (CA-AIPG), la gestion alternative représentait plus de 2 milliards d'euros d'encours gérés fin 2001 (...).* » Ces prédictions s'expliquent par les fonctions du rapport d'activité, qui doit faire état de l'actualité commerciale remarquable, mais aussi des changements affectant l'organisation du groupe bancaire.

Une autre forme de prédication attributive est utilisée, semble-t-il, moins pour catégoriser que pour prendre prétexte de la catégorisation afin de réaliser une qualification de l'entité nommée. C'est bien sûr une autre fonction importante assumée par les rapports d'activité, que de valoriser les entités responsables des résultats économiques dont ils font état. On trouve ainsi : « *En 1999, le Crédit agricole est la première banque française par le montant de ses capitaux propres.* », « *CAI est le 1er intervenant sur le marché primaire de la dette souveraine à court terme en euros.* ». Ou bien encore : « *Sofinco est devenue en 2001 l'organisme de crédit qui bénéficie de la plus grande notoriété en France, d'après le baromètre IPSOS réalisé en janvier 2002.* »

Il faut par ailleurs mentionner des expressions de la catégorisation qui sortent du cadre proposé par J.-D. Gollut et J. Zufferey. La première d'entre elles concerne les reprises anaphoriques nominales internes à la phrase ou transphrastiques. On a pu voir dans la deuxième partie que ce type de reprise n'est pas très fréquent, ce qui le rend assez difficile à repérer. Néanmoins, aux fins d'illustration, il est possible de limiter les investigations à une configuration type facilement identifiable. On a retenu pour cela la présence d'un adjectif démonstratif – dont la majuscule initiale indique qu'il est placé en tête de phrase –, suivi par un substantif. Cette configuration, repérée à l'aide d'une recherche portant sur les segments répétés du corpus commençant par « Ce », donne une trentaine de réponses que l'on présente

ci-dessous, nettoyées des informations redondantes (segments répétés de trois formes ou plus) ou des scories (par exemple, noms de société commençant par Ce, comme « Cedicam »).

Forme	Fréquence
Cette acquisition	5
Cette action	3
Cette activité	4
Cette banque	4
Cette équipe	2
Cette filiale	4
Cette hausse	2
Cette initiative	4
Cette offre	9
Cette opération	8
Cette société	3
Cette structure	3
Ce dispositif	3
Ce document	4
Ce pôle	8
Ce produit	2
Ce projet	3
Ce service	2
Ce succès	2
Ces activités	4
Ces développements	2
Ces évolutions	4
Ces fonds	3
Ces initiatives	2
Ces métiers	3
Ces résultats	3
Ces tendances	3

Figure 94.

Recherche sur les segments répétés du corpus commençant par « Ce » à partir de l'interface de constitution de groupes de formes de Lexico

Les fréquences relativement peu élevées qui caractérisent cette trentaine de segments répétés confirment la précédente remarque au sujet du caractère relativement peu étendu du phénomène. On note cependant que celui-ci est plus présent dans le cadre phrastique. On donne ci-dessous un exemple illustrant la reprise transphrastique à l'aide de « Cette filiale » : « *Les activités de grandes clientèles, de marché et de l'international du Crédit agricole sont principalement regroupées au sein de Crédit Agricole Indosuez. Cette filiale opère à travers un réseau de succursales et filiales implantées dans plus de cinquante pays.* » Un autre procédé, également utilisé dans les rapports d'activité, réside dans l'emploi d'une

unité dont la fonction catégorisante est appliquée de manière distributive aux éléments composant une même série¹ : « *Au total, avec une part de marché de 13,1 %, le Crédit agricole (par ses filiales de gestion Segespar, BFT, BGP et Indosuez) est le premier réseau distributeur de Sicav.* » Dans ce type de configurations, le lien établi entre la dénomination propre et l'unité catégorisante tend à se distendre, et les exemples suivants semblent bien constituer des cas limites : « *CA-AM a développé plusieurs nouveaux produits destinés à la clientèle des particuliers : Step PEA, Preditop 5, Energetic PEA et Nutrio, Regener, CA-AM Actions Cliquet, CA-AM Stratégie alternative, fonds "opportunité novembre 2003" (...).* », ou bien encore : « *Calyon a été retenu comme conseil de l'Etat ou de la société pour la quasi-totalité des opérations de privatisation de taille importante prévues pour 2004 ou 2005 : EDF, GDF, ADP, Snecma, Thalès, cession des 51 % de la SNET encore détenus par Charbonnages de France.* » On remarque également que c'est la seule forme de catégorisation que l'on trouve dans les contextes à dominante non syntactique. Dans ces derniers, une signalétique (couleur particulière des caractères, pictogrammes), vient parfois préciser si la société dont le nom est cité correspond à une filiale, à une participation ou à une succursale. On rencontre également, par exemple pour les caisses régionales de Crédit agricole, un intertitre mentionnant qu'il s'agit de ce type d'entités, suivi d'une liste de plusieurs dizaines de caisses régionales dont le nom est donné en version courte (« Centre-Est » pour « Caisse régionale de Crédit agricole Centre-Est »).

Enfin, ces derniers exemples permettent d'aborder le cas des dénominations propres pour lesquelles aucune catégorisation n'est proposée de manière explicite. On en trouve quelques-unes parmi les unités de basse ou de moyenne fréquence, comme dans ces exemples : « *Un mouvement de même ampleur a été en enregistré en Allemagne où il a été facilité par l'assouplissement de la politique monétaire de la Bundesbank.* », ou encore « *Le Groupe a également été conseil d'Etam dans l'offre publique sur Etam Grand -Bretagne, du Groupe Pinault dans son OPA sur Worms, et de Rallye dans l'OPA défensive sur Casino.* » Cependant, un sondage

¹ Ces séries peuvent être partiellement repérées à l'aide d'une expression régulière du type : \<[A-Za-zéèê\-\-]+, \<[A-Za-zéèê\-\-]+, \<[A-Za-zéèê\-\-]+ qui permet de rechercher une série de trois mots séparés par une virgule.

portant sur un ensemble de plus d'un millier de dénominations montre que les cas de non catégorisation sont très minoritaires dans le corpus pour les deux classes de dénominations étudiées – moins de cinq pour cent –, y compris parmi les dénominations de fréquence 1 ou 2. Les « carences » portent de manière assez systématique sur des dénominations propres de sociétés ou d'organisations possédant un minimum de notoriété économique : banque centrale d'un état européen, grands groupes financiers, industriels ou bancaires, comme dans les précédents exemples. Par ailleurs, on constate que le critère de notoriété recoupe un autre clivage entre unités catégorisées et non catégorisées dans les textes. En effet, on remarque également qu'à notoriété équivalente, la catégorisation est toujours utilisée pour les entités nommées entretenant un rapport ontologique étroit avec le groupe bancaire, alors qu'elle est moins systématique pour les autres. C'est ce que met en évidence cet extrait, dans lequel l'un des membres du conseil d'administration de Crédit Agricole S.A. est présenté : « *Gérard Mestrallet. Nommé le 29 novembre 2001. Président Directeur Général de Suez. Président de la Société Générale de Belgique et de Tractebel. Vice-Président de Aguas de Barcelona et de Hisusa. Administrateur de Saint-Gobain et de Pargesa Holding SA. Membre du conseil de surveillance de Axa, Casino, M6, Sagem SA, Société du Louvre. Représentant permanent de Fided au Conseil de Le Monde Entreprises, et de Sperans au Conseil de Fimalac.* »

La lecture de ce paragraphe peut donner l'impression d'avoir des informations sur ce membre du conseil d'administration de Crédit Agricole s.a., mais il n'en va pas de même pour les organisations, dont seul le nom est cité. Il faut aussi préciser que l'absence de catégorisation est plus fréquente en contextes non syntactiques. De ce point de vue, la catégorisation implicite qui prévaut dans ces contextes semble prendre appui sur les informations injectées dans les contextes à dominante syntactique. Les dénominations propres montrent ainsi leur capacité à fréquenter des contextes complémentaires, du moins lorsque leur place dans le référentiel véhiculé par les rapports d'activité le permet.

La diversité des manières de catégoriser confirme la nécessité de recourir à une approche robuste pour l'exploitation des données textuelles. La formule canonique « X est un Y », dans laquelle X représente une entité nommée tombant

sous le concept Y, est la moins utilisée dans le corpus étudié. Les procédés recensés dans cette section ont tous en commun de laisser relativement peu de traces explicites de catégorisation. Sont en revanche privilégiés des procédés discrets qui contribuent à entretenir le confort cognitif du lecteur sans qu'il paraisse. De ce point de vue, le comble est atteint par l'absence de catégorisation, lorsque l'énonciateur semble vouloir compter sur la seule culture économique de son interlocuteur¹. Ce clivage entre entités catégorisées et non catégorisées ne doit cependant pas gommer des distinctions plus fines liées à la nature de l'entité.

c. Catégorisations et classes d'entités nommées : entre exubérance et laconisme

Toutes les entités catégorisées ne le sont pas de la même manière. Cela va de soi au niveau le plus fin, car il est très probable que chaque entité fasse l'objet d'une catégorisation adéquate qui tienne compte de ses particularités. On peut néanmoins se demander si certaines régularités se dégagent d'une observation menée à l'échelle des classes de dénominations propres qui fournissent les unités-pilotes utilisées pour la collecte.

En effet, les dénominations propres de personnes morales sont généralement associées à deux sortes de catégorisations correspondant à deux finalités différentes. La première catégorisation est utilisée pour expliciter la nature du rapport capitaliste ou intuitionnel qui lie l'entité au groupe bancaire. On observe alors une gradation dans l'utilisation des procédés de catégorisation en fonction de la nature du lien. D'une manière générale, plus le lien est étroit avec l'organe central, plus le procédé rencontré est explicite. Le *continuum* s'étend ainsi des constructions du type « *La Fédération nationale du Crédit agricole est l'instance de concertation du Crédit agricole.* », qui s'opposent aux exemples, donnés précédemment, d'absence de catégorisation d'entités extérieures au groupe bancaire, en passant par l'usage d'une catégorie appliquée de façon distributive aux éléments composant une série, comme cela se rencontre pour les filiales, les participations bancaires ou les caisses régionales.

¹ Cette interprétation est confirmée par le traitement des toponymes qui a été présentée plus haut. En effet, on a pu constater que l'appartenance de ces unités à un référentiel partagé par les interlocuteurs dispensait l'énonciateur de recourir à la catégorisation.

La deuxième sorte de catégorisation est destinée à préciser la nature de l'entité. Elle peut prendre des formes diverses se coulant dans les modes opératoires qui ont été rappelés précédemment. On illustre ce point à l'aide d'un extrait des catégorisations associées à la dénomination propre « Sofinco » :

« La Banque Sofinco se situe dans les tout premiers rangs des établissements de crédit à la consommation en France. »

« 1999 Acquisition de Sofinco, spécialiste du crédit à la consommation. »

« Sofinco, organisme de crédit le plus connu des Français. » Lorsque la distance ontologique au centre du groupe bancaire est plus importante, on vérifie que le mode opératoire se fait plus lâche, comme dans cet exemple :

« Des succès ont été remportés sur tous les segments de clientèle :

- grands groupes entreprises : LVMH , Saint - Gobain , PPR (...). »

Pour les dénominations propres d'objets, et plus particulièrement de produits et de services, puisque ces deux sortes d'entités sont majoritaires dans cette classe, on constate une plus grande variété dans les usages de la catégorisation. Tout d'abord, celle-ci prend en charge différentes facettes de l'entité qu'elle catégorise. Ainsi, la marque « Mozaïc » est-elle présentée également comme une offre, car elle regroupe un ensemble de produits et de services bancaires destinés aux jeunes clients ; cela explique qu'elle soit aussi présentée comme une gamme. On retrouve cette alternance des points de vue dans la catégorisation pour un certain nombre de produits, dont « Agilor », présenté successivement comme offre, prêt et crédit :

« En 1997, l'offre Agilor, qui assure la moitié des financements de matériel agricole, (...). », et encore « Grâce à la simplicité des prêts Agilor, crédits prescripteurs développés avec les constructeurs et les distributeurs de matériel agricole (...). »

Dans l'exemple donné plus haut avec la dénomination « Sofinco », les syntagmes « établissement de crédit à la consommation », « spécialiste du crédit à la consommation » et « organisme de crédit » constituent autant de variations autour de la même idée. Pour les produits, la présentation de facettes différentes réalisée par la catégorisation, comme pour Mozaïc et Agilor, peut aller plus loin qu'un

déplacement de point de vue , en permettant la saisie d'une même entité au sein de domaines différents¹. Le produit Prediagri déjà cité est ainsi présenté comme placement, produit de retraite complémentaire et contrat. Cela permet à l'énonciateur d'aborder le même produit sous l'angle de l'épargne, sous celui des régimes spéciaux de retraite, et enfin, sous celui des contrats d'assurance.

Ces usages diversifiés de la catégorisation semblent être limités par trois causes. La première tient à la fréquence de la dénomination propre observée dans le corpus. À cette limitation « mécanique » due à la manière générale dont les dénominations propres d'objet sont traitées dans les rapports d'activité, s'ajoute une seconde limitation, qui tient à la construction de la dénomination. On observe ainsi que lorsque la dénomination propre est complexe et qu'elle intègre un terme, par exemple « Livret jeunes Mozaïc » ou « Pep's² », la diversité de la catégorisation tend à décroître. Tout se passe alors comme si la dénomination propre se suffisait à elle-même. Dans ce cas, le terme joue le rôle de l'unité catégorisante. Enfin, une troisième limitation a été recensée, plutôt exceptionnelle, qui apparaît lorsque l'unité catégorisante adéquate fait défaut ou qu'elle n'est pas encore utilisable dans les rapports d'activité, compte tenu de son caractère trop technique. Plutôt que de recourir à un néologisme, l'énonciateur semble alors préférer une périphrase définitionnelle s'appuyant sur des définisseurs passe-partout, comme « produit », « offre », ou, comme dans cet exemple, « outil » : « *Le Cedicam (...) a développé CA Certificat, un outil de signature électronique qui garantit l'authentification, l'intégrité et la confidentialité des transactions électroniques, notamment pour la déclaration et le paiement de la TVA.* »

En résumé, il apparaît que les réalisations de la catégorisation non seulement sont susceptibles de revêtir des formes diverses, mais aussi remplissent des fonctions

¹ La classe des objets n'est pas la seule à posséder cette propriété qui se retrouve également, mais dans une moindre proportion, pour les dénominations propres de personnes morales. On remarque que pour les unités collectées, ce phénomène de catégorisation à facettes n'est associé qu'aux dénominations propres les plus fréquentes, et correspondant à des entités particulièrement importantes dans le référentiel du Crédit agricole.

² Le « Livret Jeunes Mozaïc » correspond à un produit d'épargne réglementé dénommé « livret jeune » et commercialisé par le Crédit agricole sous la marque Mozaïc ; quant à « Pep's », il s'agit d'une famille de produits du même nom et correspondant à des déclinaisons propres au Crédit agricole d'un autre produit d'épargne réglementé, le plan d'épargne populaire, abrégé en PEP.

différentes qui peuvent varier en fonction de la nature de l'entité nommée. Or, alors que la première caractéristique complique parfois le repérage de la catégorisation, la seconde pose une difficulté d'un autre ordre, celle du choix des unités devant être retenues pour intégration dans un référentiel terminologique. Avant d'aborder ce point, il convient de présenter les caractéristiques du deuxième type de relation que l'on souhaite prendre en compte.

2. Constructions discursives du rattachement à un nom de secteur d'activité ou de clientèle

a. Domaines, marchés, secteurs : saisir le découpage de l'activité

Il convient de préciser tout d'abord que le rattachement des dénominations propres d'objets à des noms de clientèle ne sera traité ici que de façon incidente. En effet, ce rattachement présentant un intérêt moindre que celui qui fait appel à des noms de secteurs d'activité, compte tenu des finalités de la collecte, il ne sera pas traité de manière détaillée. Par ailleurs, la méthode permettant l'identification et la description des opérations discursives de rattachement étant la même dans les deux cas, une présentation complète aurait entraîné des redites. Les noms de clientèle ont cependant été collectés, et les questions soulevées par leur rattachement à un référentiel terminologique seront abordées plus loin.

L'étude du rattachement des unités-pilotes d'objet ou de personne morale à une dénomination de secteur d'activité revient à rechercher les cadres utilisés pour la restitution de l'activité bancaire qu'il est possible de capter à partir de certaines dénominations propres. C'est pourquoi, il semble que cette question ne se pose pas dans les mêmes termes que pour la catégorisation. En effet, la collecte des unités catégorisantes peut s'appuyer sur le constat global selon lequel celles-ci se trouvent dans le contexte de la dénomination propre dont elles catégorisent le référent. La portée de l'empaquetage d'informations réalisé par la dénomination propre a été évaluée a priori comme devant tenir dans le cadre du paragraphe. Les observations empiriques réalisées au cours de la collecte ont confirmé globalement ce fait, la catégorisation étant généralement réalisée dans le cadre de la phrase. En revanche, dans la relation établie entre une dénomination propre et un nom de secteur d'activité, il n'y a pas a priori de nécessité syntaxique obligeant les deux sortes

d'unités à se trouver à proximité l'une de l'autre. On peut même concevoir que la mention d'un secteur d'activité fournisse dans les textes un cadre thématique général, utilisé pour rassembler toutes sortes d'ingrédients, comprenant ou non des dénominations d'objets et de personnes morales, au gré des aléas de la restitution de l'activité bancaire. De ce point de vue, le cadre de la phrase, adopté dans les travaux précédemment cités de J.-D. Gollut et J. Zufferey, peut s'avérer trop étroit. Il reste que le travail de ces auteurs permet de rappeler que le paramétrage de l'univers romanesque à l'aide de circonstants de temps et de lieu peut parfaitement être réalisé dans l'espace exigü d'un incipit. La notion de secteur d'activité est-elle assimilable à de tels paramètres, au point d'être traduite syntaxiquement par les mêmes moyens ? C'est l'hypothèse que l'on a cherché à vérifier en repoussant toute définition a priori d'un tel cadre, et en faisant reposer l'identification de ces dénominations sur l'étude de leur construction dans les discours. Les dénominations propres-pilotes sont alors utilisées comme point d'observation de la manière dont la mention des secteurs d'activité est exprimée dans les textes du corpus et peut conférer à ces unités le rôle de délimiteurs d'univers de discours.

L'option consistant à ne pas chercher à déterminer a priori ce qui peut constituer une bonne unité candidate pour un secteur d'activité provient d'un double constat : d'une part, l'hétérogénéité et le caractère toujours orienté des définitions proposées et, d'autre part, la nécessité de promouvoir l'usage attesté comme alternative aux débats théoriques sur le sujet. Certes, un certain nombre de domaines « académiques¹ » peuvent prétendre à une réelle consistance, mais celle-ci repose pour la plupart d'entre eux sur un processus d'institutionnalisation achevé et réalisé à l'échelle de la société tout entière. Il est également possible de fournir une définition générique relativement satisfaisante des domaines d'activité : « *Un domaine d'activité permet d'identifier un champ d'action, un ensemble d'actes coordonnés, une activité réglée, une pratique. Il correspond à une activité humaine, sociale, économique, industrielle. Il est constitué d'un ensemble de procédés bien*

¹ Il s'agit, par exemple, des sortes de domaines auxquels la norme ISO 704 : 2000 fait allusion : « (...) la présente Norme internationale s'applique aux travaux terminologiques effectués dans des domaines scientifiques, technologiques, industriels, administratifs, ainsi que dans les autres domaines de la connaissance. » Norme ISO 704 :2000(F), p.1.

*définis destinés à produire certains résultats*¹. » Il reste qu'une telle définition laisse encore une grande marge de liberté à qui souhaite délimiter un tel domaine à partir de critères extralinguistiques dans une perspective terminographique : choix du point de vue, du niveau de maille, de l'étendue et, en définitive, de la dénomination. Ce que l'on cherche à capter ici est localisé dans les discours : ce sont les unités particulières qui sont utilisées dans le corpus afin de construire un découpage de l'activité bancaire. Par conséquent, on parlera de « secteur d'activité » de façon générique, sans préjuger à ce stade de la nature du découpage opéré dans les pratiques présentées dans les rapports d'activité. Les unités candidates seront rassemblées en fonction de similarités distributionnelles, pouvant être ramenées à des schémas fréquemment attestés, supports d'une proximité sémantique. Celle-ci sera validée à son tour par la mise au jour d'un trait sémantique commun. Enfin, on admettra que le principe de coopération rend peu probable la proposition par l'énonciateur d'un découpage de l'activité bancaire qui soit complètement atypique par rapport à ce que l'on peut s'attendre à rencontrer dans des rapports d'activité. Dans les limites qui viennent d'être énoncées, toutes les suggestions présentes dans les discours doivent être étudiées. En revanche, il sera nécessaire, dans un deuxième temps, de définir des critères permettant de retenir telle ou telle unité dans le cadre de la collecte.

Dès lors que l'on cherche à recenser les noms de secteurs d'activité à partir des dénominations propres-pilotes², il est nécessaire de s'interroger sur les formes que peuvent prendre les relations établies dans les discours entre ces deux sortes d'unités. On dispose pour cela de plusieurs cadres textuels : la phrase, le paragraphe, mais aussi la rubrique telle qu'elle a été définie dans la Première partie (un espace textuel ou non, délimité par deux intertitres).

La phrase fournit un premier cadre lorsque le secteur d'activité est introduit sous la forme d'un circonstant. Dans cet espace, les deux dénominations d'entité et de secteur d'activité peuvent se rencontrer selon le schéma type suivant : « l'entité X

¹ B. de Bessé (2000 : 184).

² Cette section sera illustrée principalement à l'aide de dénominations propres de personnes morales (entreprises), et il sera fait incidemment allusion au comportement observé pour les dénominations propres d'objets (produits).

exerce son activité dans le secteur Y ». Le secteur d'activité correspond dans ce schéma à une forme de circonstant indiquant le cadre plus général dans lequel s'exerce l'activité de l'entité. Le patron type dans lequel sont associés ces deux ingrédients comporte une préposition, un nom plus ou moins général dont le sens comporte au moins un trait relatif à l'idée d'ensemble ou de portion d'un ensemble, et le nom d'un secteur d'activité. Un exemple prototypique de cette construction serait « dans le domaine du crédit à la consommation », où « dans » fournit la préposition, « domaine » le nom général dont la signification comporte le trait « ensemble », et « crédit à la consommation » correspond au nom d'un secteur d'activité. Le mot « domaine¹ » peut être remplacé dans ce patron par « secteur », « métiers » ou « équipes », « marché », ou encore « activité ». Tous ces noms peuvent entrer dans le patron au singulier ou au pluriel. L'article défini placé devant le mot correspondant à une notion générale peut être remplacé par d'autres déterminants (indéfini, démonstratifs, possessifs). Lorsque « marché » est utilisé, la préposition « sur » remplace « dans ».

Cette construction possède une forme elliptique dans laquelle le mot général peut être omis. L'usage de la préposition « dans » est alors complété par celui de « en ». On rencontre ainsi avec la préposition « dans » : « dans l'assurance dommages » (3 occ.), « dans l'assurance-vie » (7 occ.), « dans l'épargne bancaire » (6 occ.), « dans l'épargne retraite » (1 occ.), « dans l'intermédiation boursière » (1 occ.), « dans la banque au quotidien » (1 occ.), « dans la banque de détail » (5 occ.), « dans la banque privée » (10 occ.), « dans la clientèle des collectivités locales » (2 occ.), « dans la gestion d'actifs » (12 occ.), « dans les financements structurés » (9 occ.), etc. Avec la préposition « en » : « en assurance-vie » (13 occ.), « en crédit à la

¹ Cet usage s'étend à des expressions voisines comme « concernant », « en matière de » qui se trouvent dans une position intermédiaire entre les introducteurs d'univers de discours et les organisateurs discursifs. De ce fait, leur utilisation comme attestation dans le cadre de la collecte est moins évidente que celle des précédentes expressions. Elles peuvent donc jouer le rôle d'indices. On note en effet un recoupement réel avec les autres expressions :

« en matière d'assurance de personnes » (1 occ.), « en matière d'épargne » (2 occ.), « en matière de banque en ligne » (1 occ.), « en matière de crédit » (5 occ.), « en matière de gestion d'actifs » (2 occ.), « en matière de prêts » (3 occ.), mais on observe également « en matière de gouvernance » (2 occ.), ou également « en matière de protection de l'environnement » (2 occ.). On remarque d'ailleurs que sont privilégiées les macro-catégories de secteurs et pas nécessairement des secteurs repris dans d'autres contextes : on est plutôt dans l'ordre des expressions non figées.

consommation » (3 occ.), « en crédit-bail mobilier » (3 occ.), « en épargne retraite » (3 occ.), « en épargne salariale » (6 occ.), « en gestion d'actifs » (6 occ.), etc.

Ces constructions fournissent les cadres à l'aide desquels l'activité de l'entité nommée est restituée. On vérifie que ceux-ci jouent un rôle communicationnel de circonstants à fonction scénique¹, tout particulièrement lorsque les prépositions sont situées en début de phrase ou de paragraphe, comme dans cet exemple : « *Dans l'affacturage, Transfact a conclu un accord de sous-traitance (...).* » De ce point de vue, les noms de secteurs d'activité sont introduits de la même façon que les repères temporels et spatiaux qui occupent fréquemment ce type de position. On rencontre, à côté des organisateurs² « En 1995 » (9 occ.) ou « En 2003 » (27 occ.), « En France » (24 occ.), « En Asie » (12 occ.), mais aussi « En assurance » (8 occ.) ou encore « En épargne retraite » (2 occ.). La préposition « Dans » est également utilisée pour introduire des secteurs d'activité, comme « Dans le domaine de l'épargne » (4 occ.), « Dans les financements aéronautiques » (2 occ.), etc.

On note enfin que ce schéma prend parfois une forme très elliptique dans laquelle les deux dénominations sont placées côte à côte, par une simple juxtaposition ou par la mise entre parenthèses de l'une des deux dénominations. C'est alors le contexte plus général dans lequel cet assemblage se rencontre qui lui donne son sens, comme dans cet exemple : « *Crédit à la consommation (Sofinco) N°2 du crédit à la consommation en France [source:ASF].* » Les contextes à dominante non syntactique (liste de filiales, organigramme financier, etc.) sont particulièrement propices aux formulations de ce type. Cependant, l'interprétation des juxtapositions est la même dans toutes les sortes de contextes ; les constructions du type « Predica (Assurance-vie) » ou « Crédit à la consommation – Sofinco »

¹ On reprend ici la terminologie de J.-M. Adam (1999 : 59), pour qui les organisateurs « *délimitent des unités en ouvrant ou en fermant des portions de textes depuis le niveau intrapropositionnel (dans le groupe nominal) jusqu'aux niveaux interpropositionnels (segmenter et lier des propositions) et textuel (segmenter et relier des pans de texte).* »

² « *On distinguera donc, selon leurs positions et le rôle communicatif qui leur est associé :*
- les circonstants à fonction scénique qui participent à la mise en place préalable du cadre de circonstances ou se situe le reste de la phrase. Ce cadre peut être conçu comme un ensemble de dimensions ouvertes qui, selon les besoins de la communication, peuvent demeurer implicites ou être saturées par des compléments circonstanciels ;
- les circonstants à fonction rhématique qui jouent le rôle de propos par rapport au thème que constitue le reste de la phrase. » M. Riegel, J.-C. Pellat, R. Rioul (2004 : 144).

doivent être interprétées comme une précision du secteur d'activité auquel se rattache la société mentionnée. On peut ajouter que ces juxtapositions doivent paraître plus évidentes au lecteur lorsque celui-ci a pris au préalable le soin de parcourir les contextes syntactiques où cette relation est explicitée. D'une manière générale, les relations établies entre noms de secteurs d'activité et noms d'entreprise sont plus resserrées que celles que l'on observe entre noms de secteurs et noms de produits¹. Cependant, même en ce qui concerne les dénominations propres d'entreprises, cette relation peut se déployer dans des espaces plus étendus que la phrase, comme le paragraphe.

Le paragraphe constitue le deuxième cadre dans lequel une relation est établie dans les discours entre une dénomination d'entité et un nom de secteur d'activité. Lorsque c'est l'activité du secteur qui constitue l'objet principal restitué dans les textes du corpus, la mention de l'entité peut apparaître dans le même paragraphe, à une ou deux phrases de distance :

« L'assurance-vie a continué de se développer à un rythme rapide. Le chiffre d'affaires de Predica, qui s'est établi à 51,3 milliards de francs, a enregistré une hausse de 11 %, supérieure à celle du marché. L'encours a augmenté, quant à lui, de 24% pour atteindre 211,2 milliards de francs. La part de marché de Predica a, ainsi, progressé de 0,7% pour atteindre 9,7% des encours. »

Dans ce paragraphe, extrait du rapport d'activité de 1995, il faut remonter de la dernière phrase à la seconde pour raccrocher par déduction l'entreprise « Predica » (car elle possède un chiffre d'affaires), au secteur d'activité de « l'assurance-vie » (car celui-ci peut être découpé en parts de marché). Dans la suite d'énoncés, cette déduction est guidée par l'isotopie du trait « accroissement » qui court sur les prédictions opérées dans les quatre phrases que compte le paragraphe : « a continué de se développer », « a enregistré une hausse », « l'encours a augmenté » et enfin, « a progressé de 0,7% pour atteindre ». Cette position ouvrante en tête de paragraphe

¹ La construction récurrente qui permet de relier une dénomination propre d'objet, correspondant au cas majoritaire du produit, est du type : Une entité X a conçu le produit Y pour la clientèle Z, le nom de clientèle étant très fréquemment en position de complément d'attribution (on renvoie à la présentation du « datif étendu » in M. Riegel *et alii* (2004 :226). La phrase constitue l'espace typique dans lequel cette relation se déploie.

est également occupée, par exemple, par les syntagmes : « Le crédit-bail » (4 occ.), « L'agriculture » (6 occ.), « L'assurance-vie » (4 occ.), « Les collectivités locales » (9 occ.), « Les entreprises » (10 occ.), « Les métiers de gestion d'actifs » (1 occ.), « Crédit-bail : » (2 occ ;), « Assurance-vie : » (3 occ.), « Collectivités locales : » (1 occ.).

Par ailleurs, on vérifie la récurrence du procédé qui apparaît quatre années plus tard, dans cet extrait du rapport d'activité de 1999 :

« Assurance : un développement soutenu. Le marché de l'assurance-vie a retrouvé, en 1999, un rythme de croissance élevé, favorisé par une stabilité de la fiscalité et une progression sensible des contrats en unités de compte. Dans cet environnement favorable, l'activité de Predica a été soutenue, avec une augmentation du chiffre d'affaires de 14% et une progression des encours gérés de 12%. Ces performances ont permis à Predica d'enregistrer une part de marché en termes d'encours de 10,8%. »

Dans cet exemple, tiré du rapport d'activité de 2003, le rattachement se fait à travers l'appartenance au groupe Crédit Agricole s.a., société mère : *« Assurances de personnes : n°2 sur le marché français. Le rapprochement juridique de Predica et de l'UAF, programmé dans le courant de l'année 2004, renforce les positions de Crédit Agricole S.A. sur le marché français : premier banque-assureur, le pôle assurances de personnes occupe la 2e place du marché de l'assurance-vie, avec une part de 15% des encours gérés. »*

Dans les trois exemples cités, l'empaquetage thématique associé à la dénomination du secteur d'activité rencontre le découpage du texte en paragraphes. La place occupée par la dénomination du secteur d'activité désigne celui-ci comme étant le sujet ou le thème d'un paragraphe ou d'une rubrique, lorsqu'il figure dans l'intertitre et qu'il fait ensuite l'objet d'une reprise à l'aide d'un catégoriseur¹. À l'intérieur

¹ Dans le chapitre 10 on a vu que, selon la définition proposée par J.-M. Marandin, le thème configuré a pour nom « le syntagme nominal en position de fermoir d'une chaîne objet ». Cette définition vaut, semble-t-il, pour tout type de texte. En ce qui concerne les textes du corpus, il semble que le découpage très fin assuré par les quelque huit cents intertitres des rubriques imposent une contrainte textuelle explicite sur le choix du nom du thème, du moins dans la majorité des cas. Le thème, annoncé dans l'intertitre, s'achève là où commence le thème proposé par l'intertitre suivant. Il y aurait donc également des thèmes configurés dont le nom est fourni par un syntagme nominal en

d'un paragraphe, la relation entre les deux dénominations est établie de manière indirecte, grâce aux déductions rendues possibles par la succession des informations apportées par les phrases.

Enfin, il s'avère parfois utile d'observer les associations à l'échelle d'une rubrique, afin de vérifier que la collecte réalisée dans les deux premiers cadres couvre bien toutes les associations entre une dénomination pilote et les différents secteurs d'activité auxquels elle est rattachée dans le corpus. Pour cela, on a recours à une partition du corpus qui reprend le découpage original des textes du corpus en rubriques puis en paragraphes. On projette ensuite les variantes de l'unité pilote sur la carte des sections de Lexico, comme dans l'exemple ci-dessous.

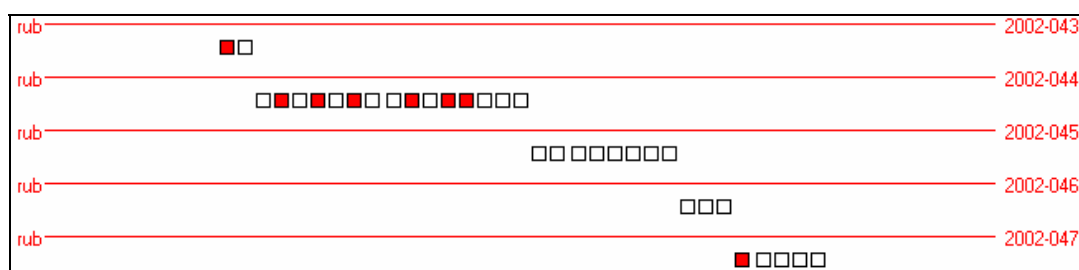


Figure 95

Carte des sections de Lexico représentant un extrait du corpus RA découpé en fonction des rubriques d'origine et occurrences des variantes de « Predica »

Commentaire de la figure 95 : Sur cette carte des sections de Lexico, chaque carré correspond à un paragraphe. Ces derniers sont répartis en fonction du découpage original des textes des rapports d'activité en rubriques ; par exemple, la notation « 2002-043 » correspond à la rubrique numéro 43 dans l'ordre d'apparition des rubriques du rapport d'activité de l'année 2002. La figure ne représente qu'un extrait du corpus limité aux rubriques 43 à 47 de la partie 2002. Les paragraphes colorés en rouge comportent au moins une occurrence de la dénomination « Predica » ou « PREDICA ».

On observe dans la figure ci-dessus que la rubrique « 2002-044 » concentre un nombre élevé d'occurrences de la dénomination-pilote, signalées par les paragraphes colorés en rouge. Néanmoins, tous les paragraphes de la rubrique ne

position d'ouvreur intertitulaire. C'est du reste un cas de figure qui se rencontre fréquemment dans différents genres (on renvoie à *Seuils* de G. Genette [1987 : 271-292]).

sont pas colorés et il peut être utile dans ce cas d'en vérifier la raison, afin de voir si ces paragraphes vierges d'occurrences ne recèlent pas des dénominations de secteurs d'activité qu'il serait intéressant d'intégrer à la collecte. Ainsi, le premier paragraphe de cette rubrique déroge-t-il au cas le plus fréquemment observé, pour lequel le nom de secteur d'activité et la dénomination propre-pilote sont rassemblés dans le même paragraphe : « *Assurance-vie : nouvelle progression de l'activité et des résultats. Après le ralentissement de l'année 2001, le marché français de l'assurance-vie a connu une croissance modérée en 2002, avec un chiffre d'affaires en progression de 1% et des encours gérés en hausse de l'ordre de 5%. L'évolution du marché a été affectée par la dégradation des marchés financiers qui a favorisé la croissance des contrats en euros (+13%) au détriment des contrats en unités de compte (-31%).* » De même, le troisième paragraphe en partant de la droite, bien que non sélectionné, présente des noms de secteurs d'activité candidats : « (...) *de saisir les opportunités de croissance sur de nouveaux marchés : la prévoyance collective, les grandes entreprises, la retraite complémentaire par capitalisation, le développement de l'activité ADI, y compris dans sa gestion* ». L'exploration manuelle des rubriques comportant des zones non couvertes par une série d'occurrences de l'unité pilote permet éventuellement d'enrichir la collecte avec des unités rares. Dans une perspective de veille, on complète ainsi utilement les sondages réalisés à partir des données obtenues à l'aide du calcul des spécificités. Cependant, dans la perspective de réutiliser ces données pour guider l'organisation de l'information, la sélection d'une unité candidate doit encore être validée à partir de critères intégrant la manière dont les noms de secteurs d'activité sont utilisés à l'échelle du corpus entier.

b. Domaines, marchés, secteurs : identifier la norme du corpus

On a indiqué précédemment que la collecte est réalisée sans idée préconçue sur ce que doit être un secteur d'activité. C'est l'étude du fonctionnement discursif de certaines unités, utilisées dans le but d'introduire ou d'évoquer des univers de discours, qui a guidé la sélection initiale des dénominations de secteurs d'activité. Ce faisant, on a été conduit à admettre que l'activité pouvait être découpée aussi bien en fonction de domaines d'activité – « le crédit à la consommation » –, qu'en fonction de marchés – « les particuliers ». Certaines de ces unités sont susceptibles de passer d'une catégorie à l'autre dans le même texte par un jeu métonymique. Il

est ainsi question du « marché de l'affacturage », du « secteur de l'affacturage », des « métiers de l'affacturage », mais aussi du « domaine de l'affacturage ». La réalisation de la collecte sur un mode itératif permet de dégager progressivement une norme pour les dénominations de secteurs d'activité dans le corpus. Cette norme est perceptible lorsque, par exemple, on hésite à retenir des syntagmes tels que « cartes privatives ». En effet, on a pu observer que parmi les composants de l'ontologie générale retenue dans ce travail, les objets bancaires, contrairement aux opérations, sont plus rarement retenus pour former des désignations de secteur d'activité. Afin de valider la sélection des unités candidates, il a donc été nécessaire d'établir des critères explicites qui reflètent les différentes matérialisations discursives de la norme dont il vient d'être question.

Pour cela, on a distingué les critères permettant de caractériser directement un syntagme comme nom de secteur d'activité, et ceux pour lesquels cette caractérisation se fait de manière indirecte. La première sorte de critères a déjà été présentée dans la section précédente : il s'agit des unités introduisant des univers de discours que l'on rencontre comme circonstants ou comme thèmes désignés en début de paragraphe. On peut ajouter à ces deux configurations principales le cas des séries. Ces dernières sont constituées lorsqu'un catégoriseur de secteur d'activité s'applique de façon distributive à une série d'unités, comme dans le dernier exemple donné plus haut (« de nouveaux marchés : »), ou bien lorsqu'une unité candidate est insérée dans une série de dénominations de secteurs déjà identifiés.

Les critères indirects correspondent à des utilisations de l'unité candidate qui suggèrent l'existence d'un secteur d'activité, sans pour autant permettre de l'appréhender directement. Ces indices ne peuvent donc jouer qu'un rôle auxiliaire destiné à étayer une candidature ayant déjà à son actif un critère direct. Le premier critère indirect correspond à la situation dans laquelle la dénomination de secteur d'activité est le complément du nom d'un catégoriseur (société, banque, filiale, partenaire, leader, produit, offre, etc.), appliqué à une personne morale ou à une autre sorte d'entité nommée, comme dans ces exemples :

« Au total, l'encours de Predica, premier bancassureur de France, a progressé de 24% à 2 11 milliards de francs. La société a ainsi accru sa part de marché de 0,7% à 9,7% des encours, confirmant sa deuxième place parmi les sociétés d'assurance-

vie de l'Hexagone. », ou bien encore, « *L'acquisition de Finaref fait du Crédit Agricole un des leaders du crédit à la consommation en France et en Europe.* »

En effet, il est possible de déduire de ces compléments qu'il existe probablement des marchés hexagonaux ou européens de l'assurance-vie et du crédit à la consommation.

Le second critère indirect rassemble une série d'indices fournissant une indication relative sur l'existence d'un secteur d'activité. Il peut s'agir d'un syntagme apparaissant dans un intertitre utilisé pour découper les textes du corpus en rubriques. On a vu plus haut qu'il peut s'agir d'un critère direct lorsque le nom de secteur figure dans l'intertitre et qu'il fait l'objet d'une reprise à l'aide d'un catégoriseur dès le premier paragraphe de la rubrique. Cette condition suffit à indiquer qu'il ne peut pas y avoir de caractérisation automatique du secteur d'activité du fait de la seule présence d'un syntagme nominal dans un intertitre. Cet indice est donc relatif. De fait, certains de ces intertitres, comme « Message du Président » ou « Chiffres clés », ne peuvent pas correspondre à un secteur. De même, le « sujet » annoncé par l'intertitre peut être, ou trop général « Le Crédit Agricole en 2000 », ou trop spécifique, comme « Fait marquant : le lancement de TLJ » pour un nom de secteur d'activité. Un autre indice est fourni par la reprise d'un tel nom dans un intitulé d'unité (direction ou département) appartenant à l'une des principales entreprises du groupe bancaire, comme « Marché des entreprises et des collectivités locales », où deux noms de marchés sont coordonnés pour former le nom d'une direction. Enfin, comme cela a déjà été noté dans la deuxième partie, l'attribution d'une majuscule à un terme désignant une sorte d'opération ou d'objet constitue une indication à prendre en compte, tout particulièrement dans les contextes à dominante non syntactique.

Lorsque les noms de secteurs d'activité ont été établis, et surtout lorsqu'il a été nécessaire pour cela de recourir à des indices *ad hoc* comme ceux qui viennent d'être mentionnés, leur intégration dans le référentiel terminologique est soumise en définitive à un critère de récurrence pondérée.

3. Critères de sélection des unités

Les critères que l'on présente ci-dessous tiennent compte de la nature des informations collectées et de leur récurrence dans le corpus. La métrique qui leur a été appliquée est par conséquent adaptée à ce que l'on recherche. En revanche, le principe consistant à établir des critères explicites, sur la base de la nature de l'information et de sa récurrence, revêt une portée plus générale. Ces critères sont différents selon qu'il s'agit de catégorisation ou de rattachement à un secteur d'activité¹. Pour la catégorisation, les unités sont systématiquement retenues en ce qui concerne les noms d'objets, compte tenu des fréquences plutôt basses qui caractérisent cette classe. Les catégorisations de personnes morales sont également largement accueillies, à condition toutefois qu'elles ne présentent pas un caractère trop anecdotique, généralement sanctionné par un statut de hapax, comme dans l'exemple précédent au sujet de Sofinco « spécialiste du crédit à la consommation ». On a vu plus haut que les unités catégorisantes permettent de présenter différentes facettes de la même entité. À ce titre, elles présentent généralement un intérêt, soit pour l'enrichissement du référentiel terminologique, soit pour illustrer la manière de catégoriser en usage dans le corpus.

Le rattachement à un secteur d'activité fait l'objet de critères particuliers, car les unités candidates sont parfois plus difficiles à caractériser, et leur sélection revêt une portée différente. En effet, retenir un secteur, c'est proposer une manière d'organiser des informations qui est susceptible d'être prise en compte pour l'organisation des sites d'un intranet (principaux embranchements de l'arborescence présents sur les pages d'accueil). La sélection des unités candidates de secteurs d'activité en vue de leur intégration dans un référentiel terminologique repose donc sur plusieurs critères. Il faut tout d'abord que l'unité candidate ait été repérée à partir d'au moins un critère direct, tel que ceux qui ont été présentés plus haut.

Un deuxième contrôle tient compte de la façon dont la relation est construite en fonction des différentes sortes d'entités nommées. On a vu que dans l'ontologie générale qui a été retenue pour ce travail, les entreprises productrices occupent une

¹ En ce qui concerne les noms de clientèles, les critères de sélection sont voisins de ceux utilisés pour les secteurs d'activité.

place première par rapport à leurs productions. On retrouve cette hiérarchie dans les textes du corpus, où la relation avec le secteur d'activité est établie de manière plus directe lorsqu'il s'agit d'une dénomination propre de personne morale. Par conséquent, pour les produits, on a adopté comme règle que les secteurs candidats relevés au moment de la collecte devaient par ailleurs être des secteurs utilisés pour la restitution de l'activité des entreprises. En ce qui concerne la collecte réalisée à partir de ces dernières, on a vérifié que l'unité candidate possède une fréquence supérieure ou égale à trois, et une récurrence constatée sur deux parties au moins du corpus. Cela permet d'éliminer des unités dont l'apparition dans le corpus peut revêtir un aspect trop conjoncturel qu'il n'a pas paru nécessaire d'introduire dans un référentiel.

En revanche, le traitement est différent pour les unités présentes dans la dernière partie du corpus. En effet, le corpus étant une série textuelle chronologique ouverte, la dernière partie mérite une attention particulière. Lorsqu'elle contient une unité candidate absente du reste du corpus, il n'est pas possible d'appliquer à cette dernière le contrôle de récurrence évoqué plus haut. Cependant, il serait dommageable de priver le référentiel d'une source d'enrichissement susceptible de révéler une évolution de l'activité bancaire. Par conséquent, les unités candidates présentes seulement dans la dernière partie du corpus ne sont pas soumises au test de récurrence, mais doivent valider les autres critères (au moins un critère direct de qualification attesté, et une fréquence supérieure à trois). Ce n'est que lorsqu'une unité candidate a rempli ces différents critères qu'elle a été intégrée au prototype de référentiel terminologique qui a été réalisé.

B. Intégration des informations collectées dans un référentiel terminologique

1. La catégorisation de l'entité nommée et la relation générique

On a vu précédemment que la collecte porte sur deux sortes d'unités : les dénominations propres utilisées comme unités-pilotes et les unités candidates, qui sont, soit des unités catégorisantes, soit des noms de secteurs d'activité ou de clientèle. Il s'agit bien d'intégrer ces deux sortes d'unités dans un référentiel terminologique, ce qui suppose que pour chacune d'elles les relations établies dans les discours soient transposées à l'aide des relations sémantiques utilisées pour relier

les termes entre eux. Dans le cadre de ce travail, on s'est limité aux questions relatives au rattachement des dénominations propres aux unités candidates. Pour cela, on a adopté comme hypothèse de travail que ces dernières (unités catégorisantes et noms de secteurs ou de clientèles) sont susceptibles soit d'être rattachées aux autres unités contenues dans un référentiel terminologique existant, soit de permettre la constitution d'un référentiel particulier¹.

Cette limitation précisée, il reste que la question du rattachement des dénominations propres pilotes aux unités candidates ne se pose pas dans les mêmes termes pour les deux sortes de relations que l'on a entrepris d'étudier. On fera donc une distinction dans le traitement selon que la relation établie dans les discours avec l'entité nommée pilote relève de la catégorisation ou du rattachement à un secteur d'activité (ou à une forme de clientèle).

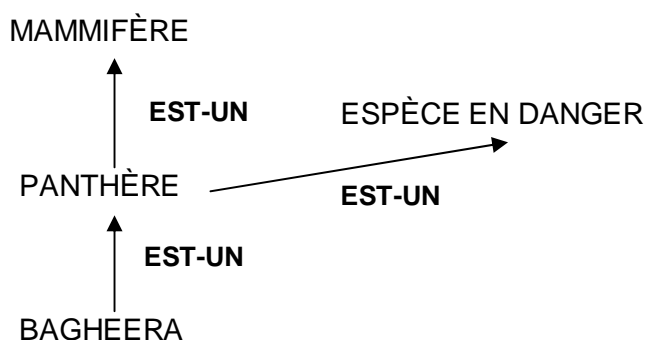
Les objets et les personnes morales dénommés à l'aide des unités-pilotes utilisées pour la collecte sont catégorisés dans les discours du corpus. Cette relation, lorsqu'elle est transposée dans un référentiel terminologique, est qualifiée de relation générique. Dans un tel référentiel, cette relation est censée relier entre eux des concepts. Or, la définition de la dénomination propre qui a été retenue plus haut ne permet pas d'avancer que toutes les unités-pilotes utilisées pour la collecte sont dotées d'un concept². Pour faire usage de cette relation dans un référentiel, il faut admettre au préalable qu'elle s'applique ici non pas à des liens établis entre concepts, mais entre des objets (les particuliers visés par les dénominations propres) et des concepts. Il s'agit par conséquent d'une relation d'instance à concept, au sens où l'on dit avec G. Frege qu'un objet « tombe » sous un concept. Cette restriction

¹ Ce point soulève une question plus générale qui porte sur la destination des référentiels terminologiques et sur leur évolution possible. Leur ouverture aux différentes strates d'un parler d'entreprise est susceptible de remettre en question certaines règles adoptées pour leur constitution. On ne tranche pas ce point ici, d'autant que la collecte réalisée est susceptible de rendre les services que l'on attend d'elle, soit dans le cadre d'un référentiel terminologique existant « réformé », soit dans le cadre d'une nouvelle sorte de référentiel dédié à la collecte et à l'étude de l'usage attesté dans un contexte. Ce point fait actuellement l'objet de débats relatifs à l'évolution des procédures d'organisation de l'information (on renvoie pour ce point à S. Dalbin [2007]).

² En effet, on a vu que les noms de produits sont susceptibles de deux saisies différentes selon le contexte, pointant tantôt le nom déposé, tantôt le terme. En revanche, les noms d'entreprises sont pris comme des noms propres prototypiques, sans qu'il ait paru nécessaire de distinguer entre dénominations pures et impures (cf. K. Jonasson [1994 : 34 et suiv.]), pour les cas où la présence de noms communs dans la dénomination pouvait suggérer une motivation qui, la plupart du temps, n'entretient un rapport que très lointain avec la réalité qu'ils sont censés décrire.

importante entraîne des limitations dès lors que l'on souhaite introduire des dénominations propres dans un référentiel terminologique.

La relation d'instance à concept caractérise donc une relation générique non pas d'inclusion mais d'appartenance¹. La relation hiérarchique générique « est un » est utilisée en terminologie pour caractériser la relation établie entre un concept incluant, le générique, et un concept inclus, le spécifique². De cette relation découlent des possibilités d'héritage exploitables par des systèmes automatisés. En effet, la relation hiérarchique peut être utilisée par un automate, par exemple, dans le domaine de la recherche d'informations, afin de réaliser des expansions de requêtes³. Pour que ces héritages soient valides, il est nécessaire que la propriété de transitivité puisse leur être appliquée, comme dans cet exemple emprunté à J.-P. Desclés :



¹ On renvoie sur ce point à G. Frege (1971 [1892]), J.-P. Desclés (1987) et G. Otman (1996 : 74-78).

² « Une relation générique existe entre deux concepts lorsque la compréhension du concept subordonné inclut la compréhension du concept superordonné et au moins un caractère distinctif supplémentaire. Le concept superordonné d'une relation générique est appelé concept générique et le concept subordonné est appelé concept spécifique.

Dans une relation générique, il existe un rapport inverse entre la compréhension d'un concept et son extension. Par conséquent, si la compréhension d'un concept est restreinte, son extension sera relativement plus large et inversement, si la compréhension est large, l'extension sera relativement plus restreinte ». 5.4.2.2 Norme Iso 704 :2000, p.6.

³ Dans un système de recherche d'informations, une expansion de requête est réalisée lorsque le mot ou l'expression saisis dans un moteur de recherche par l'utilisateur sont complétés par d'autres mots ou expressions tirés d'un dictionnaire ou d'un thésaurus et qui sont utilisés pour la recherche dans l'index du moteur. Par exemple, si la requête est « panthère » et que le moteur de recherche est associé à un automate dont les traitements s'appuient sur un réseau sémantique du type de celui qui est donné en exemple, l'index sera parcouru à l'aide de « panthère » mais aussi du terme spécifique qui lui a été préalablement associé « Bagheera », et éventuellement, du terme générique « mammifère ».

Figure 96.

Extrait du graphe de « Bagheera est-un » d'après J.-P. Desclés¹

Dans ce graphe, la relation « est-un » relie de façon correcte « Bagheera² » à « Panthère » puis à « Mammifère », car on vérifie que si Bagheera est une panthère et que la panthère est un mammifère, alors Bagheera est un mammifère. De même, il sera possible de considérer que le groupe des mammifères comprend les panthères dont l'une d'elles est nommée Bagheera. En revanche, si la panthère est une espèce en danger, il ne sera pas correct de dire la même chose de Bagheera. J.-P. Desclés illustre son raisonnement à l'aide de cet exemple afin de conclure à la non transitivité de la relation « est-un », et par conséquent, à son caractère inadapté pour les systèmes d'héritage reposant sur des relations hiérarchiques. G. Otman, pour sa part, réduit l'utilisation de la relation aux seuls rapports d'inclusion entre concepts, ce qui, comme on vient de le voir, exclut les dénominations propres qui ne peuvent être reliées aux autres unités d'un référentiel que par un rapport d'appartenance. Dans la mesure où l'on cherche néanmoins à intégrer les dénominations propres aux référentiels terminologiques, il apparaît nécessaire de bloquer les relations d'héritage les concernant au niveau n+1, ce qui revient à retenir « Panthère » mais à exclure « Mammifère » pour reprendre la relation établie dans l'exemple cité, « Bagheera » est-un « Panthère » est-un « Mammifère ». Il semble qu'à cette condition les dénominations propres puissent être reliées aux autres termes présents dans un référentiel³. Il reste à envisager maintenant le type de relation pouvant être établie entre une dénomination propre et un nom de secteur d'activité.

¹ J.-P. Desclés (1987 : 56), et G. Otman (1996 : 77-78).

² Dans les raisonnements de J.-P. Desclés et de G. Otman, il va de soi que Bagheera désigne une panthère. Dans l'approche retenue dans ce travail, une telle conclusion aurait nécessité une étude du fonctionnement discursif de la dénomination propre « Bagheera » afin d'établir son contenu. Il faut donc faire crédit aux auteurs cités de la notoriété de l'ami de Mowgli !

³ La norme ISO 704 parvient à une conclusion similaire concernant les appellations désignant des concepts uniques (concepts dont l'extension ne correspond qu'à un seul objet) : « Dans un système générique de concepts, ils occupent le plus bas niveau hiérarchique. » Norme ISO 704 :2000, 7.4, p. 29.

2. Le rattachement d'une entité nommée et les relations associatives

Comme pour la précédente relation, il est nécessaire de discuter les questions soulevées par l'usage que l'on propose de faire de la relation terminologique devant formaliser le lien établi en discours, entre une dénomination propre et un nom de domaine d'activité ou de clientèle. Pour le secteur d'activité, on pourrait être tenté de traduire ce lien par une relation de type « partie-tout », étant donné que les secteurs d'activité fournissent les cadres dans lesquels a cours le fonctionnement économique de certaines entités. Néanmoins, la norme précitée rappelle qu'une telle relation ne peut être établie que si elle repose, non pas sur une fonction discursive consistant à ouvrir des domaines de discours, mais sur les traits caractéristiques des concepts que l'on cherche à relier par cette autre relation hiérarchique : « *On considère qu'il existe une relation partitive lorsque le concept superordonné représente un tout, et que les concepts subordonnés représentent des parties de ce tout. Les parties s'assemblent pour former le tout (...)*¹. » Bien qu'abordant cette question en adoptant un point de vue lexicologique, I. Tamba considère également que la relation « partie/tout » appartient aux relations logiques, c'est-à-dire qu'elle doit apparaître dans la définition du mot et fait donc partie de sa signification. Sur ce dernier point, cet auteur distingue cependant la situation du méronyme (le concept subordonné), qui intègre normalement la relation dans sa définition, de celle des holonymes, qui en général ne l'intègrent pas. Même si cet auteur reconnaît que les propositions peuvent contribuer à conférer à un mot son statut de méronyme, cet apport externe reste dans un cadre « définitionnel ». Il en va de même lorsque l'on cherche à établir cette relation à partir de marqueurs présents dans les textes : il s'agit de retrouver certains traits définitionnels, qui sont ensuite validés dans un cadre terminologique².

Or, comme pour la relation générique, l'absence de définition, en langue commune ou spécialisée, associée aux dénominations propres, constitue un obstacle important à leur intégration dans un référentiel terminologique. Le contenu de la

¹ Norme ISO 704 :2000, 5.4.2.3, p. 10.

² Voir G. Otman (1996 : 82) pour les manifestations linguistiques de « partie de » et A. Condamines, J. Rebeyrolle (2000 : 225-241), dans une perspective de fouille de textes réalisée en vue de la constitution d'une base de connaissances terminologiques.

dénomination ne peut pas se substituer à une telle définition dans la mesure où il repose sur les seules indications fournies par les discours. Certes, l'usage tolère ce type de relation, au moins en ce qui concerne les toponymes, pour lesquels on a coutume d'emboîter les villes dans les pays, et ces derniers dans les continents ou les structures politiques ou économiques qui les englobent. Cependant, il semble difficile d'alléguer cet usage dans le cas d'un référentiel spécifique rassemblant des unités ne possédant pas, comme les noms géographiques, une fonction de repères dans un référentiel spatial fondamental, partagé par une grande partie des locuteurs. On a donc recherché une autre sorte de relation pour relier les dénominations propres aux secteurs d'activité.

En effet, on propose de relier les deux sortes d'unités par une relation associative, reposant non plus sur un emboîtement conceptuel hiérarchique, mais sur les données de l'expérience, celles-ci pouvant être de natures diverses : « *Les relations associatives ne sont pas hiérarchiques. En effet, une relation associative existe lorsque l'expérience permet l'établissement d'un lien thématique entre des concepts. Certaines relations associatives existent lorsqu'un lien de dépendance est établi entre des concepts en raison d'une proximité spatiale ou temporelle. Ces relations peuvent être du type matière première - produit, action - équipement/outil, grandeur - unité, matériau - propriété, matériau - état, matière/substance - propriété, élément concret - matériau, élément concret - forme, action - cible, action - site/emplacement, action - acteur, etc. Certaines relations peuvent porter sur des événements dans le temps tels qu'un processus en fonction du temps ou d'une chronologie; d'autres encore peuvent être du type cause - effet¹.* »

Parmi les autres exemples cités par la norme, la relation « contenant – contenu » semble convenir au type de relation que l'on souhaite formaliser dans un référentiel². Les secteurs d'activité jouent alors le rôle de contenants possibles pour

¹ Norme ISO 704 :2000, 5.4.3, p. 13.

² En ce qui concerne la relation entre dénomination propre de produit et nom de clientèle, c'est une autre sorte de relation associative qui peut être utilisée : produit – clientèle ou débouché du produit. Pour la relation établie entre dénominations propres de personnes morales et noms de clientèle, on a le schéma parallèle producteur – débouché. On voit que dans ce cas, ce que fait l'entité X pour la clientèle Z peut fonctionner parallèlement à la relation unissant l'entité X à un marché A. On peut juger ces liens redondants, et c'est pourquoi, comme le suggère la restitution de l'activité dans le corpus, on a plutôt réservé l'association au nom de clientèle pour les dénominations propres de

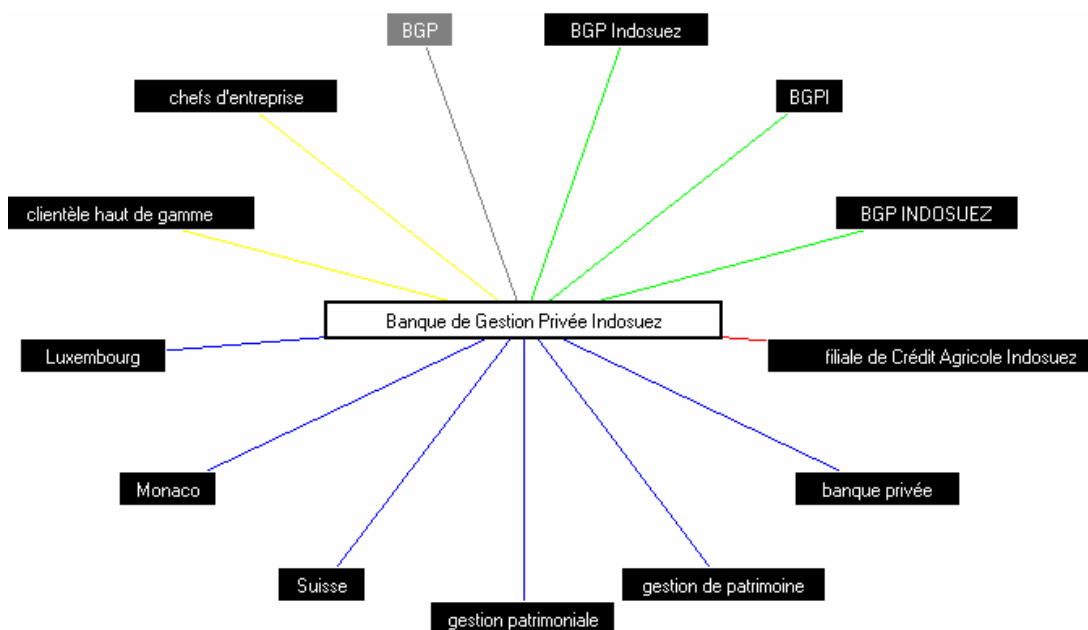
certaines objets et personnes morales. Cette souplesse a néanmoins pour inconvénient de bloquer l'héritage, ce dernier étant réservé aux relations hiérarchiques. Cette règle étant observée, il n'en reste pas moins que la relation associative est dotée d'un potentiel informationnel aussi important que la relation hiérarchique, car elle permet de multiplier les passerelles entre dénominations propres et termes.

Dès lors qu'elle est coupée de l'arbre complet des relations hiérarchiques, la dénomination propre intégrée à un référentiel terminologique apparaît dans un réseau étoilé à un seul niveau, dont elle constitue le centre. En ce sens, les relations ainsi construites reflètent l'image de la structure type de contenu élaborée à partir des relations établies dans et par les textes du corpus autour des dénominations propres. La présentation de ces relations par un programme informatique permet de restituer graphiquement cette organisation rayonnante¹. Les dénominations propres intégrées à un référentiel terminologique semblent jouer à la fois le rôle de feuilles terminales et celui de boucles rétroagissant vers les autres unités du réseau auxquelles elles offrent un nouvel accès. En effet, les unités n+1 auxquelles les dénominations sont rattachées n'occupent pas nécessairement une position terminale avant l'intégration des dénominations propres. On donne ci-dessous, à titre d'illustration, le réseau d'une dénomination propre de personne morale, établi dans le cadre d'un prototype de référentiel terminologique².

produits, et l'association au nom de secteur d'activité pour les dénominations propres de personnes morales. Il ne s'agit là que d'une option pragmatique destinée à ne pas alourdir exagérément les tâches de la collecte.

¹ On renvoie aux restitutions du référentiel terminologique données en annexe.

² Ce prototype est consultable dans les annexes. Il a été réalisé, d'une part, à partir de données stockées dans un fichier texte tabulé et, d'autre part, en important ces données dans le logiciel Inxight Star Tree Studio version 2.0 de la société Inxight Software, Inc. Il ne s'agit que d'une traduction possible des données de la collecte. Celles-ci pourraient être intégrées, par exemple, dans un gestionnaire de référentiel adapté aux langages du Web sémantique (tels que les langages SKOS - Simple Knowledge Organization System pour les thésaurus) du type du logiciel open source *THManager* (<http://sourceforge.net/projects/thmanager>). Le prototype que l'on propose a pour avantages sa robustesse, sa facilité d'utilisation et un format de restitution adapté à l'intranet et au travail des webmestres.



*Figure 97.
Représentation graphique des unités associées à « Banque de Gestion Privée Indosuez » dans un prototype de référentiel terminologique*

Commentaire de la figure 97 : la dénomination propre-pilote « Banque de Gestion Privée Indosuez » est placée au centre du réseau. Une partie des unités qui composent son contenu, dans le corpus RA lui sont rattachées par un lien dont la couleur correspond à un type de relation : vert pour les variantes (par ex. « BGPI »), rouge pour la relation « est un » (« filiale de Crédit Agricole Indosuez »), bleu pour le rattachement à un secteur d'activité (par ex. « banque privée »), jaune pour le rattachement à une clientèle (par ex. « clientèle haut de gamme ») et, enfin, gris pour mentionner une ancienne dénomination (« BGP »).

Les données affichées dans ce graphique proviennent d'une feuille de données Excel, mais il pourrait s'agir d'une base de données à part entière. Ces données sont organisées en fonction de la procédure d'import qui permet la restitution présentée ci-dessus. Cette procédure¹ consiste à répartir l'information principale sur deux colonnes, la première colonne portant l'information de départ du lien et la seconde celle de l'arrivée. De cette manière, il est possible de créer une arborescence qui se développe dans la feuille de données ligne par ligne sous la forme $A \rightarrow B$, puis $B \rightarrow C$, $C \rightarrow D$, etc. Les deux colonnes principales dans

¹ *Inxight Star Tree Studio Version 2.0 – Studio User Guide*, p. 49-76. © Copyright 1999-2001 by Inxight™ Software, Inc. All rights reserved.

lesquelles sont déclarées les unités-pilotes et les unités collectées sont complétées par des colonnes supportant certains attributs. Ces derniers se ramènent principalement à des chiffres correspondant à un codage des relations et des états. Ce codage est interprété visuellement lors de la restitution HTML. On a, par exemple, la relation « est un » qui est codée avec le chiffre « 2 » lui-même interprété et restitué à l'affichage sous la forme d'un lien rouge. De la même manière, le cartouche encadrant les unités du référentiel est noir pour celles qui sont actives et gris pour les dénominations obsolètes. On donne ci-dessous un extrait du tableau de collecte des dénominations propres de personnes morales, concernant l'unité « BGPI ».

Tableau 84.

Extrait du tableau de collecte des données pour l'unité-pilote « Banque de Gestion Privée Indosuez » destiné à une restitution graphique

Niveau n	Niveau n+1	nœud	texte	lien
Banque de Gestion Privée Indosuez	BGP Indosuez	0	1	3
Banque de Gestion Privée Indosuez	BGPI	0	1	3
Banque de Gestion Privée Indosuez	BGP INDOSUEZ	0	1	3
Banque de Gestion Privée Indosuez	filiale de Crédit Agricole Indosuez	0	1	2
Banque de Gestion Privée Indosuez	banque privée	0	1	4
Banque de Gestion Privée Indosuez	gestion de patrimoine	0	1	4
Banque de Gestion Privée Indosuez	gestion patrimoniale	0	1	4
Banque de Gestion Privée Indosuez	Suisse	0	1	4
Banque de Gestion Privée Indosuez	Monaco	0	1	4
Banque de Gestion Privée Indosuez	Luxembourg	0	1	4
Banque de Gestion Privée Indosuez	clientèle haut de gamme	0	1	6
Banque de Gestion Privée Indosuez	chefs d'entreprise	0	1	6
Banque de Gestion Privée Indosuez	BGP	12	1	12

Commentaire du tableau 84 : la première colonne « Niveau n » reprend la dénomination propre de l'unité pilote qui se trouve au centre du réseau dans le graphique. La colonne « Niveau n+1 » contient les unités collectées qui lui sont rattachées dans le référentiel terminologique. La colonne « nœud » contient le codage de la couleur de fond du cartouche : noir pour une unité active, gris pour une unité obsolète. La colonne « texte » porte le code couleur du texte, blanc pour toutes les unités. Ces couleurs sont inversées pour le cartouche central. La colonne « lien » contient le codage des relations restituées sous la forme de couleurs : 3 (vert) pour la relation d'équivalence, 2 (rouge) pour la relation « est un », 4 (bleu) pour la relation associative de rattachement à un secteur d'activité, 6 (jaune) pour la relation associative marquant la liaison avec un clientèle, 12 (gris) pour un lien rattachant une unité active à une unité obsolète.

III. ÉVALUATION DE LA COLLECTE

A. Caractéristiques générales de la collecte

1. Évaluation de la productivité de la collecte

a. La collecte varie selon le type d'entité désignée par une dénomination propre

La collecte qui a été réalisée peut être présentée dans ses aspects généraux sous deux angles. Le premier, quantitatif, livre une évaluation de la collecte des unités candidates en termes de productivité. Le second angle est plutôt d'ordre

qualitatif, dans la mesure où il propose une estimation de l'apport des données rassemblées autour d'une dénomination propre, en vue d'une réutilisation pour traiter des problématiques d'organisation de l'information.

Le premier enseignement de la collecte qui a été réalisée est relatif à la position inégale des classes de dénominations propres. En effet, on constate que les cinquante-deux dénominations propres d'objets ont permis de collecter 227 unités, alors que pour un nombre d'unités-pilotes équivalent (cinquante-cinq), les dénominations propres de personnes morales permettent d'obtenir plus du double avec 595 relations établies. Ces résultats vont dans le sens de ce qui avait déjà été obtenu au moment de l'établissement des structures-types de contenu des dénominations propres. Toutes les classes de dénominations propres ne sont pas appelées à jouer le même rôle dans la restitution de l'activité bancaire telle qu'elle est présentée dans les textes du corpus. On peut tenter de résumer ces observations à l'aide d'un coefficient de productivité de collecte pour une classe d'entités nommées. Ce coefficient correspond seulement au rapport du nombre de relations établies sur le nombre de dénominations pilotes. On obtient pour la classe des « objets » un coefficient de 4,4 et de 10,8¹ pour les « personnes morales », soit une dénomination propre d'objet pour un peu plus de quatre unités collectées, et une dénomination propre de personne morale pour environ onze unités collectées. Cette première approche de la productivité de la collecte peut ensuite être affinée en tenant compte de la manière dont chaque classe est à même de capter plus ou moins d'unités à partir de telle ou telle relation.

b. La productivité varie en fonction des relations étudiées

Le classement des relations types associant les dénominations-pilotes aux autres unités révèle de nouvelles disparités. On donne ci-dessous la répartition des types de relations établies lors de la collecte pour les deux classes de dénominations-pilotes.

¹ On a respectivement pour les objets, un coefficient de $\frac{227}{52} = 4,4$ et pour les personnes morales, un coefficient de $\frac{595}{55} = 10,8$.

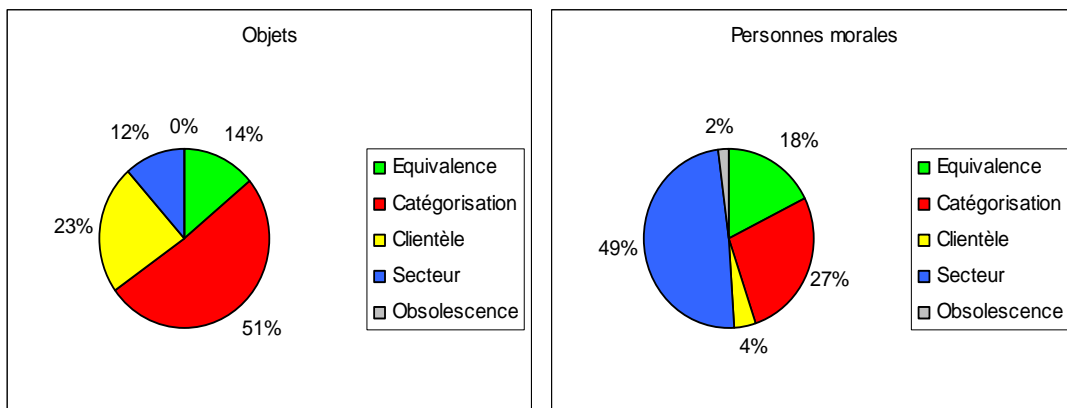


Figure 98.
Répartition des relations-types selon les classes d'unités-pilotes¹

La relation d'équivalence, utilisée pour relier entre elles les variantes d'une même dénomination propre, est la seule à se retrouver dans des proportions comparables pour les deux classes d'entités nommées. Cependant, cette proximité dans la répartition ne doit pas masquer un autre phénomène. En effet, on a déjà remarqué plus haut que l'apparition des variantes est globalement liée à la répétition de la dénomination. La reprise avec variantes peut être limitée à une seule partie, mais elle s'observe plus régulièrement d'une partie sur l'autre. Cela explique probablement que cette relation se rencontre dans des proportions moindres pour les dénominations d'objets – dont on a vu qu'elles sont de fréquences faible à moyenne, ce qui limite mécaniquement leur apparition à une ou deux parties du corpus –, par rapport aux dénominations propres de personnes morales, dont les plus basses fréquences retenues pour la collecte sont supérieures à vingt occurrences². La relation de catégorisation appelle des remarques du même ordre.

En effet, la catégorisation semble être la relation privilégiée pour les objets lorsqu'elle est comparée autres relations. Cependant, les rapports observés sont

¹ Les couleurs adoptées dans ces graphiques reprennent celles qui sont utilisées dans le référentiel tel qu'il est restitué au format HTML et dont on a donné un exemple dans la précédente section. Chaque relation est affectée du code couleur suivant : relation d'équivalence : vert ; relation de catégorisation : rouge ; relation de rattachement à une clientèle : jaune ; relation de rattachement à un secteur d'activité : bleu ; relation devenue obsolète du fait de la disparition de la dénomination d'une entité dans la période couverte par le corpus : gris.

² Pour les dénominations d'objets, on observe 31 variantes pour 52 dénominations, et pour les dénominations de personnes morales, on observe 104 variantes pour 55 dénominations.

d'une dénomination pour deux catégorisations en ce qui concerne les objets, et d'une pour trois pour les personnes morales¹. L'utilisation des catégorisations est néanmoins différente pour chaque classe. En effet, on a vu précédemment que pour les objets, ces catégorisations multiples permettent à l'énonciateur de présenter un même objet sous des aspects différents. Cela lui permet également de jouer sur des strates différentes du parler d'entreprise, comme pour le produit « Agilor » catégorisé par ordre croissant de précision comme offre, prêt, puis défini comme « crédits prescripteurs développés avec les constructeurs et les distributeurs de matériel agricole ». La catégorisation des personnes morales semble répondre au double souci de marquer le lien capitalistique qui les rattache au groupe bancaire, et de préciser leur nature. Ce dernier usage, qui par définition n'est jamais neutre, est parfois détourné explicitement à des fins de modalisation ; le Crédit agricole peut en effet être présenté comme « banque française » mais aussi comme le « banquier du logement », le « partenaire de l'agriculture » ou comme un « groupe uni et décentralisé » occupant une place de « leader ». De la même manière, le rattachement à une clientèle ou à un secteur d'activité varie sensiblement d'une classe à l'autre.

On constate ainsi, que le rattachement à une clientèle, s'il n'est pas absent pour les personnes morales, est caractéristique des objets². Ces derniers correspondant majoritairement à des produits, le « patron » sous-jacent qui a été relevé est du type « tel produit est destiné à telle clientèle ». Les deux classes s'opposent encore sous le rapport du rattachement à un secteur d'activité. Ce type de relation semble en effet être la caractéristique principale des dénominations propres de personnes morales³. Cette fois, le patron sous-jacent est « telle entreprise du groupe bancaire exerce son activité dans tel secteur ».

Enfin, on remarque que pour cette dernière classe, une relation nouvelle a été introduite afin de tenir compte du fait que sur la période couverte par le corpus, certaines personnes morales sont sorties du champ des rapports d'activité. La

¹ On a 116 catégorisations pour 52 objets et 160 catégorisations pour 55 personnes morales.

² Pour 52 dénominations propres d'objets, on enregistre 52 rattachements à une clientèle et 26 à un secteur d'activité.

³ Pour 55 dénominations propres de personnes morales, on compte 23 rattachements à une clientèle et 288 à un secteur d'activité.

principale cause de ces disparitions réside dans les fusions opérées à l'occasion de l'acquisition de nouvelles entreprises, ou dans la rupture du lien capitalistique par cession de la participation du Crédit agricole.

Dans le premier cas de figure, ces rapprochements se traduisent fréquemment par l'adoption d'une dénomination différente pour l'entité nouvellement créée. La collecte enregistre donc des données qui peuvent devenir obsolètes pour une partie de la période couverte par la série textuelle chronologique. Cette « obsolescence » de la dénomination est prise en compte dans le référentiel, car sa disparition d'un corpus ne signifie pas qu'elle ait disparu d'autres corpus (cas d'une rupture du lien capitalistique), ni qu'elle soit simultanément sortie du référentiel propre à l'entreprise et du souvenir qu'en a le personnel utilisant l'intranet. Dans le prototype de référentiel terminologique qui a été réalisé dans le cadre de cette étude, on a vu que les relations sont codées par des couleurs. L'obsolescence a été traduite par le gris que l'on a utilisé pour colorer le cartouche encadrant chaque dénomination propre, ainsi que le lien unissant la dénomination obsolète, soit avec la dénomination de la classe (personnes morales), soit avec la dénomination propre de l'entité qui lui a succédé. En revanche, les relations unissant la dénomination obsolète avec d'autres unités sont conservées telles quelles. Le référentiel terminologique permet ainsi de conserver sous une forme schématisée la mémoire des mouvements qui affectent le référentiel de l'organisation¹.

¹ La question de l'obsolescence des mots, pendant de la néologie, a été identifiée depuis longtemps en lexicologie, même si c'est surtout la néologie qui a été mise en avant. Sa description et sa prise en compte se rencontrent dans de nombreux travaux. On ne citera ici que les exemples de Peter J. Wexler (1955) qui, à propos du vocabulaire des chemins de fer, rend compte des proliférations synonymiques liées à un état de la technique, et qui évoluent vers une certaine uniformité de la terminologie (1955 :129) ; ou bien encore J. Dubois (1962), qui décrit le même phénomène à propos du vocabulaire politique et social sous le Second Empire et au début de la III^e République (1962 : 129 et suiv.). R. Eluerd (1993 : 369-370) propose le terme de paléologie pour les unités lexicales qui sortent de l'usage. Cet aspect est traité avec parcimonie dans les normes relatives aux langages documentaires (NF Z 47100 : 1981, 7.2.2., p. 18, et ISO 2788 : 1986, 10.5.3, p. 30), et sous l'aspect utilitaire du service rendu ou non rendu par le terme pour la recherche ou l'indexation. Dans la norme ISO 704 : 2001, ce point est abordé de manière indirecte comme conséquence de la normalisation terminologique qui impose une hiérarchie entre les désignations (un terme peut être privilégié, toléré ou déconseillé) : « *L'introduction d'une désignation dans une norme de terminologie doit entraîner l'abandon implicite des autres désignations comme synonymes dans le domaine en question. Il est sage de considérer ces termes et de les identifier de manière explicite comme des termes tolérés ou déconseillés et d'exposer, dans la mesure du possible, les motifs de l'abandon.* » (Ibid. p. 39). On voit

2. Données collectées et organisation de l'information

a. Catégorisations multiples

La collecte met en évidence que les dénominations propres d'objets et de personnes morales sont respectivement associées à deux ou trois catégorisations du référent qu'elles désignent¹. Comme on l'a vu précédemment, ces proportions relativement voisines correspondent à des utilisations assez différentes : présentations des facettes de l'objet dans le premier cas, nature et lien capitalistique dans le second. On sait que G. Kleiber a fait l'hypothèse à plusieurs reprises « *que les particuliers porteurs d'un nom propre sont saisis prioritairement (et non nécessairement !) dans une catégorie basique*². » Sans que l'on ait cherché à vérifier cette hypothèse à l'aide de la collecte qui a été réalisée, cette dernière permet de compléter les termes de l'hypothèse.

Tout d'abord, on remarque que plusieurs catégorisations appartenant vraisemblablement au même niveau – ce point peut en effet être débattu –, sont disponibles pour un même particulier : « Predica » est une compagnie d'assurance-vie, un bancassureur et une filiale. Mais on remarque également la présence de niveaux de catégorisation différents : « Floriane » est un produit, un contrat, un contrat d'assurance-vie et un contrat d'assurance-vie multisupport. Enfin, les catégorisations multiples ont fréquemment pour fonction de présenter plusieurs facettes du même objet : « Pleinchamp » est un site, un site Internet, une communauté virtuelle et un service bancaire sur Internet.

que la disparition d'une raison sociale pose la question dans des termes nouveaux pour les référentiels terminologiques.

¹ Pour les objets, on dénombre 116 catégorisations pour 52 dénominations ; pour les personnes morales, on compte 160 catégorisations pour 55 dénominations propres.

² G. Kleiber (1996 : 582) et (2004 : 137). La théorie d'Eleanor Rosch à laquelle se rattache la notion de concept de base a été discutée par G. Kleiber dans *La sémantique du prototype* (1990), ouvrage auquel on renvoie tant pour l'exposé et la discussion des deux versions de la théorie que pour la bibliographie. On se borne à rappeler ici que les différentes versions de cette théorie tentent de répondre à la question : comment et pourquoi range-t-on tel objet dans telle catégorie ? Les concepts susceptibles d'être utilisés sont organisés en trois ensembles hiérarchiques : niveau superordonné, niveau de base et niveau subordonné. G. Kleiber illustre cette organisation à l'aide des exemples suivants :

SUPERORDONNÉ	animal	fruit	meuble
NIVEAU DE BASE	chien	pomme	chaise
NIVEAU SUBORDONNÉ	boxer	golden	chaise pliante

G. Kleiber (1990 : 83).

En résumé, on constate la surdétermination de la catégorisation. À une échelle macroscopique, celle-ci est déterminée par la situation de communication dans laquelle elle s'inscrit, et qui est représentée ici par le genre des rapports d'activité. Ce dernier induit par ailleurs un traitement discursif différent pour les différentes classes de dénominations propres. Perceptible lors de la mise au jour des structures-types, ces différences de traitement sont vérifiées par la productivité de la collecte. À un niveau intermédiaire, constitué par les textes ou leurs grandes subdivisions, on remarque que l'éclairage contextuel est susceptible de mobiliser des catégorisations concurrentes pour un même particulier. On verra plus loin que la dimension temporelle doit également être prise en compte pour expliquer les variations introduites dans la manière de catégoriser le même objet.

Enfin, on a également noté que la catégorisation peut être déterminée par la dénomination propre elle-même, en ce sens que l'utilisation d'une signalétique particulière est susceptible de bloquer une catégorisation qui pourrait apparaître redondante, comme pour « Livret jeune » ou « Crédit Agricole Epargne Salariale ». L'établissement d'un référentiel terminologique ne prétend naturellement pas trancher les questions théoriques soulevées par ce point. En revanche, il vise à restituer les usages dans leur diversité, quand bien même ils sont limités à une même série textuelle. On retrouve ce phénomène dès lors que l'on examine le rattachement d'une entité aux secteurs d'activité.

b. Diversité des rattachements d'une même entité

Si le rattachement des entités nommées à des secteurs d'activité ou à des types de clientèles semble être la règle observée pour les unités-pilotes, la précédente section a permis de mettre en évidence que la réalisation de cette relation varie de manière importante en fonction des classes. En effet, ces rattachements se font dans une proportion de 1 à 5 pour les dénominations propres de personnes morales, alors que le rapport n'est que d'un pour un en ce qui concerne les dénominations d'objets. Les noms de secteurs d'activité revêtent une importance particulière pour les problématiques liées à l'organisation de l'information. Ils traduisent, sans la refléter, une conception particulière du découpage de l'activité d'un groupe bancaire. Or, on a de bonnes raisons de penser que ce découpage entretient un rapport étroit avec les manières d'organiser l'information dans une

situation de communication semblable. La collecte et son résultat organisé dans le référentiel offrent alors une vision synthétique des usages attestés disponibles pour l'organisation de l'information relative à la restitution de tel ou tel pan d'activité bancaire. Comme pour la catégorisation, le fait saillant est la grande diversité des rattachements observables. Il est néanmoins possible de distinguer certaines régularités.

En effet, le rattachement d'une personne morale aux secteurs d'activité se fait selon deux axes principaux. Le premier propose un découpage de l'activité en domaines et en marchés. Ainsi, la « gestion d'actifs », secteur d'activité, peut être orientée vers le marché des particuliers ou celui des entreprises. Ces marchés sont susceptibles d'être découpés en zones géographiques correspondant également à des marchés ou à des zones économiques, comme on l'a vu dans le chapitre consacré aux toponymes. Ce premier découpage est complété par une présentation plus ou moins détaillée de l'activité. Pour chaque niveau de détail, un nom de secteur plus ou moins précis est ajouté. On a ainsi la « banque de financement et d'investissement », les « financements de projets » et en définitive, juste avant la présentation de telle ou telle opération d'envergure, les « financements aéronautiques », ou « maritimes ». De la même manière, on rencontre des emboîtements du type « banque de proximité », « moyens de paiement », et « cartes bancaires », qui chacun à leur niveau servent de cadre à une restitution plus ou moins détaillée de l'activité. Outre la coexistence d'approches concurrentes pour organiser la restitution de l'activité bancaire, la collecte permet ainsi de mettre au jour dans un même contexte des présentations complémentaires, ayant des degrés de précisions différents et dotées de leurs propres dénominations.

Ces deux caractéristiques révélées par la collecte, catégorisations et rattachements multiples des unités pilotes doivent maintenant être évalués par rapport aux attentes initiales. On rappelle que celles-ci visaient à obtenir, d'une part, un échantillon d'unités fidèle aux manifestations du parler d'entreprise observées dans le corpus et, d'autre part, l'accès aux principaux découpages de l'activité véhiculés par les rapports.

B. L'accès au vocabulaire du corpus

1. Rappel des limites apportées par les procédures utilisées pour la collecte

Un autre apport de la collecte doit consister à fournir des unités représentatives de l'utilisation du parler d'entreprise tel qu'il se manifeste dans les textes du corpus. Or, dans la mesure où la procédure de collecte à partir des unités pilotes s'appuie sur un certain nombre de filtres, on peut se demander si l'utilisation de ces derniers n'a pas eu pour conséquence de privilégier une strate au détriment des autres.

Cette évaluation de la collecte ne vise pas à déterminer si elle a permis de couvrir en volume ou en variété un échantillon représentatif du vocabulaire du corpus. Il s'agit plutôt de vérifier que les variétés de langues sont bien restituées parmi les unités collectées. Afin d'étudier ce point, il convient de rappeler les procédures de sélection des unités pilotes qui, de fait, condamnent la collecte à n'être pas représentative de l'ensemble du vocabulaire.

On a vu dans les précédents chapitres que la collecte est organisée à partir de dénominations propres pilotes et des relations qu'elles entretiennent avec d'autres catégories d'unités. Les relations sélectionnées dans les structures de contenu types des dénominations propres d'objets et d'agents correspondent à deux types d'opérations réalisées dans les discours : celle de catégorisation et celle de rattachement à un secteur d'activité ou à une clientèle. Par cette procédure, certains composants de l'activité bancaire sont exclus a priori¹. C'est le cas des opérations, des ensembles complexes, des propriétés et d'une partie des éléments composant l'ensemble hétéroclite baptisé « divers ». Par ailleurs, parmi les composants retenus, une double sélection a encore été opérée. Parmi les « agents », seules les personnes morales ont été retenues, la structure de contenu type des dénominations propres de personnes physiques ayant paru moins intéressante pour la collecte à réaliser. Enfin, seules les dénominations propres d'objets et de personnes morales ayant respectivement une fréquence supérieure à 3 et à 20 ont été retenues.

¹ Cela ne signifie pas que des unités désignant des objets appartenant à ces classes ne puissent pas être collectées lorsqu'elles entretiennent avec les unités-pilotes l'une des deux relations étudiées.

On voit qu'il n'y aurait pas grand sens à rechercher une quelconque représentativité de la collecte après ces opérations de sélection successives. Il reste que la question de la représentativité peut être posée sur un autre plan, celui des variétés de langue. Dans la deuxième partie, certaines strates appartenant au parler du groupe bancaire avaient été recensées. On résume ces observations dans le tableau ci-dessous.

*Tableau 85.
Variétés de langue attestées dans le corpus Rapports d'activité*

Variété de langue	Présence dans les textes du corpus
Langue commune	bien représentée <i>Ex. : banque</i>
Restrictions de sens et « économismes »	bien représentés <i>Ex. : développement [économique]</i>
Termes banalisés et usages « génériques »	bien représentés <i>Ex. : cartes bancaires</i>
Terminologie bancaire générale	assez bien représentée <i>Ex. : banque privée</i>
Terminologie bancaire précise	mal représentée <i>Ex. : acceptation de banque</i>
Terminologie propre au Crédit agricole	peu représentée <i>Ex. : caisse locale</i>

Ce sont ces catégories que l'on utilise dans les sections suivantes, afin d'évaluer la manière dont elles sont représentées par les unités collectées. Les observations sont organisées à partir des relations de catégorisation et de rattachement identifiées dans les structures-types des contenus des dénominations propres d'objets et de personnes morales.

2. Variétés de langue et catégorisation

D'une manière générale, on peut considérer que les variétés de langue présentes dans le corpus sont représentées par les unités collectées dans des proportions similaires à celles caractérisant le vocabulaire pris dans son ensemble. Il convient cependant de nuancer cette observation en tenant compte des deux classes de dénominations pilotes. Pour ce qui est des dénominations de produits, les unités catégorisantes appartiennent massivement à la langue commune, soit directement

(marque, offre, produit), soit avec une restriction de sens courante adaptée au domaine économique (contrat, gamme, service). Les termes banalisés sont également bien représentés (cartes, fonds, prêts). À elles seules, ces trois catégories couvrent 90% des unités utilisées pour catégoriser les produits.

Les proportions sont moindres en ce qui concerne les personnes morales. En effet, ces trois catégories couvrent seulement, si l'on peut dire, les deux tiers des unités catégorisantes. On peut attribuer ce phénomène au fait que, de façon générale, les noms de catégories d'entreprises opérant dans le secteur bancaire sont moins répandus que ceux des produits qui touchent directement un large public. On constate ainsi que la terminologie bancaire générale est assez utilisée pour caractériser les entreprises du groupe bancaire (« banque commerciale », « banque d'affaires »). On relève également des caractérisations propres au Crédit Agricole, mobilisées lors de la présentation des organes constitutifs du groupe bancaire (« caisse centrale du Crédit Agricole »), ainsi que des termes utilisés dans la banque « de gros » (arrangeur, lead manager), tout aussi peu connus par le grand public.

3. Variétés de langue et rattachement des entités nommées

Comme pour la catégorisation, les variétés de langue mobilisées par les unités collectées à partir de la relation de rattachement se rencontrent dans des proportions voisines de celles qui ont été observées pour le vocabulaire du corpus. Il faut noter cependant que, d'une manière générale, les unités appartenant à la langue commune, ou dont l'usage dans le corpus repose sur une sélection de traits sémantiques liés au domaine économique, sont présentes dans des proportions moindres. En effet, elles occupent respectivement pour les dénominations d'objets et de personnes morales les trois quarts et la moitié seulement des effectifs collectés.

En ce qui concerne les référents des unités appartenant à la classe des objets, leur rattachement à une clientèle ne peut se faire pratiquement que s'il s'agit de produits. Or, cinq d'entre eux n'en sont pas, puisque l'on recense deux indices boursiers, un document, un ratio et un texte de loi. Par ailleurs, on constate qu'un tiers des produits ne sont pas rattachés à un nom de clientèle mais de secteur d'activité. Cela concerne principalement des produits de filiales spécialisées (assurance-vie, produits financiers), dont la présentation n'est pas réservée à

l'exposition de l'offre destinée aux différentes clientèles, mais sert surtout à illustrer l'activité de tel ou tel secteur. Or, on constate par ailleurs que les noms de secteurs d'activité se répartissent à parts égales entre langue commune et terminologie bancaire générale.

En effet, les dénominations de personnes morales sont rattachées à des noms de secteurs d'activité relevant pour la moitié d'entre eux de la terminologie bancaire générale ou précise. Cette proportion est d'autant plus significative que l'on a pris pour option de comptabiliser les noms de pays, utilisés pour délimiter des marchés nationaux, parmi les unités de langue commune, en considérant que le référentiel géographique est le domaine des dénominations propres le mieux partagé entre interlocuteurs. Comme pour les unités catégorisantes, la présentation des activités relatives à la banque de gros est responsable pour une large part de la présence de noms de secteurs relevant d'une terminologie bancaire précise et peu connue du grand public. On rencontre ainsi parmi les dénominations de secteurs, des unités telles que « capital investissement », « financements structurés », « multigestion », « négoce international », « titrisation », etc.

La présentation des secteurs d'activité recourt également à la terminologie bancaire générale. Contrairement à ce que l'on observe pour les noms utilisés afin de désigner des sortes de produits ou des types de clientèles, qui sont des entités globalement visibles et connues pour un large public, les dénominations de secteurs d'activité bancaires semblent bien avoir pour centre de gravité l'univers de la banque. Ce phénomène s'explique en partie par le fait que le découpage de l'activité dans le cadre des rapports d'activité est aussi une affaire de positionnement par rapports aux autres établissements bancaires. Pour l'énonciateur, il convient en effet de montrer que l'on occupe au moins les mêmes terrains que les concurrents, quitte à réutiliser leur terminologie.

Cet écart relatif par rapport aux caractéristiques du vocabulaire du corpus est fréquemment contrebalancé à l'échelle des unités-pilotes par des catégorisations et des rattachements multiples. Le cas de Crédit Agricole Indosuez permet d'illustrer ce point. En effet, parmi les unités catégorisantes on apprend que cette « banque d'affaires internationale » (terminologie bancaire générale) est également un « spécialiste en valeurs du Trésor » (terminologie bancaire précise), mais aussi une

« filiale » du Crédit agricole (langue commune avec sélection de traits relatifs au domaine économique). Pour le rattachement aux secteurs d'activité, il est indiqué que cette banque opère dans les domaines de la « dette souveraine » et de la « dette-change » (terminologie bancaire précise), ainsi que dans le secteur de la « banque de grandes clientèles » (terminologie bancaire générale) et celui des « marchés boursiers » (terme banalisé). On voit avec cet exemple, mais également avec ceux qui sont présents dans la base de données du référentiel ainsi que dans sa restitution sous la forme de réseau sémantique, que les unités-pilotes permettent d'accéder aux différentes variétés de langue présentes dans les textes du corpus. Il est maintenant nécessaire d'évaluer les apports de la collecte relativement à la restitution des façons de découper l'activité présentée dans les textes du corpus.

C. Mise en évidence de la complexité du découpage de l'activité

1. Construction d'un modèle théorique

L'un des apports importants de la collecte réalisée à partir des unités pilotes consiste dans la mise en évidence de la complexité du découpage de l'activité telle qu'elle est restituée dans les textes du corpus. Afin d'évaluer cet apport, on propose de construire un modèle théorique des différents secteurs dans lesquels l'activité d'une entreprise bancaire telle que le Crédit agricole est susceptible de se déployer, en tant que banque, mais aussi en tant qu'institution et unité économique. Ce découpage macroscopique sera ensuite confronté avec des unités caractéristiques de la collecte.

L'étude d'un lexique, comme l'élaboration d'outils dédiés à l'organisation de l'information, cherche à s'appuyer sur un découpage de l'activité humaine en cadres ou en domaines d'activité¹. Cependant, depuis les travaux de lexicologie sociale de G. Matoré² ou ceux d'E. Wüster³ pour la terminologie, force est de constater qu'un

¹ B. de Bessé (2000 : 184).

² G. Matoré (1953 et 1973 pour la préface de la deuxième édition). Celui-ci a d'ailleurs reconnu qu'il n'existe pas de cadre satisfaisant susceptible d'être utilisé afin de classer les faits de vocabulaire (1953 : 70 et suiv.).

³ E. Wüster (1974) et H. Felber (1987).

cadre structural a priori¹ n'a pas pu être dégagé de façon consensuelle. Néanmoins, dans un but heuristique, on peut chercher à construire un modèle que l'on fera contraster avec la façon dont les discours proposent des manières d'organiser l'activité bancaire, en ouvrant au fil des textes des univers de discours. Ce modèle vise à schématiser ce que pourrait être idéalement la présentation de l'activité d'un groupe bancaire tel que le Crédit agricole dans des rapports d'activité produits vers l'année 2000. On a vu précédemment que l'appareil des entrées de sommaires et des intertitres présents dans les textes du corpus ne peuvent pas fournir un tel découpage, soit parce qu'ils sont modifiés par des visées autres que le découpage de l'activité, comme la mise en valeur ponctuelle de certains événements, soit parce que l'on a constaté que leur absence dans les sommaires ne signifie pas que le pan d'activité qu'ils désignent ne se trouve pas dans les rapports d'activité. Il a donc été nécessaire de construire un modèle.

La construction d'un tel schéma a priori s'est heurtée à la difficulté prévisible, compte tenu de ce qui a déjà été indiqué plus haut, à disposer d'une source unique susceptible de couvrir l'ensemble des secteurs relatifs à l'activité bancaire, au fonctionnement institutionnel et à l'organisation interne de l'entreprise bancaire. Par conséquent, pour chacun des ces grands domaines, on s'est appuyé sur des sources externes différentes². Il s'agit tout d'abord du *Code monétaire et financier* daté de mai 2005. Ce code est mis à jour en permanence sur le site Legifrance³ de l'administration française. L'activité bancaire étant pour diverses raisons particulièrement encadrée et réglementée, il nous a semblé que la source primaire que constitue ce code ne devait pas être négligée. Cependant, il n'était pas possible d'exiger d'une telle source qu'elle intègre des tendances économiques n'entrant pas nécessairement dans son champ, mais qui constituent néanmoins une partie de la réalité de l'activité bancaire. On a donc utilisé pour cela une autre source

¹ À titre d'illustration, on peut citer la définition de « domaine » que propose la *Norme ISO 1087-1 :2001* : « *branche spécialisée de connaissance. Note : Les limites d'un domaine sont définies selon un point de vue particulier lié à l'objectif visé.* » On peut constater à l'aide de cette seule définition que la problématique s'est déplacée d'un modèle général d'inspiration sociologique ou de type wüsterien vers des considérations terminographiques *ad hoc*, sans pour autant que la question ait été tranchée de façon satisfaisante.

² Le commentaire détaillé des renseignements tirés des sources citées est donné en annexe.

³ *Code monétaire et financier* à 2005, disponible sur le site <http://www.legifrance.gouv.fr>.

tenant dans deux ouvrages de synthèse, *Le système bancaire français* et *Le système bancaire et financier – Approches française et européennes*¹. Enfin, on a pensé utile de compléter cette vision portant principalement sur l'activité extérieure d'un groupe bancaire par un éclairage portant sur les éléments liés au fonctionnement de l'entreprise bancaire. Autrement dit, on présuppose que la présentation de l'activité orientée vers la production de biens et de services, contrairement à l'impression qui se dégage des textes du corpus, pourrait également comprendre les moyens humains et matériels mis en œuvre. Afin de dresser un schéma sommaire de ces moyens, on a utilisé une partie de la hiérarchie du *Thesaurus Delphes*, de la Chambre de commerce et d'industrie de Paris². Cet ouvrage est utilisé pour la description du fonds documentaire de cet organisme collectant, entre autres, de l'information sur les entreprises et les secteurs économiques.

Compte tenu de ce que sont les usages contemporains, il est possible de distinguer d'un point de vue abstrait un certain nombre de secteurs englobés par la notion d'activité bancaire – on parlera également d'opérations bancaires. Celles-ci regroupent les activités de banque, principales et connexes, ainsi que les services d'investissement relatifs à la gestion des instruments financiers³. En outre, en dehors de ces activités bancaires à strictement parler, il existe des secteurs para-bancaires dans lesquels les banques diversifient leur activité depuis une vingtaine d'années⁴. Il est également nécessaire de tenir compte des caractéristiques propres aux banques mutualistes qui, quoique généralistes, se distinguent des banques commerciales traditionnelles par leur organisation⁵. Enfin, l'activité d'une banque ne se réduisant pas au seul exercice des métiers bancaires, il a été nécessaire d'examiner également les postes d'activité liés au fonctionnement d'une entreprise bancaire et qu'un

¹ A. Choinel et G. Rouyer (1996) et A. Choinel (2001).

² *Thesaurus Delphes* de la CCI de Paris, édition 2004.

³ On s'est appuyé sur les articles L311-1, L311-2, L321-1 et L321-2 du *Code monétaire et financier*, à 2005. Ce code est accessible à l'adresse <http://www.legifrance.gouv.fr>.

⁴ Pour l'énumération de ces formes de diversification de l'activité bancaire, on s'est appuyé sur A. Choinel (2001 : 206 et suiv.).

⁵ Le *Code monétaire et financier* a également été sollicité sur ce point. En effet, le Chapitre II du Livre V de ce code présente les caractéristiques et les mesures s'appliquant aux banques mutualistes ou coopératives dont le Crédit Agricole fait partie (en particulier, l'article L512-20, L512-34).

rapport d'activité serait susceptible d'aborder¹. Il existe différents classements relatifs à ce sujet, tous influencés par un point de vue particulier² (économie, gestion, comptabilité, droit, etc.); celui que l'on propose ici n'a pas d'autre prétention que de fournir une organisation commode pour le but que l'on s'est fixé.

On propose dans les tableaux ci-dessous une organisation idéale qui pourrait servir à la restitution de l'activité d'un groupe bancaire, sans qu'il soit tenu compte des contraintes s'appliquant au genre des rapports d'activité, qui ont recensées dans la première partie.

¹ Pour ce point, c'est une partie de la nomenclature du *Thésaurus Delphes* de la CCI de Paris, édition 2004 qui a été utilisée (les modules concernant l'entreprise et sa gestion : Entreprise, Gestion de l'entreprise, Ressources humaines, gestion de la production, gestion commerciale et gestion financière).

² On peut également ajouter que les notions d'« entreprise » et de « gestion » n'ont été abordées d'un point de vue théorique que relativement récemment, ce qui peut expliquer leur manque de stabilité. Pour un point de vue critique, on renvoie aux deux articles du *Dictionnaire de sociologie*, Le Robert – Seuil, Paris, 1999, p. 188-189 et 242-243.

Tableaux 86, 87 et 88

Découpage théorique de la restitution de l'activité d'un groupe bancaire mutualiste

L'activité bancaire	
Activité bancaire et opérations connexes	réception des fonds du public opérations de crédit moyens de paiement opérations de change valeurs mobilières et placements financiers conseil en gestion de patrimoine conseil en gestion financière location simple de biens services d'investissement gestion d'instruments financiers gestion de portefeuille pour le compte de tiers prise ferme et placement conservation d'instruments financiers conseil aux entreprises location de coffres-forts négociation de marchandises sous-jacentes à certains instruments financiers
Spécialisation de l'activité bancaire	crédit-bail mobilier crédit-bail immobilier crédit à la consommation crédit à moyen et long terme aux entreprises affacturage
Diversification et développement de l'activité bancaire	capital investissement assurance IARD assurance-vie gestion d'actifs banque en ligne international partenariats et acquisitions

Fonctionnement institutionnel	
Statuts	banque mutualiste société anonyme société cotée en bourse
Structures	groupe bancaire entités appartenant au groupe organigramme de l'organe central

L'entreprise bancaire	
Stratégie de l'entreprise	développement durable déontologie projet d'entreprise organisation et management communication interne, institutionnelle et financière
Gestion financière	risques comptabilité contrôle de gestion
Gestion du personnel / ressources humaines	conditions de travail rémunération embauches formation politique sociale
Gestion de la production	innovation / recherche qualité productivité organisation de la production
Gestion commerciale	marketing promotion commercialisation / vente
Logistique	télécommunications informatique locaux achats maintenance
Sécurité	personnes matériel biens immatériels

On peut maintenant tenter de confronter ce modèle abstrait avec les secteurs d'activité tels qu'ils ont pu être identifiés au cours de la collecte réalisée à l'aide des dénominations propres pilotes présentes dans les textes du corpus.

2. Confrontation du modèle avec les unités collectées

a. Occultation de l'entreprise bancaire

Les secteurs les moins bien couverts dans les textes du corpus appartiennent à la présentation de l'entreprise bancaire. Hormis certains points relatifs à la stratégie de l'entreprise, aux ressources humaines, ainsi qu'à la gestion financière, qui sont traités dans la partie financière des rapports annuels, les autres secteurs relevant de ce bloc sont généralement passés sous silence. On a vu dans la première partie que cela tient dans une large mesure au genre des rapports d'activité. En effet, dans ce type de discours, les moyens de production sont largement occultés pour des raisons de confidentialité, mais aussi parce qu'il est admis implicitement que la

communication ne se fait que sur les résultats et non sur les moyens mis en œuvre pour y parvenir. Dans les dernières années du corpus, la vogue du « développement durable » est responsable de l'apparition de nouvelles entrées relatives aux ressources humaines, donnant ainsi au secteur une nouvelle visibilité, alors que le sujet est évoqué dès les premières parties du corpus. Il en va de même pour le mécénat culturel et caritatif.

En ce qui concerne l'activité bancaire et les opérations connexes, un certain nombre de lacunes doivent également être constatées. Elles portent sur des opérations peu valorisées, soit à cause de leur faible importance économique, soit parce qu'elles revêtent un caractère plus qualitatif et sont de ce fait moins facilement chiffrables, ce qui dans le contexte des rapports d'activité a pour effet de les rendre moins visibles. Parmi les premières, on peut citer les « opérations de change » (celles qui sont réalisées au guichet, car ce qui relève de la « banque de gros » est traité avec ce qui est relatif aux marchés financiers), la conservation de titres, la location simple de biens ou la location de coffres-forts qui n'apparaissent pas dans le corpus. Les secondes portent essentiellement sur les services de type « conseil ».

De cette première confrontation, il ressort que les rapports d'activité ne constituent pas une ressource pertinente dès lors que l'on cherche à appréhender les secteurs d'activités relevant du fonctionnement interne de l'entreprise. D'autre part, le phénomène de valorisation-occultation, présenté dans les deux parties précédentes, fournit une explication valable pour certaines lacunes. Enfin, elle a permis de déceler des flottements terminologiques, par exemple pour la partie relative aux instruments financiers. En effet, on constate que la terminologie de la source (ici le *Code monétaire et financier*) n'est pas reprise en tant que telle, les rapports d'activité lui préférant une terminologie plus vague, comme « gestion d'actifs¹ ».

b. Spécialisation et diversification de l'activité bancaire

¹ « gestion d'actifs » ne compte que 34 occurrences dans un corpus réunissant les rapports d'activité de trois autres banques sur la même période, alors que le syntagme lexicalisé compte 184 occurrences, toutes graphies confondues, dans le corpus RA des rapports d'activité du Crédit Agricole pour la période 1995-2003.

Il reste qu'un certain nombre de secteurs sont couverts par la collecte conformément au modèle. On le constate pour les domaines traditionnels de la banque (crédit, épargne et moyens de paiement), comme pour les secteurs relevant d'une certaine spécialisation ou d'une diversification de l'activité bancaire. On en donne ci-dessous l'illustration pour ces deux derniers points.

*Tableau 89
Confrontation d'un découpage théorique de l'activité bancaire en secteurs avec les unités afférentes collectées*

Modèle théorique		Unités collectées
Spécialisation de l'activité bancaire	crédit-bail mobilier	crédit-bail mobilier
	crédit-bail immobilier	crédit-bail immobilier
	crédit à la consommation	crédit à la consommation
	crédit à moyen et long terme aux entreprises	banque de financement et d'investissement
	affacturation	affacturation
Diversification et développement de l'activité bancaire	capital investissement	capital investissement
	assurance IARD	assurance IARD
	assurance-vie	assurance-vie
	gestion d'actifs	gestion d'OPCVM ; gestion d'actifs, assurances et banque privée ; gestion d'actifs
	banque en ligne	banque à accès multiples ; banque à distance
	international	participations / bancaires / internationales / prises de , banque de grandes clientèles et de l'international ; international et grandes clientèles ; négoce international ; financements spécialisés
partenariats et acquisitions	partenariats ; participations	

La couverture complète de ces deux pans d'activité du modèle théorique par les unités collectées peut recevoir au moins deux explications. La première tient aux sources qui, en raison de leur nature – il s'agit d'ouvrages de synthèse consacrés à l'activité bancaire –, sont probablement beaucoup plus réceptives à la terminologie

ambiante du secteur que le *Code monétaire et financier*, qui est tenu à un respect strict de la terminologie juridico-financière. La seconde explication tient probablement au fait que la spécialisation des activités, ainsi que leur diversification, font en quelque sorte partie du « standing » d'un groupe bancaire important dont les rapports d'activité sont chargés de valoriser l'activité. Cette bonne couverture permet également de confirmer que le niveau de maille retenu pour les secteurs d'activité dans le modèle théorique correspond, au moins en partie, à celui des rapports d'activité. C'est cependant l'apport principal de la collecte que de faire ressortir, parallèlement à cette relative concordance, des différences importantes de granularité dans les noms de secteurs qui ont été collectés.

c. Microsecteurs et macrosecteurs d'activité

En effet, la collecte réalisée à partir des unités-pilotes révèle l'utilisation parallèle, d'une part, de microsecteurs, servant à découper un secteur de niveau intermédiaire, comme ceux qui sont recensés dans le modèle théorique, et d'autre part, de macrosecteurs permettant de regrouper plusieurs de ces secteurs intermédiaires. Les trois niveaux de secteurs attestés dans le corpus sont complétés par des secteurs transversaux. Ces derniers opèrent une coupe horizontale dans la précédente organisation et permettent de rassembler des informations communes à des secteurs distincts. Enfin, ce double découpage de l'activité révèle toute sa complexité lorsque l'on met en évidence des dénominations concurrentes pour un même secteur. Ces croisements multiples se combinent généralement pour mettre les dénominations propres de personnes morales au centre d'un maillage révélant plusieurs facettes de leur activité.

Les microsecteurs apparaissent lorsqu'un secteur de niveau intermédiaire est détaillé. Ainsi, les « moyens de paiement » peuvent être découpés en « cartes bancaires », en « monétique » ou en « gestion des flux » selon l'actualité. De même, le « crédit-bail » est plus ou moins subdivisé selon les années, sans que l'on puisse déceler une progression qui pourrait correspondre à un développement de l'activité. On peut trouver à côté de « crédit-bail », des dénominations de sous-secteurs tels que « crédit-bail mobilier » et « crédit-bail immobilier », mais aussi « crédit-bail matériel », « location de longue durée » ou encore, « location de longue durée

automobile ». À l’opposé de ces microsecteurs, les rapports d’activité découpent également l’activité en macrosecteurs.

Ces secteurs englobants permettent de rassembler plusieurs secteurs de niveau intermédiaire sans pour autant les faire disparaître. C’est, par exemple, le pôle « assurances » qui vient chapeauter « assurance-vie » et « assurance IARD », ou le pôle « services financiers spécialisés » qui est placé au-dessus de « crédit à la consommation », « crédit-bail », et « affacturage ». Ces macrosecteurs possèdent la double propriété de correspondre fréquemment à des créations redondantes¹ et tardives, puisque les secteurs d’activité intermédiaires leur préexistent généralement dans la série textuelle chronologique, comme l’illustre ce graphique de ventilation des fréquences :

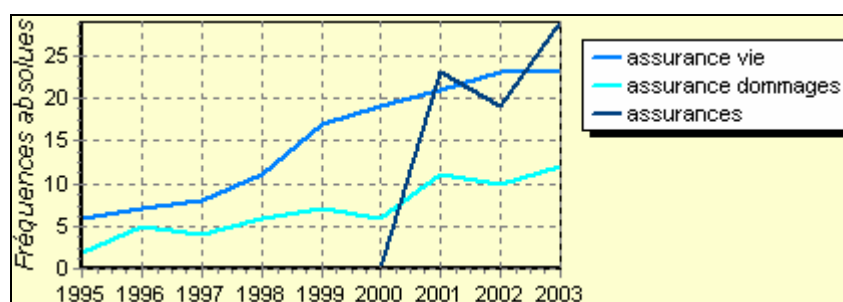


Figure 99.

Ventilation en fréquences absolues de trois groupes de formes² sur les neuf parties du corpus de rapports d’activité

Ces regroupements semblent pouvoir s’expliquer par deux causes. La première tient vraisemblablement au souci de présenter une image équilibrée du groupe bancaire, qui tient compte de son développement. L’acquisition d’Indosuez en 1996, qui a conféré au groupe bancaire une véritable dimension internationale, a entraîné un regroupement des marchés locaux traditionnels du Crédit agricole sous

¹ On rencontre également des dénominations telle que « Gestion d’actifs, assurances et banque privée » qui cumule des défauts rédhibitoires en matière de maniabilité, en incorporant une virgule, une coordination et la juxtaposition de dénominations existantes.

² Les trois groupes de formes correspondent respectivement à « assurance vie » et « Assurance vie », « assurance dommages », « Assurances dommages » et « IARD », et pour le dernier, à « assurances » et « Assurances ».

la dénomination « Banque de proximité¹ ». On observe le même phénomène lors des acquisitions importantes marquant l'arrivée ou le renforcement de la banque dans un secteur d'activité. Ces regroupements semblent résulter également d'un alignement sur les standards internationaux qui servent à caractériser l'activité des établissements bancaires cotés en bourse. Le changement de statut de l'organe central du groupe bancaire en 2001 s'accompagne de la mise en avant de nouveaux macrosecteurs, comme la « banque de détail à l'étranger ».

Parallèlement à ces découpages hiérarchisés de l'activité d'un secteur, deux découpages horizontaux sont utilisés pour regrouper les informations relatives à un secteur. Mais avant de les présenter, on peut considérer que le découpage temporel propre aux rapports annuels, qui séquence la restitution de l'activité année après année, fournit un macro-découpage transverse. Cependant, compte tenu, d'une part, de leur caractère global (c'est toute la matière relative à une année qui est rassemblée dans un rapport annuel), et d'autre part, de leur relative neutralité par rapport à la restitution de l'activité des différents secteurs, les millésimes ne se présentent pas comme des secteurs, au même titre que l'affacturage ou le crédit à la consommation. En revanche, il n'en va pas de même pour les noms de pays qui sont utilisés comme découpeurs de secteurs. En effet, à la suite de ce qui a été constaté dans le chapitre dédié à l'usage des toponymes dans le corpus, les noms de pays (et de certaines entités présentant en outre un caractère politique et économique), peuvent assumer ce rôle, lorsqu'ils sont utilisés comme marchés, en plus de leur fonction de repères géographiques. Ils fonctionnent dans les textes à la manière des secteurs d'activité, en ouvrant des univers de discours, « *En Grèce, le Crédit Agricole a conclu un accord stratégique (...)* », en permettant des regroupements de dénominations d'entités, « *Portugal : Banco Espírito Santo, Credibom, Tranquilidade* », ou en étant intégrés à des dénominations de secteurs d'activité dont ils précisent l'extension, comme « *banque de proximité en France* » ou « *le marché français de l'assurance-vie* ».

¹ Dans la première partie, lorsqu'il a été question du découpage des rapports d'activité en rubriques, on a pu voir que ces bouleversements sont répercutés sur les entrées de sommaires. Par ailleurs, voir le chapitre 6.

Ce découpage géographico-mercantile fonctionne comme une sorte de modèle génétique des coupes opérées en fonction des débouchés économiques, qui constituent une autre forme de répartition transversale de l'activité. Deux sortes de dénominations de marchés sont utilisées dans les textes du corpus. La première repose sur une utilisation métonymique du découpage des populations clientes de la banque en grands ensembles, afin de désigner les marchés correspondants. C'est, par exemple, le syntagme « la clientèle des entreprises » (qui ne se limite pas aux entreprises clientes), qui fournit le nom de secteur « marché des entreprises ». Cette dénomination sert à ouvrir des univers de discours, comme dans cet exemple « *Sur le marché des entreprises, les capitaux gérés en épargne longue des salariés (...)* ». Son ellipse connaît le même emploi, comme dans : « *Les entreprises : consolider des relations fondées sur proximité et qualité des services (...)* ». Enfin, la dénomination de marché fonctionne parallèlement comme dénomination propre d'unité comme, par exemple, « *Entreprises, Collectivités locales & Professionnels* ». L'autre sorte de dénominations de marchés est construite à l'aide du même mécanisme métonymique, mais cette fois ce sont les noms d'objets associés au fonctionnement de la banque et des marchés financiers qui sont utilisés. On rencontre alors aussi bien le « marché de l'habitat » que les « marchés d'actions » ou les « marchés d'obligations ».

Ce foisonnement de découpages est en quelque sorte redoublé par la coexistence de plusieurs dénominations pour un même secteur. C'est ainsi que « crédit-bail », dont il a été question plus haut, fonctionne de façon parallèle avec son équivalent anglais « leasing », dont les occurrences sont néanmoins cinq fois moins importantes. Par ailleurs, cette cohabitation concerne également les quasi-synonymes, dont certains ont été présentés dans la deuxième partie, à propos de la phraséologie des rapports d'activité. Ainsi, la dénomination « banque à domicile » désigne-t-elle la même chose que « banque à distance », ces deux dénominations étant par ailleurs très proches de « banque à accès multiples¹ » et de « banque en

¹ La ventilation de « banque à distance », « banque en ligne » et « banque à accès multiples » est donnée dans la seconde partie, dans le chapitre 6.

ligne » ; de même, « banque proximité » cohabite avec « banque de détail », comme le montre le graphique de ventilation ci-dessous :

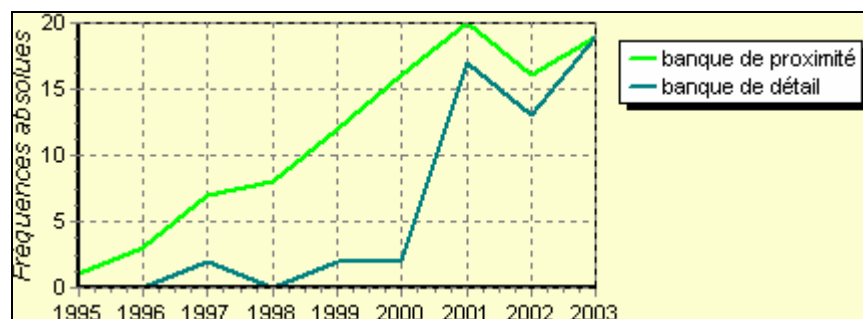


Figure 100.

Ventilation des occurrences de « banque de proximité » et « banque de détail » sur les parties du corpus de rapports d'activité

Les dénominations de personnes morales se trouvent par conséquent rattachées de façon presque systématique à plusieurs noms de secteurs d'activité¹. Chacun d'eux présente un intérêt dès lors qu'il s'agit d'évoquer l'activité de l'entité nommée. De ce point de vue, ils constituent autant de points d'accès à la dénomination. On donne ci-dessous, à titre d'illustration, une représentation des relations entretenues par une entité nommée et ses différents secteurs d'activité.

¹ Il existe bien sûr des exceptions. Celles-ci permettent d'isoler des entités qui n'interviennent pas au même titre que les autres dans les différents secteurs d'activité, soit que leur fonction soit principalement institutionnelle, comme la Fédération Nationale de Crédit Agricole, soit qu'elles n'appartiennent pas au même cercle ontologique, comme l'État.

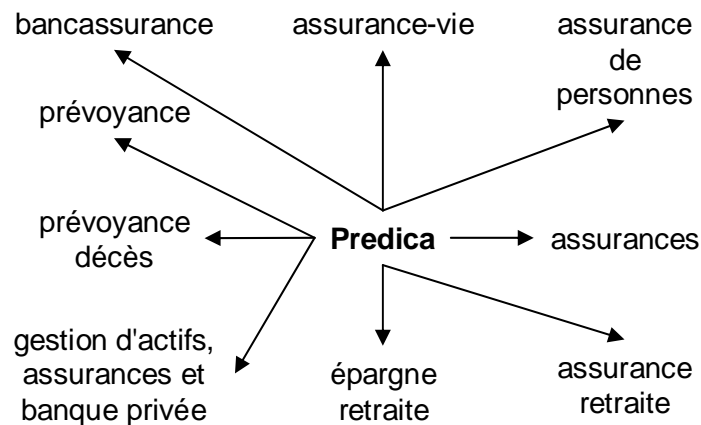


Figure 101.

Noms de secteurs d'activité rattachés à la dénomination propre « Predica » dans le corpus de rapports d'activité

Les différents rattachements dont la dénomination propre « Predica » fait l'objet mettent en évidence l'utilisation de plusieurs découpages hétérogènes utilisés pour la présentation d'un secteur d'activité. La restitution de l'activité emprunte ainsi aussi bien aux macrosecteurs (« bancassurance », « assurances » et « gestion d'actifs, assurances et banque privée ») qu'aux microsecteurs (« prévoyance », « prévoyance décès »). On note également la présence de secteurs parallèles (« assurance retraite », « épargne retraite »), pour lesquels on retrouve le même flottement terminologique que pour « assurance-vie » et « assurance de personnes ». Dans cet ensemble, le nom de secteur correspondant au niveau intermédiaire (« assurance-vie ») apparaît non plus comme le seul moyen de caractériser l'activité de l'entreprise Predica, mais comme une possibilité parmi d'autres. La collecte réalisée à partir d'une dénomination pilote offre ainsi une restitution schématisée des différentes facettes d'un secteur d'activité tel qu'il est évoqué dans les discours du corpus. La formalisation de ces rattachements permet ensuite de documenter l'organisation de l'information sur des sites, en proposant les principaux découpages utilisés relativement à un pan de l'activité du groupe bancaire dans une situation de communication donnée. Une autre propriété de la collecte concerne la possibilité qu'elle offre de suivre l'évolution d'un référentiel sur une période donnée.

D. L'accès aux flux d'information

1. Cartographie de la restitution des secteurs d'activité

a. Présentation générale

Dans les deux sections qui précèdent, on a proposé une évaluation de la collecte réalisée à partir des unités-pilotes, tant pour ce qui concerne les manifestations d'un parler d'entreprise, que pour les manières de découper l'activité de la banque. Pour cela, les relations de catégorisation et de rattachement à un secteur d'activité ont été envisagées pour toute la période couverte par la série textuelle, soit une dizaine d'années. Il convient maintenant d'examiner les apports de la collecte sous l'angle diachronique, ce qui permettra en définitive d'évaluer une procédure recourant à un corpus organisé sous la forme d'une série textuelle chronologique.

Dans une précédente section, on a vu que l'analyse factorielle des correspondances appliquée au corpus C² permet de rassembler par affinités les différents stocks de co-occurents associés aux dénominations propres-pilotes dans les textes du corpus RA. On a ainsi utilisé les résultats de l'AFC afin de sélectionner des stocks de co-occurents homogènes. Le nombre de stocks de co-occurents pris en compte est alors passé de 55 à 47. Il s'agit maintenant de proposer une lecture et une interprétation de ces résultats à la lumière des informations collectées à partir des unités-pilotes. Cependant, il convient de rappeler au préalable les grands principes de l'analyse factorielle des correspondances¹ en rapport avec l'utilisation qui en est faite dans le cadre de ce travail, afin de préciser les fondements sur lesquels reposent les interprétations que l'on va proposer.

L'analyse factorielle des correspondances réalise ses calculs à partir du tableau lexical entier du corpus. Ce tableau est composé de lignes, correspondant à des formes graphiques (suites de caractères séparées par deux espaces et/ou signes

¹ Il ne s'agit pas ici de réaliser une présentation de l'AFC qui requerrait des compétences dépassant celles de l'auteur de ces lignes. Pour une présentation de l'AFC accessible et complète aux non mathématiciens, on renvoie à L. Lebart et A. Salem (1994), ainsi qu'à l'étude pionnière d'A. Prost (1974) réalisée avec la collaboration avec l'Institut de Statistique de l'Université de Paris (J.-P. Benzécri) et le Centre de Lexicologie politique de l'École normale supérieure de Saint-Cloud (R.-L. Wagner et M. Tournier).

de ponctuation dans les textes), et de colonnes, correspondant aux différentes parties du corpus. Chacune des cases de ce tableau contient un nombre correspondant à la fréquence observée pour telle forme dans telle partie du corpus. Ce nombre varie donc de 0, lorsqu'une forme est absente, à la fréquence la plus élevée observée parmi les formes du corpus pour une partie donnée. Dans le cas que l'on étudie ici, les formes proviennent des formes isolées et des formes incluses dans les segments répétés co-occurents d'une unité-pilote et de ses variantes dans le corpus des rapport d'activités (RA)¹. On se souvient que le corpus C² rassemble les stocks de co-occurents de 47 unités-pilotes. Les parties de ce corpus C² correspondent donc chacune à l'un de ces 47 stocks.

Les facteurs sont calculés uniquement à partir des données rassemblées dans le tableau lexical entier ; ils tiennent donc compte simultanément des lignes et des colonnes, mais de rien d'autre, ce qui exclut toute prise en compte de données syntaxiques ou sémantiques externes. Ils permettent ici de faire apparaître des liaisons entre les colonnes du tableau qui seraient imperceptibles autrement. Les facteurs représentent chacun une partie de l'information contenue dans le tableau de données. Le premier facteur correspond à la plus importante quantité d'informations extraites relativement aux autres facteurs. Le second facteur offre une couverture moindre, et ainsi de suite. Théoriquement, le nombre de facteurs correspond au plus petit nombre observé sur les bordures du tableau, soit en lignes, soit en colonnes. Il serait donc ici de 47. En fait, seuls les premiers facteurs sont significatifs et ils s'avèrent suffisants pour produire une analyse. On se limitera ici aux deux premiers facteurs. Il faut ajouter que chaque facteur est orienté, de gauche à droite ou de bas en haut par rapport à une valeur zéro présente en son milieu. La présentation de deux facteurs sous la forme de droites perpendiculaires découpe un plan en quadrants, dont le centre constitue le point d'inertie des facteurs représentés.

Les stocks de co-occurents sur lesquels une AFC a été pratiquée sont donc répartis de gauche à droite, en fonction de leur position sur le premier facteur, et de bas en haut, en fonction de leur position sur le second facteur. La proximité ou

¹ On se souvient que parmi ces co-occurents, seuls ceux qui possèdent un coefficient de spécificité positif supérieur ou égal à 4 ont été retenus.

l'éloignement des stocks de co-occurents représentés par des étiquettes doit donc être comprise relativement aux deux facteurs représentés, même si pour des raisons de commodité, le commentaire des deux axes factoriels peut être réalisé séparément. Ce que les résultats de l'analyse factorielle présentent, ce sont donc des affinités entre stocks de vocabulaire pour une certaine quantité d'informations extraites du tableau lexical entier et représentée par chaque facteur. C'est ce résultat qu'il s'agit d'interpréter maintenant à la lumière des informations collectées à partir des unités-pilotes. Deux lectures peuvent alors être réalisées : l'une statique et l'autre dynamique.

b. Les métasecteurs d'activité

Les métasecteurs fournissent une image globale de l'activité du groupe bancaire telle qu'elle est restituée par les rapports d'activité. La lecture des informations organisées en fonction des deux premiers facteurs peut être réalisée à deux niveaux. Le premier permettra de dégager des métasecteurs qui structurent de façon très globale la restitution de l'activité. Le second fera réapparaître les principaux macrosecteurs d'activité repérés au moment de la collecte. On redonne ci-dessous le résultat de l'analyse factorielle appliquée au corpus C². Cette fois, le graphique a été complété à l'aide de deux sortes d'étiquettes. Les premières correspondent au nom de l'entité, repérée dans le précédent tableau à l'aide d'une simple étiquette numérotée, quant aux secondes, elles consistent en un certain nombre de zones dessinées et numérotées, qui feront ensuite l'objet d'un commentaire.

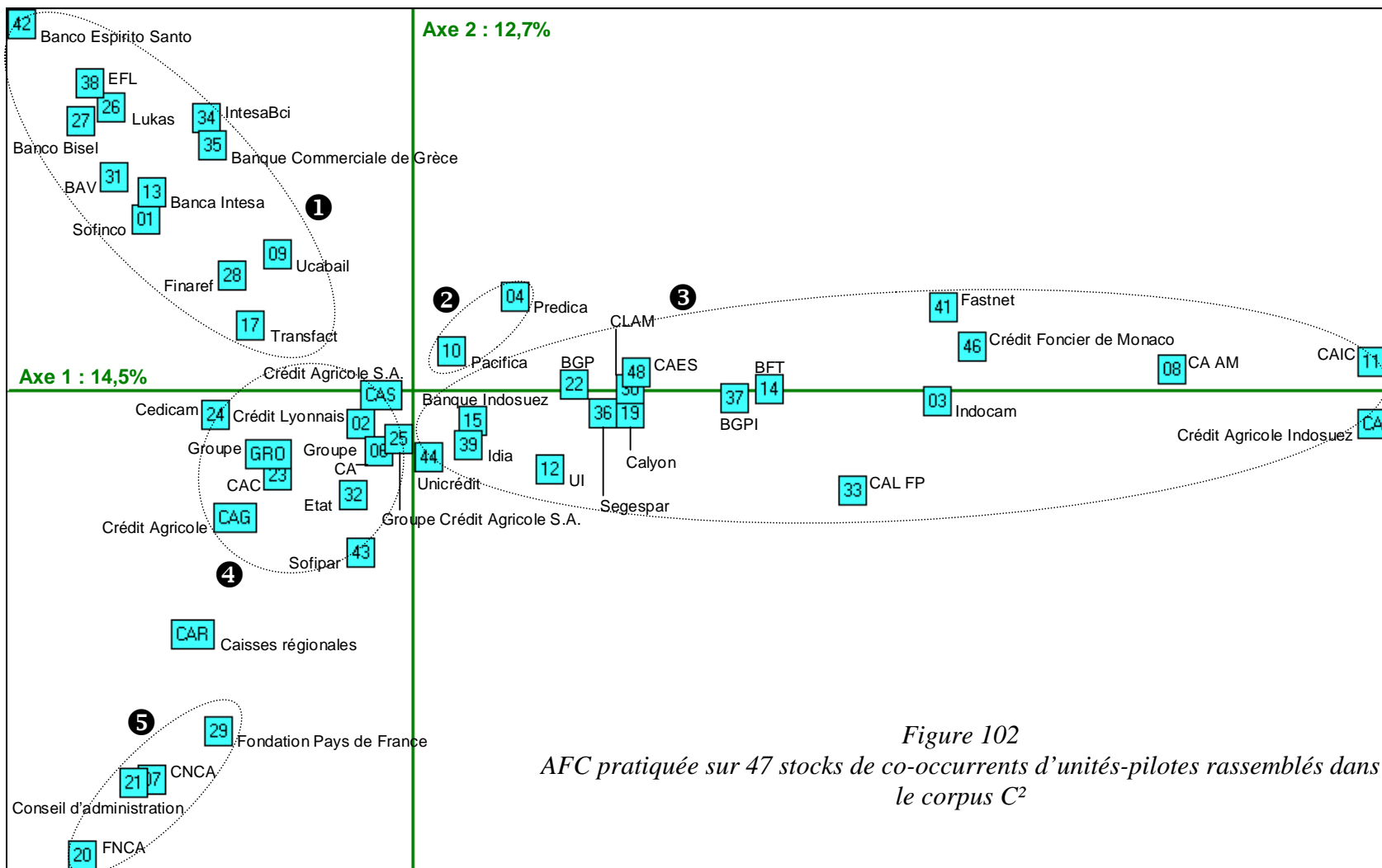


Figure 102
 AFC pratiquée sur 47 stocks de co-occurents d'unités-pilotes rassemblés dans le corpus C²

L'examen de la répartition des stocks de co-occurents le long du premier facteur (axe 1 horizontal), qui rassemble 14,5 % de l'information du tableau lexical, fait apparaître une bipartition assez nette. À la gauche du deuxième axe sont regroupés les stocks d'unités-pilotes dont le référent participe de près ou de loin à l'activité bancaire classique de banque de détail. C'est dans cette partie que l'on retrouve « Crédit Agricole », mais aussi « Crédit Lyonnais ». La partie droite rassemble quant à elle les stocks d'unités appartenant plutôt à la sphère des activités de marché et de la gestion ou du financement d'actifs. En effet, on trouve à l'extrémité de cet axe « Crédit Agricole Indosuez » et « Crédit Agricole Indosuez Cheuvreux », intervenant respectivement dans la banque de financement et d'investissement et dans l'intermédiation boursière. Toutes les entités représentées par leur stock entre ces extrêmes et l'intersection des deux axes, interviennent à des titres divers dans cette sphère d'activité, qu'il s'agisse d'Unicredit, pour les financements structurés, de Predica pour l'assurance-vie, de la BGPI (Banque de Gestion Privée Indosuez) pour la banque privée, ou encore de CA AM (Crédit Agricole Asset Management), pour la gestion d'actifs.

Le deuxième axe (axe 2 vertical, représentant 12,7 % de l'information du tableau lexical) disperse les stocks de co-occurents principalement dans les deux quadrants gauches. Il oppose les stocks proches de l'institution (partie basse), c'est-à-dire les unités pour lesquelles le traitement discursif produit dans les rapports d'activité concerne plus particulièrement la vie de l'institution, à ceux qui s'en éloignent à divers titres (partie haute). Le centre est toujours occupé par « Crédit Agricole » et un certain nombre de stocks apparentés (« Groupe », « Groupe Crédit Agricole », « Crédit Lyonnais », « Crédit Agricole S.A. », etc.). Le quadrant situé en haut à gauche rassemble, au sein de l'activité bancaire traditionnelle, les entités dont certaines caractéristiques les éloignent de celles qui incarnent plutôt la norme. On trouve ainsi les banques étrangères, filiales ou participations, qui exercent le même type d'activités mais dans des pays étrangers (Banco Espirito Santo, banque portugaise, Lukas, banque polonaise ou encore Banca Intesa, banque italienne, etc.). On trouve également rassemblées ici les entités dont les activités correspondent à une spécialisation de l'activité bancaire, comme « Sofinco » et « Finaref » pour le

crédit à la consommation, « Ucabail » pour le crédit-bail, ou encore « Transfact » pour l'affacturage.

Ces observations peuvent être combinées afin de proposer une autre lecture. On peut en effet distinguer un croissant reliant les stocks du « Banco Espirito Santo », situé dans le coin supérieur gauche de la figure, à celui de « Crédit Agricole Indosuez », placé à l'extrémité droite du premier axe, en passant par « Pacifica », proche de l'intersection des deux axes. Ce croissant s'oppose à un ensemble en forme de clou situé dans le quadrant en bas à gauche de la figure, et dont la pointe est constituée par les stocks « FNCA », « Conseil d'administration », « CNCA » et « Fondation Pays de France », et la tête par l'ensemble central rassemblé autour de « Groupe ». Le croissant correspond globalement aux différentes diversifications de l'activité bancaire du Crédit Agricole : banque de détail à l'étranger et services financiers spécialisés (ensemble 1), assurance (ensemble 2) et banque de financement, de marchés et gestion d'actifs (ensemble 3). Cet ensemble rassemblant les entités supportant la diversification de l'activité bancaire du Crédit Agricole semble avoir pour socle les entités représentant la part traditionnelle de cette activité dévolue à la banque de détail (ensemble 4), y compris dans sa dimension institutionnelle (ensemble 5).

On voit que le traitement discursif du stock « Caisses régionales » le place à mi-chemin des ensembles institutionnel et historique. Cela correspond en effet au double statut de ces entités, tel qu'il est intégré à la restitution de l'activité bancaire : avec les caisses locales, elles forment le socle institutionnel du Crédit Agricole, en même temps qu'elles détiennent le capital de son organe central ; mais dans la mesure où ces banques régionales exercent le même type d'activités, la restitution opérée par les rapports d'activité met en avant le Crédit Agricole comme groupe bancaire et non comme addition d'entités distinctes. L'analyse factorielle vient ainsi donner du poids à des regroupements macroscopiques apparus, entre autres, dans les sommaires des rapports d'activité. En effet, leur fonction ne se réduit pas, comme on aurait pu le croire, à une tentative d'ordre cosmétique de la part d'un énonciateur soucieux de rendre ce foisonnement plus lisible à ses lecteurs. Les noms de macrosecteurs viennent en réalité formaliser des mouvements de vocabulaire plus profonds que la comparaison des stocks de co-occurrents a fait ressortir.

La position du stock « Caisses régionales » permet d'insister sur le fait qu'un tel graphique livre une photographie sur laquelle chaque étiquette occupe une position qui est donnée, non pas tant en fonction de l'activité réelle d'une entité, qu'en fonction de la manière dont celle-ci est restituée dans les rapports d'activité. Il est donc fort probable qu'il ne s'agisse pas d'une simple vue statique, mais d'une sorte « d'arrêt sur image », figeant un processus en permanente évolution. Pour s'en convaincre, il suffit d'imaginer le résultat d'une telle analyse appliquée à des stocks issus d'un corpus d'origine traitant de la décennie précédente. Sans aller jusqu'à une telle comparaison, la seule observation du graphique laisse apparaître des scories dont l'analyse va se révéler éclairante.

c. Photographie d'une dynamique centrifuge

Lorsque l'on observe de façon plus détaillée la répartition des différents stocks de co-occurents, certaines positions deviennent problématiques. En effet, comment expliquer les distances séparant « Segespar » d'« Indocam » et ce dernier stock de celui étiqueté « CA AM », dans la mesure où les entités représentées par ces stocks sont catégorisées de la même manière – il s'agit de sociétés de gestion d'actifs et de filiales du Crédit agricole –, et où elles sont rattachées au même secteur de la gestion d'actifs ? La question se pose dans les mêmes termes pour les stocks « Finaref » et « Sofinco », représentant deux entités opérant dans le crédit à la consommation, comme pour « Banque Indosuez », « Calyon » et « CAI » pour la banque de financement et d'investissement. On remarquera enfin que les stocks occupant une position centrale ne sont pas confondus pour autant et que la position véritablement centrale est occupée par Crédit Agricole S.A., alors que l'on aurait pu s'attendre à voir une telle position tenue par le stock « Crédit Agricole ».

On a vu précédemment que les deux axes correspondant aux deux facteurs se coupent en leur point d'inertie. Un groupe de co-occurents qui occuperait cette position – et l'on peut voir que le stock étiqueté « Crédit Agricole S.A. » n'en est pas très éloigné –, serait composé majoritairement de formes communes à tous les autres stocks. Autrement dit, il s'agit d'une position neutre par rapport à la norme présente dans le corpus. Cependant, cette absence d'originalité peut aussi correspondre à une position centrale d'un autre ordre : dans le corpus, tout ce qui est dit des autres entités l'est aussi, d'une manière ou d'une autre, de ou à propos de

« Crédit Agricole S.A. ». Que cette position singulière soit occupée par l'entité qui est également l'émetteur des rapports d'activité vient conforter les observations réalisées dans la première partie. Le changement de statut de l'émetteur opéré en 2001 est doublé par un décentrement du discours produit sur l'ontologie du groupe bancaire. En effet, alors que l'organe central de la banque occupait avant 2001 une position instrumentale de porte-parole au service du Crédit Agricole (on remarque que le stock « CNCA » fait partie de l'ensemble « institutionnel »), à partir de 2001, tout se passe comme si l'activité du groupe bancaire était devenue l'objet d'un discours produit par un émetteur se mettant lui-même en scène comme centre du dispositif. On a vu précédemment que le rôle dévolu à une classe d'entités nommées dans les discours détermine la richesse de la structure de leur contenu, on peut constater à présent que ce rôle détermine également pour une bonne part la position assignée aux stocks de co-occurents dans l'espace factoriel.

On peut alors comprendre les positions occupées par les autres stocks comme autant d'étapes d'un processus centrifuge. Plus le stock possède un vocabulaire qui lui est propre, plus il aura tendance à se tenir éloigné du centre. Cependant, il ne faut pas oublier que cette autonomie se réalise ici dans un discours sur l'activité. On a vu que le groupe a diversifié ses activités de manière progressive en procédant à des acquisitions : Banque Indosuez en 1996, Sofinco en 1999, Finaref et Crédit Lyonnais en 2003. Chacune de ces acquisitions n'a pas entraîné un bouleversement du vocabulaire des rapports d'activité de l'année correspondante. Tout se passe au contraire sur le mode d'une infusion progressive. En effet, l'année de l'acquisition met surtout en avant le point de vue de l'acquéreur et quelques points remarquables de la société nouvellement acquise. Les années suivantes peuvent consacrer l'implantation partielle d'une terminologie propre à la filiale, mais dans des proportions variables en fonction, entre autres, de son importance économique et la pente prise par son développement. Selon ce processus, l'incorporation de la nouvelle entité se traduit d'abord par la proximité de son stock de co-occurents avec celui de l'entité intégratrice. Cependant, les positions occupées par les stocks de co-occurents, regroupés dans les ensembles dessinés sur le précédent graphique, obéissent rarement à un principe explicatif unique. Au contraire, plusieurs sortes d'explications sont souvent nécessaires, comme c'est le cas dans l'ensemble n° 4.

La position surprenante du groupe « Crédit Lyonnais », que l'on pouvait s'attendre à trouver moins proche du centre, tient au phénomène d'intégration qui a été évoqué plus haut : il est surtout question en 2003 des étapes de l'acquisition et des questions soulevées par l'intégration au groupe Crédit Agricole S.A. d'un réseau concurrent de celui du Crédit Agricole. En effet, la seule proximité thématique due à la présentation d'activités similaires dans la banque de détail n'aurait pas suffi à expliquer cette proximité. Les stocks « Groupe CA » et « Groupe Crédit Agricole S.A. » sont également proches du centre car ils partagent avec Crédit Agricole S.A. la propriété d'être les entités au nom desquelles l'activité est présentée. De ce point de vue, « Crédit Agricole S.A. » a très certainement pris la place occupée avant 2001 par « Crédit Agricole » et « Groupe » (forme habituelle et assez spécialisée de reprise de « Crédit Agricole »).

Trois structures, peu visibles pour les marchés, car intégrées de près au fonctionnement du groupe bancaire, « Sofipar », « Crédit Agricole Consultants » et « Cedecam », sont placées par leurs stocks de co-occurents dans le « giron discursif » de l'organe central, comme autant de satellites. Enfin, le stock de co-occurents « État » ne se trouve pas proche du centre « historique » du Crédit Agricole...pour des raisons historiques, même si l'organe central du Crédit Agricole a été jusqu'en 1988 sous la tutelle de l'Etat. Cette proximité s'explique par les multiples rencontres opérées entre celui-ci et la « banque verte » dans le domaine économique telles qu'elles sont restituées dans les rapports d'activité : données macroéconomiques relatives au secteur bancaire, distribution des prêts bonifiés à l'agriculture, privatisation d'entreprises publiques, formation et sauvegarde du patrimoine. Il reste que les positions occupées par certains stocks témoignent de l'existence d'un processus centrifuge dominant.

La projection sur l'axe principal des positions des stocks Sofinco et Finaref, Banque Indosuez, Calyon et Crédit Agricole Indosuez, ou encore, Segespar, Indocam et CA AM va permettre d'éclairer ce point.

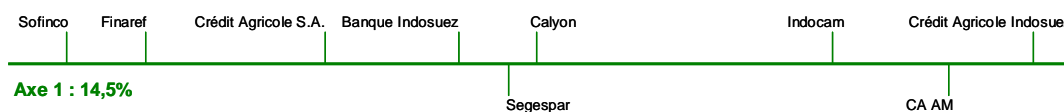


Figure 103.

Projection des positions de 8 stocks de co-occurents sur l'axe 1 de l'AFC réalisée sur les 47 stocks de co-occurents du corpus C²

Comme dans le précédent graphique, « Crédit Agricole S.A. » occupe la position centrale de l'axe factoriel, légèrement décalé sur la gauche. L'application du principe explicatif d'une autonomisation progressive des stocks de co-occurents semble convenir pour Finaref (acquis en 2003) et Sofinco (acquis en 1999), ce dernier stock occupant une position moins centrale que le premier. Il s'applique également à la série « Segespar », « Indocam » et « CA AM », mais d'une façon plus surprenante, car il s'agit en fait de trois avatars de la même entité qui se sont succédés sur la période couverte par le corpus. En comparant les trois stocks, l'AFC les a éloignés du centre en « respectant » la chronologie en fonction de laquelle les trois dénominations se sont succédées, comme le montre graphique de ventilation des occurrences de chaque dénomination.

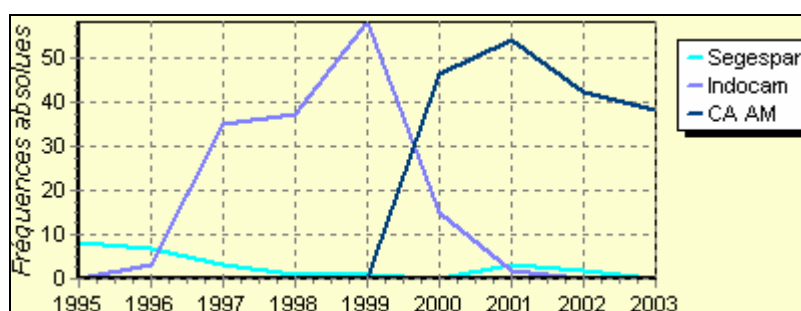


Figure 104.

Ventilation des dénominations « Segespar¹ », « Indocam » et « CA AM » et de leurs variantes sur les neuf parties du corpus principal RA

¹ Les activités de la société « Segespar » ont été fusionnées avec celles d'Indosuez Asset Management en 1997. Le nom et la structure ont néanmoins été conservés jusqu'en 2002. Mais passé 1997 les occurrences de la dénomination ne sont plus liées à une présentation de l'activité de gestion d'actifs. Elles s'expliquent par des mentions dans certains organigrammes ainsi que par des rappels historiques.

Les positions des trois ensembles de co-occurents mettent en évidence le figement du processus centrifuge pour les stocks qui, du fait de la disparition de la dénomination propre qu'ils représentent au cours de la période couverte par la série textuelle, sont comme arrêtés dans leur éloignement du centre du facteur.

Les choses paraissent cependant moins évidentes, lorsque l'on tente d'appliquer la même explication au dernier groupe que l'on a choisi d'étudier (« Banque Indosuez », « Crédit Agricole Indosuez » et « Calyon »). En effet, le groupe « Calyon » devrait être placé après « Crédit Agricole Indosuez », puisqu'il lui est postérieur dans la succession chronologique. La contradiction n'est pourtant qu'apparente, si l'on considère que l'apparition de la dénomination « Calyon » est contemporaine de l'achat du Crédit Lyonnais en 2003. On a observé antérieurement que Calyon résulte de la fusion des activités de banque de financement et d'investissement du Crédit Lyonnais avec celles de Crédit Agricole Indosuez. Par conséquent, ce qui caractérise le plus massivement le stock « Calyon » dans la dernière partie du corpus, c'est la naissance de la nouvelle entité présentée par son acquéreur. On se trouve pour ce stock dans le même schéma d'intégration que pour le stock « Crédit Lyonnais » présenté plus haut.

L'analyse factorielle des correspondances permet ainsi non seulement de positionner chaque stock de co-occurents par rapport aux autres, fournissant ainsi des pistes d'exploration thématique, mais elle propose aussi une cartographie « dynamique » du développement du groupe bancaire, dès lors que l'on recherche des schémas explicatifs capables de rendre compte du résultat produit à partir de la comparaison de stocks de vocabulaire. Elle sert ainsi d'introduction à une compréhension plus profonde de l'évolution d'un vocabulaire qu'une collecte limitée à une approche synchronique aurait pu ignorer.

2. Suivi des manières de découper l'activité, ou l'acculturation des terminologies exogènes

a. Évolution du contenu d'une dénomination propre

L'évolution du vocabulaire concerne aussi bien les unités collectées dans le cadre de la relation de catégorisation que dans celui du rattachement à un secteur d'activité. On a illustré ailleurs¹ ce phénomène à l'aide du contenu de la dénomination propre « UI », également utilisée comme unité-pilote dans le cadre de ce travail. On donne ci-dessous un tableau récapitulatif de l'évolution de la catégorisation et du rattachement à des secteurs d'activité pour cette unité.

Tableau 90.
Variations affectant certains des éléments composant le profil de l'entité « UI »

Extrait du contenu de « UI »	Parties du corpus								
	95 ²	96	97	98	99	00	01	02	03
Dénomination									
UI (F : 82 / f : 82 ³)									
Catégorisation									
banque d affaires (F : 12 / f : 2)									
société d investissement (F : 7 / f : 6)									
(de ses) filiales ⁴ (F : 21 / f : 3)									
structures (F : 31 / f : 2)									
Domaine d'activité									
capital développement (F : 8 / f : 7)									
(en) fonds propres ⁵ (F : 25 / f : 16)									
capital investissement (F : 20 / f : 15)									
Private Equity (F : 10 / f : 4)									
Banque d investissement (F : 11 / f : 2)									

Commentaire du tableau 90 : Les cases en noir indiquent que, pour une partie du corpus donnée, telle unité est associée de façon exclusive à la dénomination-pilote. Le quadrillage indique que l'unité est associée dans plus de cinquante pour cent des cas aux dénominations de l'entité étudiée et apparaît en outre au sein d'une même partie, dans des contextes où ne figure aucune autre dénomination propre d'entité. Les lignes horizontales indiquent que l'unité est associée dans moins de cinquante pour cent des cas aux occurrences de l'unité-pilote étudiée et se trouve associée à au moins une autre dénomination propre de personne morale dans la même partie du corpus. Enfin, une case blanche indique que l'unité est soit

¹ F. Erlos (2006). C'est également le même corpus (RA) qui a été utilisé.

² Parties du corpus correspondant au rapport d'activité d'une année.

³ « F » correspond à la fréquence total de la dénomination et de ses variantes observée dans le corpus, « f » correspond à la fréquence observée pour les paragraphes dans lesquels la dénomination propre ou ses variantes sont présentes.

⁴ Segment entrant dans les expressions « Par l'intermédiaire de ses filiales (...) » ou « Avec l'appui de ses filiales (...) ».

⁵ Ce segment répété entre dans les syntagmes lexicalisés « investissement(s) en f. p. », « financement(s) en f. p. », « apport(s) en f. p. ».

absente, soit sortie du champ du contenu de la dénomination propre. Ce tableau rassemble un extrait du contenu de la dénomination propre dont la version complète figure dans le fichier de collecte ainsi que dans sa restitution au format HTML (prototype de référentiel terminologique).

Les éléments rassemblés dans le tableau ci-dessus permettent d’opposer à la permanence de la dénomination propre les variations affectant la catégorisation et le rattachement à un domaine d’activité, sur une période d’une dizaine d’années seulement. Sans être banal, cet exemple n’est pas pour autant exceptionnel. Le rassemblement de ces informations suggère que le domaine d’activité et la catégorisation partagent une même caractéristique : celle d’indiquer, à travers l’emploi des unités qui leur sont rattachées, la dilution progressive du rôle attribué à l’entité étudiée. En effet, celle-ci, catégorisée comme « banque d’affaires » au début du corpus, est finalement devenue une simple « structure », qui plus est, au sein d’un domaine d’activité qui s’est probablement étoffé de nouvelles entités (le nombre de paragraphes dans lesquels la dénomination de l’entité étudiée apparaît de façon exclusive va diminuant).

Il reste que l’évolution du rattachement aux noms de secteur d’activité semble être la plus riche d’enseignement. En effet, elle donne accès à des indications précises sur la façon dont un domaine d’activité est construit en discours. On décèle des variations de dénominations, révélatrices d’évolutions terminologiques (apparition du calque français d’*investment capital*, puis d’un synonyme anglais *private equity*). On est introduit à des mouvements d’arrière-plan, comme l’élargissement des interventions en fonds propres (passage du capital de développement à celui, plus générique, du capital d’investissement), etc.

Ces indications sont en accord avec les résultats de la collecte, qui ont mis en évidence que la relation de rattachement à un secteur d’activité présente un plus grand intérêt relativement à l’évolution du vocabulaire de l’organisation de la restitution de l’activité bancaire. En effet, comme on l’a observé précédemment, la catégorisation subit une double contrainte : pour les objets, elle n’est pas systématique et elle n’est observable que pour des périodes courtes d’une à deux années ; en ce qui concerne les personnes morales, elle est cantonnée à deux fonctions principales, ce qui limite de fait la variété des unités collectées ainsi que leurs éventuelles variantes. On ne trouve pas ces limitations dès lors que l’on

s'intéresse à l'étude du rattachement des personnes morales à des secteurs d'activité. La productivité de la collecte pour cette relation et cette classe de dénominations est en effet de 5 en moyenne, ce qui fournit un volume conséquent de données.

Dans les sections précédentes, l'utilisation des unités-pilotes ayant pour référent une personne morale a mis en évidence l'existence de rattachements multiples possibles pour une même entité. Dans le prolongement de l'analyse qui a été donnée plus haut en ce qui concerne l'évolution du vocabulaire, on voudrait montrer ici que ce mouvement, décelable à l'échelle des stocks, est également perceptible dans les relations de rattachement captées à partir des unités-pilotes. On sera alors introduit à l'une des principales sources d'évolution du parler d'entreprise tel qu'il se manifeste dans les rapports d'activité. Or, le groupe bancaire a développé ses activités principalement en acquérant des sociétés externes. On sait par ailleurs que les rapports d'activité ont pour finalité, entre autres, de construire une image de l'institution qui soit à la fois rassurante pour les investisseurs et dans laquelle les composantes du groupe puissent se retrouver. Dans le contexte des nombreuses acquisitions réalisées en une dizaine d'années, cette double finalité ne peut qu'être source de tensions. Comment concilier, en effet, l'image d'un développement cohérent avec la prise en compte des particularités des sociétés nouvellement acquises ? Comme on l'a indiqué antérieurement, on limitera ici l'étude de cette question à ses éventuelles répercussions sur la manière de découper l'activité ou de nommer différemment des découpages préexistants.

Lorsqu'elle a pour source une série textuelle chronologique ouverte, la collecte est normalement enrichie à chaque fois qu'une nouvelle partie vient mettre à jour le corpus de suivi. Il s'agit donc de vérifier maintenant si l'enrichissement des relations de certaines unités-pilotes réalisé à partir de la relation de rattachement à un secteur d'activité se fait concomitamment à l'apparition de nouvelles entités. Comme on va le voir, que la rencontre des deux phénomènes trouve au moins trois sortes d'issues attestées qui participent toutes à l'évolution du vocabulaire du corpus, du moins en ce qui concerne les noms de secteurs d'activité. Afin d'illustrer ce point, on s'appuiera sur les dénominations propres de personnes morales repérées à l'aide des spécificités évolutives de coefficient élevé dans la dernière partie du corpus (2003), ce coefficient constituant un bon indice de nouveauté.

Comme cela a été montré dans la première partie, les dénominations propres de personnes morales concernées sont : « Crédit Lyonnais », « Calyon », « CLAM », « CA AM », « Finaref », « UAF », « Eurofactor » et « Lixxbail ». Les trois dernières dénominations ne font pas partie des unités-pilotes utilisées plus haut, dans la mesure où leur fréquence est inférieure à vingt occurrences. Elles ont néanmoins été intégrées à cette dernière analyse, compte tenu de leur présence caractéristique pour l'année 2003. En revanche, les commentaires relatifs au « Crédit Lyonnais » risquant d'être redondants avec ceux qui ont déjà été produits plus haut à propos du groupe de co-occurents, cette entité ne sera pas utilisée pour les développements qui suivent.

b. Phénomène de substitution

Le phénomène de substitution consiste dans le remplacement d'une dénomination de secteur d'activité par une autre, opérée brutalement d'une année sur l'autre. Le rapprochement des activités de banque de financement et d'investissement du Crédit Lyonnais avec celles de Crédit Agricole Indosuez en donne une bonne illustration. En effet, le remplacement d'un nom de secteur d'activité par un autre se trouve dans ce cas associé au changement de dénomination de l'entité résultant de la fusion des activités des deux banques. Les deux bouleversements se font ainsi écho. Depuis 2001, la « banque de grande clientèle » constituait le nom de secteur attribué de Crédit Agricole Indosuez. Cette dénomination de secteur n'était cependant pas la seule à faire partie de la structure de contenu de cette banque. On y trouve, entre autres, « banque de d'investissement ». Comme cela est représenté dans le graphique ci-dessous, la naissance de « Calyon » va entraîner la disparition d'un nom de secteur installé au profit d'une nouveauté.

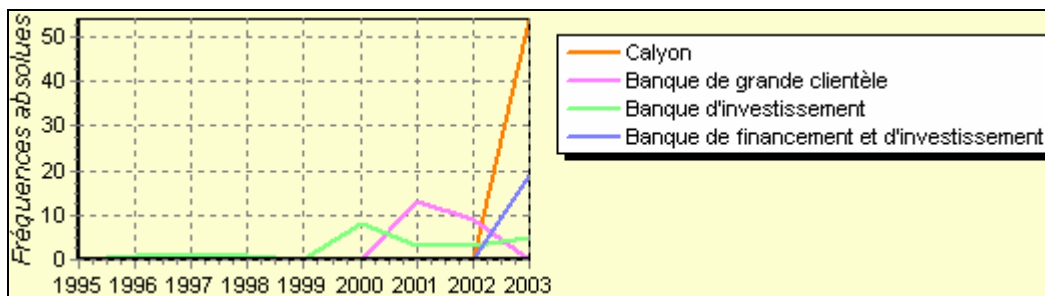


Figure 105.
Ventilation sur les neuf parties du corpus RA de « Calyon » et de trois noms de secteurs d'activité

Les ventilations des trois groupes de formes correspondant aux noms de secteurs d'activité « Banque de grande clientèle¹ », « Banque d'investissement² » et « Banque de financement et d'investissement³ » montrent que le premier de la série disparaît au moment où l'usage du dernier se manifeste, en 2003. Cette concomitance est liée directement à la naissance de la nouvelle entité représentative de ces secteurs d'activité au sein du groupe Crédit Agricole. Sur le même graphique on peut également distinguer que les occurrences du nom « Banque d'investissement » n'étaient pas pour autant tout à fait absentes des précédentes parties du corpus. En effet, ces surgissements terminologiques apparaissent généralement sur un terreau favorable. Il reste que les dénominations concurrentes peuvent cohabiter au-delà de l'événement fondateur. C'est ce que l'on peut caractériser comme une forme de complémentation.

c. Phénomène de complémentation

La complémentation se caractérise par la coexistence de plusieurs dénominations de secteurs d'activité successivement attachées à la même unité-pilote. L'élément déclencheur est encore la restitution de l'intégration d'une nouvelle entité au sein du groupe bancaire. Cependant, la réaction n'est pas aussi

¹Le groupe est constitué des segments répétés « Banque de grande clientèle » (19 occ.) et « banque de grande clientèle » (3 occ.).

²Le groupe est constitué des segments répétés « Banque d'investissement » (11 occ.) et « banque d'investissement » (11 occ.).

³Le groupe est constitué des segments répétés « Banque de financement et d'investissement » (9 occ.) et « banque de financement et d'investissement » (10 occ.).

tranchée que dans le cas de figure précédent. En effet, l'arrivée de la nouvelle entité entraîne un usage accentué d'une dénomination déjà présente avant l'événement. Ce point est illustré à l'aide de deux exemples appartenant pour l'un à la gestion d'actifs, et pour l'autre, au domaine de l'assurance-vie. Ces deux secteurs sont, avant 2003, associés aux activités respectives de Crédit Agricole Asset Management et de Predica. L'intégration de nouvelles filiales appartenant au groupe Crédit Lyonnais sert à nouveau d'élément déclencheur.

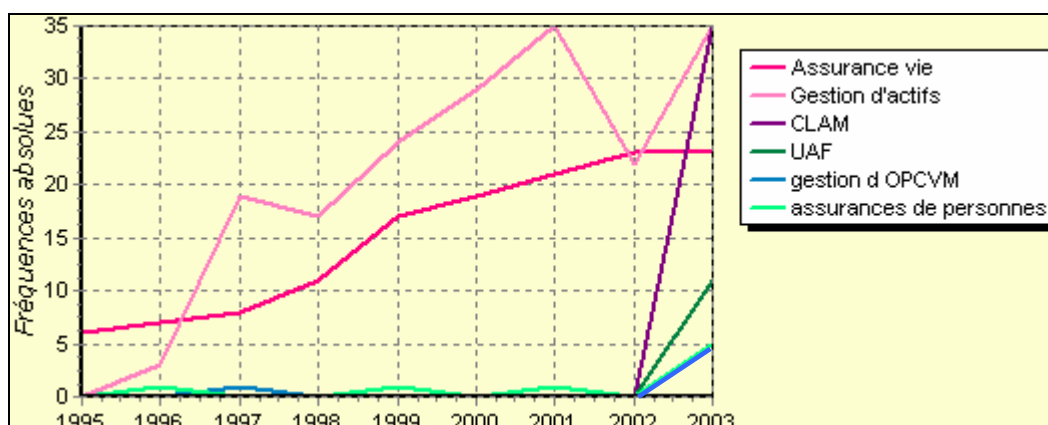


Figure 106.

Ventilation sur les neuf parties du corpus RA de « CLAM » et de « UAF » ainsi que de quatre dénominations de secteurs d'activité

L'apparition, en 2003, de deux nouvelles entités opérant dans les secteurs de l'assurance-vie et de la gestion d'actifs entraîne parallèlement la mise en avant de deux autres noms de secteurs moins visibles auparavant. En effet, ces deux noms de secteurs¹ sont utilisés non seulement pour le rattachement des nouvelles entités mais également pour celui des entités mentionnées avant 2003. Réciproquement, les noms de secteurs dominant avant 2003² viennent compléter la structure de contenu des nouvelles dénominations propres « UAF » et « CLAM ». On a ici une nette manifestation de l'un des processus responsables de l'enrichissement des structures

¹ « Gestion d'OPCVM » est un segment répété de fréquence 6, et « assurances de personnes » correspond au groupe de forme de même nom rassemblant les segments répétés « assurances de personnes » (3 occ.) ; « assurance de personnes » (3 occ.) et « Assurances de personnes » (2 occ.).

² « Assurance vie » correspond aux occurrences du groupe de formes rassemblant les segments répétés « assurance vie » (114 occ.) et « Assurance vie » (21 occ.). « Gestion d'actifs » correspond aux occurrences des segments répétés « gestion d'actifs » (117 occ.), « Gestion d'actifs » (59 occ.) et « GESTION D'ACTIFS » (8 occ.) rassemblés dans le même groupe de formes.

de contenu, au moins pour ce qui concerne les dénominations de secteurs d'activité. Un troisième phénomène doit enfin être mentionné. Il concerne plus particulièrement l'apparition de métasecteurs destinés à donner une certaine lisibilité aux évolutions du groupe bancaire.

d. Phénomène de subsomption

Le phénomène se caractérise non pas par une simple juxtaposition, comme cela a été montré dans la précédente section, mais par l'apparition d'un nom de secteur d'activité englobant. Dans ce cas, les noms de secteurs d'activité employés au préalable, afin d'assurer la lisibilité de l'activité des entités du groupe bancaire, ne suffisent plus. On sait que cette volonté de clarification attribuée à l'énonciateur est très certainement mêlée, dans le contexte des rapports d'activité, au souci de valoriser certains aspects de cette activité. Il n'en reste pas moins que ce type de création semble apparaître lorsqu'un certain seuil a été atteint, soit en termes de visibilité économique, soit en termes d'« encombrement ».

Afin d'illustrer ce dernier point, on utilisera les trois dénominations « Finaref », « Eurofactor » et « Lixxbail », dont les fréquences les plus élevées se situent dans la dernière partie du corpus. Ces trois entités interviennent respectivement dans les secteurs du crédit à la consommation, de l'affacturage et du crédit-bail, où les ont précédé dans les rapports d'activité les sociétés Sofinco, Transfact et Ucabail. On présente dans le graphique ci-dessous les ventilations de l'ensemble de ces groupes de formes¹.

¹ Les groupes de formes sont présentés dans leur ordre d'apparition dans la légende du graphique : « Crédit à la consommation » est composé des segments répétés « crédit à la consommation » (128 occ.) et « Crédit à la consommation » (15 occ.) ; « Sofinco » est composé des formes « Sofinco » (165 occ.) et « SOFINCO » (9 occ.) ; « Affacturage » est composé des formes « affacturage » (59 occ.), « Affacturage » (12 occ.) et « AFFACTURAGE » (1 occ.) ; « Transfact » est composé des formes « Transfact » (60 occ.) et « TRANSFACT » (7 occ.) ; « Crédit bail » est composé de la forme « BAIL » (1 occ.) et des segments « crédit bail » (80 occ.) et « Crédit bail » (19 occ.) ; « Ucabail » est composé des formes « Ucabail » (92 occ.) et « UCABAIL » (9 occ.) ; « Finaref » est composé des formes « Finaref » (30 occ.) et « FINAREF » (1 occ.) ; « Eurofactor » est composé des formes « Eurofactor » (15 occ.) et « EUROFACTOR » (1 occ.) ; « Lixxbail » est composé des formes « Lixxbail » (9 occ.) et « LIXXBAIL » (1 occ.) ; enfin, « Services financiers spécialisés » est composé des segments répétés « Services financiers spécialisés » (8 occ.) et « services financiers spécialisés » (7 occ.).

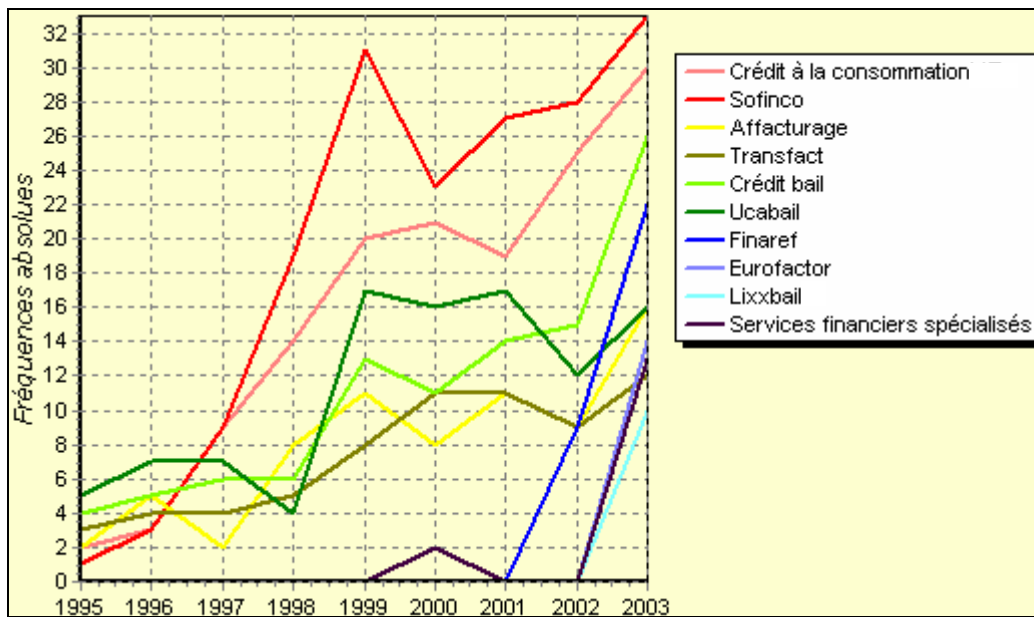


Figure 107.

Ventilation sur les neuf parties du corpus RA des noms de neuf sociétés ainsi que des noms des principaux secteurs d'activité auxquels elles sont associées

Le graphique met en évidence les variations parallèles des fréquences des dénominations propres de sociétés avec celles des noms de leurs secteurs d'activité de rattachement. Les trois noms d'entités dont les fréquences se manifestent principalement dans la dernière partie (« Finaref », « Eurofactor » et « Lixxbail ») sont associés aux noms de secteurs préexistants. Mais il semble que l'apparition au même moment du nom du macrosecteur « services financiers spécialisés » soit déclenché, ou rendu possible, par le renforcement du poids économique des secteurs existants. À nouveau, l'apparition de nouveaux noms de personnes morales permet de suivre finement l'évolution du découpage de l'activité. Si ce phénomène ne permet naturellement pas d'expliquer à lui seul les évolutions du vocabulaire du corpus, il vient néanmoins illustrer le fait plus général selon lequel, dans le cadre d'un parler d'entreprise, ces évolutions paraissent moins déterminées par une évolution terminologique interne que par des causes externes. On pourrait probablement avec un corpus plus étendu trouver dans les années 1970 l'amorce d'une première transformation du Crédit agricole qui s'est achevée au début des années 1990 : le passage d'une banque agricole à une banque universelle motivant une restitution de l'activité en fonction des principales clientèles de la banque de détail – particuliers, dont les jeunes, entreprises, professionnels, dont les

agriculteurs, collectivités publiques. La période retenue pour le corpus correspond à une nouvelle étape de l'histoire du groupe bancaire : la diversification et l'internationalisation de ses activités. Il semble que l'approche par les unités-pilotes permette de suivre ces évolutions de manière assez détaillée pour ce qui est de leur traduction dans le vocabulaire des rapports d'activité. En cela, le référentiel terminologique semble bien être en prise avec l'évolution du référentiel de l'institution.

La présentation et l'évaluation de la collecte réalisée à partir des unités-pilotes constituent une mise en application des points abordés dans les précédents chapitres. Parmi les dénominations propres préalablement recensées, certaines ont été sélectionnées, d'une part, en fonction des propriétés de la structure-type de contenu de la classe à laquelle elles se rattachent, et d'autre part, en fonction de la fréquence de leurs occurrences dans le corpus RA.

Toutes les propriétés révélées lors de l'établissement des structures-types de contenu n'ont pas été retenues pour la réalisation de cette collecte. Le choix de la catégorisation des référents nommés à l'aide de dénominations propres, ainsi que celui de leur rattachement à des secteurs d'activité ou à des types de clientèles ont été orientés par l'utilisation que l'on compte faire des données collectées en matière d'organisation de l'information.

Le contenu des dénominations propres étant construit au fil des discours, il a été nécessaire de décrire la manière dont les deux relations étudiées sont matérialisées dans les textes du corpus. C'est à partir de ces observations qu'il a ensuite été possible d'établir des critères capables de guider de manière opérationnelle la réalisation de la collecte. Ces critères consistent donc en la caractérisation des moyens syntaxiques et graphiques utilisés dans les textes afin de réaliser des catégorisations des référents des dénominations propres pilotes, ou de rattacher ceux-ci à des secteurs d'activité ou des formes de clientèles. Ces critères ont permis de fiabiliser le repérage d'unités candidates. Toutes n'ont cependant pas été intégrées au prototype de référentiel terminologique créé pour l'occasion.

En effet, l'intégration des unités candidates a ensuite été évaluée en fonction des types d'informations tels qu'ils sont représentés dans le corpus pour chaque classe d'unités-pilotes. Pour les dénominations propres d'objets, qui ont des fréquences faibles à moyennes, les unités catégorisantes ont été retenues systématiquement, alors que pour les dénominations propres de personnes morales, dotées de fréquences moyennes à élevées, les catégorisations hapax n'ont pas été retenues. De même, le rattachement aux secteurs d'activité étant principalement une caractéristique des dénominations propres de personnes morales, les rattachements de dénominations d'objets n'ont été enregistrés que dans le cas où ils étaient déjà

identifiés pour au moins une personne morale. De plus, cette information étant susceptible d'influencer la manière d'organiser l'information sur les sites d'un intranet, on a veillé à ce que sa présence soit bien enracinée dans les textes du corpus. Pour cette raison, un nom de secteur d'activité n'a été intégré que dans la mesure où il possédait une fréquence égale ou supérieure à trois et une ventilation sur au moins deux parties du corpus (qui en compte neuf).

L'intégration dans un référentiel terminologique soulève, par ailleurs, une question d'un autre ordre, dès lors qu'il s'agit de dénominations propres. En effet, celles-ci ne bénéficiant pas de définition en langue, mais seulement d'un contenu construit dans et par les discours, il a été nécessaire de justifier l'utilisation de relations formelles destinées en terminographie à relier entre eux des concepts. On a alors pu déterminer que la transposition de la catégorisation observée dans les discours peut être réalisée dans un tel référentiel sous la forme d'une relation d'instance. En revanche, les rattachements à un secteur d'activité ou à une forme de clientèle ne semblent pouvoir être traduits que par des relations associatives. Celles-ci sont rendues possibles grâce à l'établissement de la précédente relation d'appartenance, elle-même contrôlée par les catégorisations enregistrées pour le référent des dénominations propres. Dans tous les cas, la spécificité des dénominations propres entraîne deux types de restrictions pour l'utilisation d'un référentiel terminologique : il doit s'agir d'unités terminales (sans niveaux subordonnés) ; par ailleurs, l'héritage de la relation d'instance doit être bloqué au niveau immédiatement supérieur à la dénomination propre (la question de l'héritage ne se pose pas pour les relations associatives qui ont généralement une valeur informative).

La collecte a été établie à partir de cinquante-deux dénominations d'objets et de cinquante-cinq dénominations de personnes morales. Elle a permis de collecter, à partir des seules relations de catégorisation et de rattachement, plus de huit cents unités, près d'un millier si l'on compte également les dénominations-pilotes. L'enregistrement des données a confirmé le rôle prédominant des dénominations propres de personnes morales qui, comme cela a été montré lors de la détermination

des structures types de contenu, établissent les connexions les plus nombreuses et les plus fournies avec les autres unités du vocabulaire du corpus.

Ces données ont été rassemblées dans les feuilles d'un tableur de manière à faciliter leur saisie mais aussi leur contrôle. Cela est rendu d'autant plus nécessaire que les données sont volumineuses et qu'un codage numérique a été utilisé afin de distinguer les différentes relations identifiées (équivalence entre dénominations, relation d'appartenance à une classe - traduite par " X est un Y" où X représente un objet et Y un concept -, relations associatives exprimant le rattachement d'une entité nommée à une clientèle ou un secteur d'activité). Par ailleurs, l'utilisation d'un tableur (ou d'une base de données) permet de réaliser des exports et des imports de données dans des formats standard de type « csv » (comma separated values) ou "texte tabulé". En l'occurrence, cela a rendu possible une restitution HTML d'un prototype de référentiel publiable sur intranet et donc consultable par les webmestres désireux de l'utiliser.

L'évaluation des résultats obtenus doit prendre en considération deux aspects principaux. Il s'agit tout d'abord de la méthodologie textométrique utilisée et, partant, du logiciel Lexico. En second lieu, il convient d'évaluer la portée des résultats obtenus au regard des exigences formulées au début de ce travail.

En ce qui concerne le premier point, l'approche textométrique apporte une robustesse et une souplesse qui permettent d'envisager des situations exploratoires diversifiées. La collecte peut porter aussi bien sur des unités de faible que de forte fréquence, les segments répétés permettent l'identification et la délimitation des syntagmes lexicalisés candidats sans analyse morphosyntaxique ; enfin, l'analyse factorielle des correspondances fournit une restitution des mouvements de vocabulaire pertinente pour l'étude du contenu des dénominations propres. Cette méthodologie permet en outre d'asseoir les jugements sur des procédures et une métrique reproductibles autant de fois que nécessaire, ce qui constitue une garantie pour la fiabilité des résultats obtenus.

D'autre part, les résultats doivent être évalués par rapport aux attentes formulées dans le premier chapitre, en matière d'organisation de l'information. Les limites générales de l'approche seront abordées dans la conclusion générale. On se contentera ici de répondre dans le cadre du seul corpus utilisé et sur les points

concernant le fait de capter un parler d'entreprise et de saisir les évolutions d'un référentiel sous une forme exploitable en termes d'organisation de l'information.

L'étude de la relation de catégorisation a mis en évidence que les dénominations propres-pilotes sont insérées dans des paradigmes désignationnels dont l'établissement fournit le moyen de capter des unités appartenant à différentes strates d'un parler d'entreprise. On dispose ainsi d'un bon observatoire de l'utilisation d'un tel parler dans une situation de communication type. Les dénominations propres et les unités captées peuvent ensuite servir de références pour conduire de nouvelles explorations destinées à illustrer des situations dans lesquelles d'autres strates d'un même parler sont mobilisées. Par ailleurs, la mise au jour des unités utilisées pour découper la restitution de l'activité a permis de collecter des indications tout aussi utiles en matière d'organisation de l'information. Outre l'identification de l'utilisation de strates différentes d'un même parler d'entreprise, la collecte fournit ici, pour une même entité, des manières différentes de découper la restitution de l'activité, mais aussi des découpages réalisés à des niveaux de précision variables dans le même type de contexte. Enfin, l'observation de variations diachroniques, rendue possible par l'utilisation d'un corpus organisé sous la forme d'une série textuelle chronologique, a permis de mettre en évidence des évolutions dans la catégorisation et le rattachement aux secteurs d'activité. De ce point de vue, l'organisation du corpus a permis d'identifier une cause importante de variation que restitue une collecte réalisée à partir de dénominations propres : l'introduction des terminologies propres aux entités acquises par un groupe bancaire se développant par achats successifs. De ce point de vue, même si cette approche est limitée sur certains aspects, la collecte qui a été réalisée atteste sa productivité et sa capacité à saisir un référentiel en mouvement.

CONCLUSION GÉNÉRALE

La mise en place de référentiels terminologiques adaptables au contexte est rendue nécessaire, d'une part, pour organiser la diffusion de l'information sur les intranets et, d'autre part, parce que les solutions existantes qui auraient pu être utilisées se sont révélées inadaptées. Or, la recherche de ces outils suppose de rompre avec certaines conceptions dominantes concernant l'usage de la langue et l'organisation de l'information. Les difficultés ne sont pas tant d'ordre technique que communicationnel. On a pu montrer dans ce travail que leur résolution nécessite de passer par la prise en compte des usages linguistiques des publics d'un site, de manière à documenter les tâches liées à l'organisation de l'information.

L'approche que l'on a retenue consiste à utiliser les traces discursives laissées par une situation de communication réelle afin de documenter une autre situation de communication, liée au partage et à la diffusion de l'information. On ne cherche donc pas à atteindre d'emblée un certain degré de généralité. Au contraire, on a considéré que les propriétés du contexte d'utilisation des référentiels terminologiques doivent être prises explicitement comme points d'appui. Parmi les principales, on en a retenu trois.

La première propriété correspond au postulat de l'existence d'un référentiel d'entreprise diversement partagé par les interlocuteurs au sein d'une organisation. Ce référentiel correspond à l'ensemble des objets (concrets ou abstraits, réels ou imaginaires) dont un locuteur peut parler dans une langue donnée et qui ont un rapport direct ou indirect avec l'exercice de ses activités dans une entreprise. Un tel référentiel permet de limiter le champ des investigations et de le centrer sur les évocations d'une ontologie particulière. À titre d'illustration, on a pris pour contexte organisationnel un groupe bancaire, le Crédit agricole, et plus précisément les sites intranets de son organe central et de sa maison mère, Crédit agricole s.a.

La deuxième réside dans les caractéristiques pragmatiques des noms propres désignant des objets appartenant spécifiquement à un tel référentiel. Celles-ci ont

pour socle le lien dénominatif stable établi entre un référent, qu'un groupe de locuteurs a souhaité distinguer, et le nom spécifique qui lui a été attribué : au moins une partie des objets qui occupent une place importante dans un référentiel sont ainsi dénommés. Par ailleurs, l'usage d'un nom propre suppose que les interlocuteurs partagent la connaissance du référent désigné. En retour, l'utilisation d'un nom propre dans diverses situations de communication contribue à définir un groupe de locuteurs. Enfin, la mise en exergue de certains objets à l'aide de ce type de dénomination est repérable dans les textes grâce à l'utilisation de la majuscule, au moins dans les domaines français et anglais concernés par la constitution d'un tel référentiel terminologique dans le contexte qui a été retenu. Leur identification est donc aisée : elle facilite le contrôle du partage effectif de ces unités entre les contenus publiés sur un site intranet et les textes utilisés comme sources pour les référentiels terminologiques. Ces unités ont donc été utilisées pour guider la collecte nécessaire à la construction des référentiels terminologiques.

La troisième propriété contextuelle à laquelle on a eu recours consiste en l'utilisation de ressources discursives produites dans le même cadre que les activités d'information qu'il s'agit de documenter. Les textes sont choisis en fonction des usages linguistiques que le référentiel terminologique doit restituer. Dans cette étude, on a retenu la situation dans laquelle un site quelconque de l'intranet de Crédit agricole s.a. est visité par un public de néophytes. Parmi les productions discursives locales, on a sélectionné les rapports d'activité qui servent à présenter au grand public les divers aspects de l'activité d'une entreprise.

L'adéquation entre les propriétés du contexte de production des rapports d'activité et celles de la consultation de sites intranets par des néophytes a fait l'objet d'une évaluation du point de vue des finalités, des locuteurs, du propos ainsi que des lieux et moments légitimes de la communication. Le résultat obtenu fait apparaître un ensemble de points communs qui justifient l'adoption d'une telle source. Néanmoins, certaines limites ont aussi été mises en évidence. Cette explicitation est nécessaire, dans la mesure où une telle démarche ne peut procéder que par valeurs approchées. Quoi qu'il en soit, on évite ainsi un autre écueil, qui consiste à justifier la constitution de ressources textuelles en fonction de leur

appartenance à la "littérature du domaine", celle-ci étant censée être adéquate quels que soient le lieu et le moment de l'énonciation. Dans la mesure où l'on a retenu l'année 2004 comme point de référence pour la construction du référentiel terminologique, on a constitué un corpus de rapports d'activité du Crédit agricole couvrant la période 1995-2003, de manière à embrasser un pan significatif du référentiel de cette banque. Ce corpus de suivi a vocation à être augmenté par l'ajout de nouvelles parties.

Cette série textuelle chronologique a ensuite été convertie dans un format électronique exploitable par des programmes informatiques. Les logiciels textométriques, tels que Lexico 3, permettent de restituer les dispositifs textuels originaux à l'aide de balises. On a ainsi pu utiliser différentes sortes de partition du corpus : les parties correspondant aux rapports d'activité, l'organisation des textes en rubriques, la répartition des informations entre contextes à dominante syntactique et non syntactique, ou encore, l'agencement original des paragraphes.

Cependant, l'observation des usages linguistiques sur une période plus ou moins étendue ainsi que la mise à jour du corpus ont conduit à prévoir une évaluation de ce type de ressource textuelle. Un tel contrôle précède, et doit normalement éclairer, les analyses qui portent ensuite sur les aspects thématiques. On a proposé un test textométrique destiné à suivre les évolutions des deux principaux dispositifs textuels présents dans le corpus (contextes à dominante syntactique et non syntactique). Ce test consiste à identifier le vocabulaire typique de ces contextes. Le vocabulaire typique d'un contexte est obtenu en sélectionnant, parmi les formes spécifiques ayant le coefficient positif le plus élevé, celles qui sont presque exclusives de l'un ou l'autre contexte. Ces formes sont projetées sur la carte des sections (paragraphes) du logiciel Lexico. On identifie alors que si les rubriques sont composées presque exclusivement des occurrences typiques de l'un ou l'autre des contextes en début de période, les dernières parties du corpus sont caractérisées par une répartition moins tranchée. On met alors en évidence l'apparition d'un nouveau contexte mixte à partir de 2001.

D'autre part, le contrôle de l'évolution du vocabulaire dans une série textuelle chronologique peut être rapporté à un modèle théorique dans lequel le vocabulaire se

renouvelle régulièrement d'une partie du corpus à l'autre. Si ce renouvellement est irrégulier, le résultat de l'analyse factorielle des correspondances déforme la répartition des parties du corpus correspondant au modèle théorique (il s'agit d'une hyperbole). On constate pour le corpus de rapports d'activité que l'année 2001 introduit une rupture avec l'ensemble homogène correspondant à la période 1995-2000. Ce phénomène est confirmé dans des proportions différentes pour les années 2002 et 2003.

On a donc cherché à expliquer cette évolution du vocabulaire. En effet, l'apparition d'une nouvelle sorte de contexte, présentant les propriétés cumulées de ceux qui avaient été identifiés précédemment, a pu être mise en relation avec l'adaptation du dispositif textuel à un nouveau public à partir de 2001. On dispose d'indications montrant que ce changement a tendance à s'infléchir dans la dernière partie du corpus. Ce faisant, on a identifié un point de contrôle qui devra être effectué lors de l'intégration de nouvelles parties. Du point de vue du vocabulaire, on a pu établir un parallèle entre les changements constatés à partir de cette même année 2001 et la recomposition du groupe bancaire, qui a été accompagnée d'une modification du statut de l'émetteur des rapports d'activité. Dans les deux cas, les noms propres apparaissent comme des formes spécifiques des périodes où interviennent ces transformations. Ces indications confirment la capacité de ces unités à donner accès aux transformations affectant un référentiel d'objets.

Ce type de contrôle a vocation à être recommencé à chaque ajout d'une nouvelle partie. Il s'agit toutefois d'un moyen rapide et fiable de détecter à un niveau encore macroscopique des changements intervenus dans le traitement de l'information pour le corpus utilisé comme source. On dispose, en plus d'une forme de contrôle, d'indications sur des pistes d'investigation lorsque l'étude du vocabulaire est approfondie.

On a indiqué précédemment que la construction de référentiels terminologiques adaptables au contexte s'appuie sur la collecte des noms propres. Toutefois, l'usage d'un parler d'entreprise ne peut pas être réduit aux seuls noms propres utilisés dans un corpus de textes. Un référentiel d'objets est verbalisé à l'aide d'autres unités qu'il est nécessaire de collecter pour la constitution d'un tel référentiel

terminologique. Par conséquent, on a proposé d'utiliser les noms propres comme unités-pilotes, ou pivots, afin de capter une partie du vocabulaire apparaissant dans leur voisinage textuel. On a alors estimé nécessaire de disposer d'indications afin d'évaluer si une telle collecte permet effectivement de sélectionner des unités caractéristiques d'un vocabulaire. C'est pourquoi, on a cherché à caractériser le vocabulaire du corpus comme expression de l'usage d'un parler d'entreprise. Celui-ci correspondant à un ensemble hétérogène empruntant aussi bien à la langue commune qu'au jargon propre à une entreprise, il a fallu identifier les usages linguistiques les mieux représentés dans le vocabulaire du corpus.

On a d'abord entrepris de repérer l'usage qui est fait de la langue commune en isolant des unités caractéristiques. Le moyen que l'on a employé consiste à comparer les fréquences observées pour des stocks de vocabulaire aux propriétés différentes. La base de données *Frantext* a fourni les fréquences observées pour les mots apparaissant dans les textes appartenant à divers genres littéraires, collectés pour la période allant de 1950 à 1990 (cette comparaison est proposée par le logiciel Hyperbase). On dispose avec ces indications d'une approximation de la langue commune littéraire pour la période considérée. La comparaison avec le corpus de rapports d'activité n'a pas été faite dans le but de mettre au jour, par contraste, une thématique bancaire assez prévisible : on a utilisé cette comparaison afin d'identifier les types de mots plus particulièrement utilisés pour évoquer l'activité bancaire dans le corpus. De ce fait, les termes trop spécifiques, absents ou rares dans le corpus *Frantext* ont été éliminés. Seuls les mots communs aux deux corpus ont été comparés.

Les cinquante premiers résultats de la comparaison ont ensuite été contrôlés de manière à écarter les nombreuses scories (60%) engendrées par une comparaison mécanique. Il s'agit d'homonymes ou de mots employés comme termes au quasi-termes dans le corpus de rapports d'activité, ou encore de formes simples dont l'usage dominant dans le corpus fait qu'elles sont presque toujours incorporées à des dénominations complexes sur lesquelles la comparaison n'a pas pu porter (seules les formes simples sont comparées).

À l'issue de ce traitement, il a été possible d'établir pour une vingtaine de mots que leur sélection reposait véritablement sur une surreprésentation de mots

relevant de la langue commune dans le corpus de rapports d'activité. Ces mots permettent d'évoquer les principaux composants de l'activité bancaire (objets, agents, opérations et propriétés) à l'aide d'un vocabulaire de base que l'on retrouve majoritairement présent dans la nomenclature d'un dictionnaire des mots fondamentaux datant du milieu des années 1960. On met ainsi en évidence l'un des procédés utilisés par l'énonciateur pour se faire comprendre de son public : il recourt à des mots de la langue commune, censés être partagés par tous les locuteurs.

Comme on l'a rappelé, ce type de test génère de nombreuses scories qui doivent être soigneusement éliminées avant d'entreprendre toute forme d'interprétation. Cependant, ce défaut est minoré par le fait que dans le cadre de ce travail, il ne s'agit pas d'une étude thématique, qui n'aurait pas grand sens, mais de la mise au jour de l'usage d'une composante d'un sociolecte. Pour un tel objectif, ce type de test offre un moyen de comparaison rapide et reproductible.

Une autre composante importante d'un parler d'entreprise réside dans l'utilisation de la terminologie du domaine ou du secteur d'activité. Cependant, il n'a pas été possible de disposer du même type de comparaison que celle qui a été réalisée pour la langue commune. On a dû procéder différemment pour faire apparaître l'usage de la terminologie bancaire dans les rapports d'activité.

La démarche que l'on a adoptée pour traiter ce point s'apparente à une étude d'implantation terminologique. On a identifié une source possible de comparaison dans le *Lexique de la banque et des marchés financiers*. Cet ouvrage a été retenu, bien qu'il adopte un point de vue plus juridique qu'économique sur la terminologie bancaire. Il recense et définit 2400 termes relativement courants dans le domaine, sans entrer dans un niveau de détail trop fin, qui aurait été inadapté pour une comparaison avec le vocabulaire du corpus. Pour chacune des entrées de ce lexique, on a procédé non seulement à la vérification de la présence de la dénomination dans les textes du corpus, mais on a aussi contrôlé la compatibilité de sa définition avec l'usage qui en est fait. Comme les usages sont rarement systématiques, on a retenu une proportion de deux tiers des utilisations pour classer un mot dans la catégorie des termes. Ces usages ont été contrôlés pour les formes ou les expressions peu fréquentes à l'aide des concordances, alors que les unités dotées d'une fréquence

élevée ont été contrôlées à partir de la liste des segments répétés dans lesquels elles sont incluses. On a alors identifié environ 300 termes attestés dans le corpus, soit un huitième de la nomenclature du lexique pris comme référence. De cet ensemble, on a choisi d'étudier les termes récurrents dans le vocabulaire du corpus, c'est-à-dire ceux dont la fréquence est égale ou supérieure à vingt occurrences et dont la présence est attestée dans plusieurs parties consécutives. Cette soixantaine d'unités totalise 90% des occurrences des 300 termes attestés.

La répartition des termes récurrents entre unités simples et complexes se fait dans des proportions inverses de celles observées pour les termes du lexique et pour les termes attestés : trois quarts d'unités simples pour un quart d'unités complexes. Le traitement des composants de l'activité bancaire révèle également des différences de répartition. On constate que les termes relevant des propriétés représentent 40% de la nomenclature du lexique et moins de 5% des termes attestés récurrents. Ces éléments sont concordants pour que l'on interprète la sélection des termes opérée dans les textes du corpus comme visant majoritairement la sélection de notions générales au détriment de termes désignant des points de détail relatifs à l'activité bancaire. On a ensuite cherché à caractériser plus précisément la sélection opérée par l'énonciateur. Pour cela, on a contrôlé le traitement de la soixantaine de termes récurrents dans deux dictionnaires de langue contemporains, le *Trésor de la langue française* et le *Nouveau Petit Robert* dans leurs versions électroniques. Le premier niveau de contrôle a porté sur la présence des termes dans les nomenclatures des deux dictionnaires : une dizaine de termes en sont exclus. Pour les termes présents, on a adopté comme indice de banalisation l'absence de marque de domaine. Même si un tel indice doit être manipulé avec précaution, on constate que les termes les plus utilisés sont également ceux qui sont banalisés ou en passe de l'être. On dispose donc d'une nouvelle indication sur la sélection opérée sur le *continuum* du parler d'entreprise dans le vocabulaire des rapports d'activité. Les termes les plus utilisés sont également ceux qui sont probablement les plus répandus du fait de leur banalisation.

Cette caractérisation porte sur une partie seulement des termes attestés dans le corpus. Cependant, on note qu'elle permet de caractériser la sélection opérée par un locuteur qui privilégie les termes généraux et plutôt banalisés. L'étude

d'implantation est assez longue à réaliser. En effet, même si elle est assistée par ordinateur, le dépouillement reste en grande partie manuel. Cependant, la mise en place d'une procédure automatique aurait nécessité que le texte fût lemmatisé ou que les termes du lexique utilisés pour la comparaison fussent déclinés a priori avec toutes leurs variantes. Or, le repérage manuel à l'aide des formes simples montre que les usages ne sont pas si facilement prédictibles et qu'il peut être coûteux d'envisager à l'avance toutes les possibilités de réalisation. En revanche, le repérage à l'aide des segments répétés incluant les formes simples permet d'identifier rapidement la variante qui est la plus utilisée dans le corpus.

Les relevés réalisés au cours des deux premières évaluations du vocabulaire du corpus ont mis en évidence la présence d'une phraséologie propre au parler d'entreprise. Afin de la caractériser, on ne disposait ni d'un corpus de référence, ni d'une terminologie déjà en partie recensée, comme pour les deux cas précédents. Il a donc été nécessaire de proposer une approche qui soit en accord avec le but poursuivi, c'est-à-dire, qui permette d'identifier les propriétés remarquables de la phraséologie des rapports d'activité sans entreprendre un recensement exhaustif.

Dans ce but, on a eu recours aux données du tableau lexical entier. Celui-ci permet, entre autres, d'identifier les formes qui sont présentes dans toutes les parties du corpus et qui appartiennent de ce fait au vocabulaire commun. Les cinquante premières formes thématiques (on a exclu de ce recensement les mots grammaticaux) se répartissent en quatre strates. On retrouve des unités appartenant à la langue commune et à la terminologie bancaire, mais aussi des noms propres ainsi que des formes n'appartenant à aucun des précédents ensembles.

La répartition de ces unités entre les composants types de l'activité bancaire est différente de celle des termes récurrents : les propriétés et les unités diverses (dont les noms propres) y occupent une place plus importante, au détriment des opérations et des objets. Dans le but de caractériser cette phraséologie, on a utilisé trois unités appartenant aux trois strates déjà identifiées (langue commune, terminologie bancaire, mots pour lesquels une partie de leur aire sémantique est utilisée, sans pour autant qu'ils correspondent à des termes), ainsi qu'un quatrième ensemble de contrôle dont les éléments appartiennent aussi au vocabulaire commun.

On a alors recherché les syntagmes ayant de bonnes chances d'être lexicalisés et dans lesquels apparaît l'une des 12 formes retenues pour le test. Pour identifier ces syntagmes, on a recherché les formes simples incluses dans les segments répétés. Ces derniers ont fourni, outre le syntagme assorti de sa fréquence, des extraits de phrases incluant le syntagme et permettant ainsi de le délimiter avec précision.

Enfin, la récurrence de chaque unité phraséologique a été contrôlée à partir de la ventilation de ses occurrences sur les différentes parties du corpus. Plus d'une centaine de syntagmes ont alors été identifiés, tous ayant une fréquence égale ou supérieure à 10. La phraséologie des rapports d'activité est caractérisée par la présence d'unités servant à désigner des réalités économiques aux contours plus ou moins bien dessinés. Elles paraissent compréhensibles tant que l'on ne cherche pas à en préciser le contenu. Elles se répartissent sur un *continuum* allant de la langue commune au jargon du Crédit agricole, en passant par la terminologie bancaire et un vocabulaire au service d'une vulgate de l'activité bancaire que l'on retrouve dans les rapports d'activité d'autres établissements bancaires français. Les unités appartenant à ces différentes variétés de langue sont enchevêtrées dans les discours et se font concurrence. Ce dernier point est mis en lumière lorsque l'on projette les occurrences des syntagmes lexicalisés sur les différentes parties du corpus. Les causes de ce renouvellement semblent aussi bien être internes qu'externes, lorsqu'elles sont concomitantes au développement d'activités nouvelles.

Encore une fois, on n'a pas cherché à atteindre un relevé systématique, mais plutôt une caractérisation. Celle-ci, pour limitée qu'elle soit dans son étendue, a néanmoins permis d'établir méthodiquement la présence de diverses variétés de langue mises à contribution par l'énonciateur, afin de produire une restitution de l'activité bancaire qui se veut accessible à un large public. On dispose alors d'indications concordantes qui ont permis d'évaluer le résultat de la collecte destinée à l'élaboration d'un référentiel terminologique adaptable au contexte.

Les observations réalisées dans les points précédents pourraient laisser penser que l'énonciateur collectif des rapports d'activité atteint son objectif communicationnel en recourant principalement à des expressions connues par un large public. Autrement dit, pour évoquer les activités multiples d'une banque auprès

d'un public de néophytes, il suffirait de choisir des mots simples dont le sens est transparent pour les interlocuteurs.

Cependant, deux points viennent contredire une telle conception de l'usage d'un parler d'entreprise. Le premier tient aux variations de fréquence qui sont susceptibles de rendre plus ou moins visibles certaines unités. Si les mêmes mots simples suffisaient d'une année sur l'autre pour évoquer la même réalité, comment expliquer les variations de fréquence qui affectent ces unités ? Il semble que l'usage d'un vocabulaire doive alors être rapporté aux caractéristiques du genre de discours. Il a donc fallu définir les moyens de mettre en lumière une telle articulation. Un second point vient également contredire l'impression d'évidence que l'énonciateur des rapports d'activité cherche à entretenir. On a vu précédemment que certains termes, présents dans les textes, ne sont pas encore banalisés. Compte tenu de la visée des rapports d'activité – communiquer des informations relatives à l'activité d'une banque auprès d'un public de néophytes –, il est possible que ces termes fassent l'objet d'une prise en charge énonciative destinée à faciliter la compréhension des lecteurs, comme c'est le cas dans les discours de vulgarisation. On a donc cherché à vérifier ce point.

L'observation des fréquences ainsi que l'interprétation de leurs variations a nécessité que l'on précise au préalable l'existence de motifs de variation directement liés aux propriétés du genre discursif. En comparant la période d'apparition de certains événements majeurs intervenus dans le secteur bancaire avec l'évocation de ces événements dans les rapports d'activité, on met en évidence trois causes de variations autres que thématiques. L'effet de discrétion peut ainsi minorer la fréquence d'une expression (taux d'intérêt), dès lors qu'elle désigne un phénomène peu valorisant (baisse des taux sur la période 1995-2003). A contrario, l'effet de valorisation est susceptible de mettre en exergue d'autres expressions parce qu'elles sont liées à l'actualité institutionnelle (cotation en 2001), quand bien même la conjoncture boursière n'est pas bonne (effondrement du CAC 40 en 2001). Enfin, certaines variations de fréquences peuvent être attribuées à des phénomènes mécaniques, tels que le changement d'unité monétaire ou la disparition du dispositif textuel qui constituait le contexte principal d'apparition d'une expression. Ces trois causes de variation des fréquences ayant été explicitées, on a cherché à préciser la

manière dont les variétés de langue ont été sélectionnées durant la période couverte par le corpus.

Afin d'étudier ce point, on est reparti de la cinquantaine de mots identifiés précédemment comme appartenant au vocabulaire commun thématique du corpus. On a alors observé la répartition de quatre sous-ensembles représentant respectivement la langue commune, la terminologie bancaire, les mots dont l'usage restreint les facettes sémantiques, ainsi qu'un quatrième groupe plus hétérogène. On a alors trouvé la confirmation d'observations réalisées précédemment. Par ailleurs, l'examen de la manière dont certains termes non encore banalisés sont pris en charge dans les discours a nécessité que l'on vérifie la présence de procédés de reformulation dans les textes du corpus. Les indices typographiques et métalinguistiques de reformulation ont été recherchés de façon systématique pour certains d'entre eux, comme la présence de parenthèses, l'utilisation de noms d'unité, ou encore, l'expression "est un" et ses variantes. On a également recherché l'inclusion des termes dans des paradigmes désignationnels qui ont été explorés à l'échelle des paragraphes à travers le repérage des reprises anaphoriques.

En ce qui concerne l'évolution de l'usage de certaines variétés de langue, on a mis en évidence, tout d'abord, une augmentation de la fréquence des unités relevant de la langue commune à partir de l'année 2001. On est alors frappé par le parallèle qui peut être établi entre le renforcement de la diffusion des rapports d'activité, vers un public de plusieurs millions d'actionnaires individuels, et le recours accentué à des expressions relevant de la langue commune afin d'évoquer les aspects les plus commerciaux de l'activité bancaire. L'usage de la langue commune est donc étroitement associé à la réalisation de la visée communicationnelle des rapports d'activité. En revanche, on a constaté que la terminologie bancaire a une portée thématique plus marquée et que son usage varie avec l'évocation de certains aspects de l'activité du secteur. L'ensemble formé par les noms propres fonctionne comme un bon indicateur des changements institutionnels de la banque. Les éléments du quatrième groupe d'unités, qui est plus hétérogène, se répartissent en fonction des trois comportements qui viennent d'être mentionnés.

Par ailleurs, on a pu établir que les termes spécialisés, dont on peut penser qu'ils ne sont pas connus du grand public, ne font pas l'objet de reformulations

destinées à en expliciter le sens, hormis quelques très rares exceptions. On fait alors l'hypothèse que c'est le procédé adopté par l'énonciateur qui bloque le mécanisme de prise en charge. En effet, la présence explicite de marques de reformulation risquerait de rompre l'impression d'évidence entretenue par un discours factuel et lisse. On observe cependant que le groupe des noms propres fait exception à cette règle, en bénéficiant d'une relative prise en charge.

Il ressort de cet examen que l'interprétation des variations de fréquences nécessite que l'on connaisse l'influence que peuvent avoir sur elles les principales propriétés du genre de discours. On a tenté de repérer certaines articulations, mais celles-ci nécessitent à chaque fois que l'examen du fonctionnement discursif soit rapporté à des événements extérieurs dont il est nécessaire d'avoir la connaissance. D'autre part, si les discours de communication auxquels se rattachent les rapports d'activité sont bien destinés à informer un public de néophytes, ils ne possèdent pas toutes les caractéristiques d'un discours de vulgarisation. En effet, leur finalité consistant plus à rassurer les lecteurs qu'à les instruire, ces discours n'utilisent pas certains procédés de reformulation caractéristiques. Ce point aurait pourtant présenté un intérêt pour la collecte, car l'un des procédés de la reformulation consiste à proposer une paraphrase ou un synonyme plus accessibles pour expliciter le sens d'un terme difficile. En mettant en relation deux variétés de langue, l'une spécialisée, l'autre mieux adaptée aux lecteurs non initiés, les traces laissées par la reformulation auraient pu offrir une piste pour la collecte du vocabulaire caractéristique du corpus. On constate, néanmoins, que les noms propres sont les seuls à bénéficier de ce type de prise en charge, même si elle n'est pas systématique. Ainsi, on a une confirmation de l'intérêt que représentent ces unités pour la constitution d'un référentiel terminologique adaptable au contexte.

Dès lors que l'on connaît les principales caractéristiques du vocabulaire du corpus, on dispose d'indications qui sont utiles pour l'évaluation d'une partie importante de la collecte. Celle-ci est à double détente, en ce sens qu'elle comporte d'abord l'extraction des dénominations propres présentes dans le corpus, puis la sélection de nouvelles unités appartenant aussi au vocabulaire du corpus et apparaissant dans le voisinage des noms propres. Dans un premier temps, il a été

nécessaire de définir le périmètre de la collecte de premier niveau ainsi que les méthodes permettant de la réaliser.

En effet, dans ce travail, on propose une méthode de repérage, de délimitation et d'extraction des dénominations propres reposant sur les techniques textométriques. Au préalable, il a été nécessaire de préciser que les noms propres collectés correspondent en fait aux dénominations propres utilisées pour référer à un objet unique, quelle que soit la morphologie des dénominations et quel que soit le type d'objet désigné. On a donc collecté des dénominations propres de personnes physiques et morales, des toponymes, des noms de marque et de produits ainsi que des noms d'artéfacts divers. En adoptant une perspective terminologique étendue aux unités qui soulèvent plus de questions d'ordre cognitif que linguistique, on a élargi la collecte des dénominations propres aux unités dotées d'une majuscule initiale et pour lesquelles la caractérisation ne pouvait se faire sans contrôle. Dans la mesure où l'on a souhaité éprouver l'approche retenue sur le plus grand nombre de cas possible, on a visé une collecte exhaustive dont les taux de rappel et de précision indiquent qu'elle tend vers l'exhaustivité. Le résultat obtenu peut ensuite être réutilisé sous la forme de dictionnaire associé à un système de recherche d'information.

La première étape de la collecte consiste à sélectionner toutes les formes simples dotées d'une majuscule initiale ou imprimées en capitales, en dehors des mots grammaticaux identifiables sans ambiguïté. Dans la seconde étape on identifie les dénominations complètes et leurs variantes telles qu'elles apparaissent dans les textes. Deux méthodes ont été utilisées selon que la fréquence des formes est élevée ou faible. Dans le premier cas, l'identification s'est faite en repérant les segments répétés incluant la forme simple. Les segments étant classés par ordre de fréquence inverse, la recherche de la forme simple fournit deux sortes d'indications. Les premières correspondent aux dénominations et les secondes à des morceaux de phrases incluant le segment-dénomination et permettant de le délimiter avec précision. Les faibles fréquences ont été traitées par lots à partir des concordances qui ont fourni leur contexte phrastique immédiat. Ces dernières permettant d'afficher ordinairement une trentaine de lignes de contextes, pour les formes ayant une dizaine d'occurrences on a procédé par lots de trois, et pour les hapax, c'est une trentaine de formes qui ont été projetées en même temps. Les dénominations propres

identifiées ont ensuite été stockées dans un tableur avec leur fréquence. Ce travail a également permis la sélection de variantes qui, autrement, auraient pu être difficiles à repérer. À l'issue de cette étape, on a recensé un peu plus de 2800 dénominations propres. Un tiers de ces dénominations sont des hapax et 9% ont une fréquence égale ou supérieure à dix occurrences. D'une manière générale, c'est parmi les unités possédant au moins cette fréquence que l'on relève le plus de variantes.

On a ensuite précisé la place de ces unités dans le corpus : on a constaté que les dénominations propres occupent une place importante, à la fois par leur nombre, par leur fréquence et par leur répartition dans la topographie des textes. On a alors noté que le renouvellement des dénominations propres se fait dans les mêmes proportions que le vocabulaire pris dans sa globalité. La surreprésentation des dénominations propres dans les contextes non syntactiques indique la capacité de ces unités à figurer dans différents dispositifs textuels, sans que cela hypothèque leur présence dans les contextes à dominante syntactique. La catégorie est décomposable en classes hiérarchiquement organisées par les fréquences. Les principales classes qui ont été identifiées concernent les agents, les toponymes et les objets. Les agents correspondent majoritairement à des dénominations propres de personnes morales et les objets à des noms de produits. On dispose alors d'une indication relative aux limites de la collecte : dans ce corpus, seuls certains composants de l'activité bancaire sont traités par les dénominations propres. Enfin, cette collecte est de nature à préciser les relations qui s'établissent dans un genre de discours donné entre dénominations propres et référentiel. En effet, on dispose ainsi d'indications, d'une part, sur les pans de référentiel qui sont privilégiés pour la restitution de l'activité bancaire et, d'autre part, sur la hiérarchie des entités à l'intérieur d'une sphère d'activité. Il a alors été possible de réaliser la deuxième étape de la collecte.

On convient que si la collecte vise l'exhaustivité, elle ne l'atteint pas complètement. Cependant, les résultats obtenus sont comparables à ceux que l'on constate dans d'autres contextes de repérage et de catégorisation manuels des dénominations propres. Par ailleurs, la collecte des dénominations propres met en évidence une autre limitation de l'approche retenue : si les noms propres sont présents de façon notable dans le corpus, ils ne sélectionnent qu'une partie des composants de l'activité bancaire. Il reste que cette limitation a le mérite d'être

explicite. Elle indique les directions qui pourront être empruntées si l'on souhaite sélectionner des sources à même de compléter le référentiel terminologique. Une telle entreprise sera néanmoins limitée aux sortes d'objets auxquels des noms propres sont attribués dans le contexte étudié. Il s'agit là d'une limitation d'une portée plus générale dont toute collecte doit tenir compte.

La deuxième étape de la collecte repose sur la façon dont les dénominations propres sont utilisées dans les discours que l'on a rassemblés en corpus. Afin de présenter la manière dont on peut caractériser le rôle que l'énonciateur fait jouer à certaines unités, on a utilisé une classe de dénominations propres qui possèdent certaines propriétés spécifiques. Les toponymes font partie des classes de dénominations les mieux représentées dans le corpus. Ils constituent probablement les noms propres les mieux partagés entre l'énonciateur collectif des rapports d'activité et les lecteurs. On a cherché à savoir si ces indications suffisent à prédire le rôle joué par ces unités dans les textes du corpus.

Afin de vérifier ce point, on a organisé la classe des toponymes en sous-classes correspondant à un découpage courant. Trois ensembles ont alors été sélectionnés : les odonymes, les noms de villes et les noms de pays. Afin d'identifier les rôles tenus par ces trois sous-classes de toponymes, on a eu recours aux techniques textométriques suivantes : fréquence et récurrence de la dénomination et de ses variantes, répartition de leurs occurrences entre les deux principales sortes de contextes identifiées dans le corpus (contextes à dominante syntactique et non syntactique). Enfin, le voisinage du toponyme sur les axes syntagmatique (segments répétés) et paradigmatic (co-occurents) a également été étudié.

On a alors pu observer que les odonymes, majoritairement présents dans les contextes de type non syntactique, sont utilisés dans des dispositifs d'adresses postales. Cependant, ces dernières ne sont pas limitées à une fonction utilitaire, car elles peuvent être utilisées pour amplifier la dimension internationale du groupe bancaire, ou pour marquer son enracinement national, au gré des visées communicationnelles du locuteur. Les noms de ville peuvent servir de simples repères géographiques, mais ils désignent également, pour certains d'entre eux, des places financières. En effet, la plasticité sémantique du nom propre permet au

locuteur de choisir certains aspects du référent en fonction de son propos. De ce point de vue, les noms de pays occupent de nombreux rôles : repères géographiques, ils sont incorporés dans certaines dénominations propres de personnes morales. Mais surtout, ils servent à désigner des marchés économiques ; ce faisant, ils permettent à l'énonciateur d'ouvrir des espaces de discours à partir desquels la restitution de l'activité bancaire est organisée.

À partir de ces éclairages, l'analyse factorielle des correspondances appliquée à une sélection des noms de pays les plus fréquents permet de dresser la carte du développement de l'activité économique du groupe bancaire. On observe alors que l'opposition « France-monde », est progressivement remplacée par « France-Europe », ce qui marque le passage d'un développement économique souhaité à une réalité en train de se mettre en place. L'étude des toponymes présente donc un intérêt en soi, dès lors que l'on cherche à saisir les évolutions d'un référentiel à travers le fonctionnement discursif d'une classe de dénominations propres. Mais l'enseignement majeur pour la collecte que l'on souhaite réaliser réside dans la mise au jour de l'importance des rôles que les dénominations propres sont susceptibles de revêtir dans les discours, soit individuellement, soit collectivement. Ces rôles ont été définis comme la manifestation, pour un genre de discours, d'une pragmatique propre à l'énonciateur, relativement à une ou à plusieurs classes de dénominations propres. Ils sont déterminés non seulement par le genre de discours dans lequel ils apparaissent, mais aussi par l'usage communicationnel qui est fait des contextes dans lesquels les dénominations sont présentes. Ce faisant, ils permettent de caractériser la manière dont ces unités sont intégrées aux discours à travers les relations qu'elles entretiennent avec les autres composantes du vocabulaire. Alors que les toponymes sont généralement exclus des référentiels terminologiques, leur fonctionnement comme noms de marchés plaide en faveur de leur introduction dans ce type d'outils.

L'étude a porté seulement sur les principaux toponymes. On constate, en effet, qu'il est plus aisé de contrôler les investigations lorsque l'on raisonne sur les principaux représentants d'une classe ou sous-classe. Quoi qu'il en soit, on peut toujours procéder à des sondages de manière à évaluer la régularité des observations conduites à partir des unités les plus fréquentes. On note également que si les

toponymes viennent enrichir la nomenclature d'un référentiel terminologique, ils ne permettent pas de capter une partie du vocabulaire caractéristique du corpus en dehors des dénominations propres. Ce phénomène semble directement lié au genre de discours. Il est donc nécessaire d'étudier les rôles qui résultent de l'articulation entre la sémantique des toponymes et les exigences du genre. On en déduit que, pour les rapports d'activité, les toponymes ne peuvent que jouer un rôle secondaire en matière de collecte. En effet, ils sont soit trop spécifiques, lorsqu'ils jouent le rôle de simples repères spatiaux, soit trop généraux, quand ils servent à regrouper une partie de la restitution des activités bancaires.

Avec les toponymes on dispose d'un exemple de fonctionnement discursif de dénominations propres partagées par les interlocuteurs. Les réalités désignées fonctionnent le plus souvent comme des repères destinés à ancrer le discours dans une géographie, sans que celles-ci aient le statut d'objets de discours. Il en résulte un traitement elliptique dans lequel il est admis que différentes facettes d'une même réalité soient évoquées successivement sur un mode implicite. En revanche, on peut penser que les objets spécifiques du référentiel de l'entreprise sont intégrés plus étroitement au propos. C'est cette intégration au discours que l'on a alors cherché à étudier, dans la mesure où l'on a pensé qu'elle permettrait d'identifier les classes de dénominations propres susceptibles d'être à la source d'une collecte du vocabulaire caractéristique du corpus qui soit productive.

On a alors tenté de préciser le statut d'une telle collecte du point de vue de la sémantique des noms propres. Les dénominations propres possèdent un sens instructionnel qui indique aux locuteurs qu'ils doivent aller chercher dans leur mémoire stable le référent qui est évoqué. On a vu avec les toponymes qu'un tel sens est mobilisé efficacement à l'aide de la seule mention de la dénomination propre. Cependant, si les interlocuteurs ne partagent pas la connaissance du référent ou de certaines de ses facettes, le processus de communication risque d'échouer. Normalement, le locuteur prend alors en charge l'apport cognitif nécessaire pour parvenir à se faire comprendre. On a considéré que tous ces ajouts réalisés dans les discours viennent compléter le contenu de la dénomination propre. La limite théorique de ces informations réside dans les relations tissées par chaque

dénomination propre avec d'autres unités du fait de son insertion dans les classes-objets produites dans et par les discours. Une telle option peut se heurter aux difficultés représentées par une trop grande variété ou instabilité des relations ainsi établies. On a alors fait l'hypothèse que pour chaque classe de dénominations propres présentes dans un corpus, il était possible d'établir une structure type de contenu. Celle-ci serait structurée par une question topique, elle-même déterminée par la nature des référents, leur importance dans le référentiel évoqué et les propriétés du genre de discours.

Toujours dans l'optique d'une réalisation terminographique, on a cherché à mettre au jour de telles structures à partir des représentants les plus fréquents pour les classes de dénominations propres les mieux représentées dans le corpus : noms de personnes morales, de lieux, de personnes physiques et d'objets. La construction des structures-types s'est alors appuyée sur un repérage des occurrences réalisé sur un échantillon du corpus. La recherche de nouvelles relations-types a été arrêtée lorsque la saturation de la structure a été atteinte.

Ainsi, on a mis en évidence l'existence d'une organisation récurrente des informations types associées aux dénominations propres dans les textes du corpus. Les noms de personnes morales et d'objets ont été retenus, car ces deux classes possèdent les relations les plus riches, ce qui les rend potentiellement plus productives en termes de collecte d'information. Par ailleurs, il a été possible d'utiliser pour la collecte deux sortes de relations répondant bien aux problématiques d'organisation de l'information : la catégorisation du référent de la dénomination propre et son rattachement à un secteur d'activité, à un marché ou à une forme de clientèle.

Le contenu d'une dénomination propre est donc lié à un corpus source. C'est le parti pris de départ pour la construction des référentiels terminologiques adaptables au contexte que de s'appuyer sur des séries textuelles homogènes. En revanche, pour une dénomination propre donnée, le contenu peut être étoffé dans le référentiel à partir des informations collectées à partir d'autres corpus. En dehors de cette contrainte, on a cherché à fiabiliser le repérage des informations au-delà de l'établissement de la structure-type. La réponse proposée consiste à utiliser les

marques linguistiques présentes dans les discours pour l'établissement de telles relations.

Les unités utilisées pour la collecte sont appelées unités-pilotes. Elles sont employées pour capter une partie du vocabulaire du corpus qui apparaît dans leur voisinage. Cependant, on a vu que cette collecte est limitée aux unités entretenant avec les dénominations propres deux sortes de relations : la catégorisation des référents, ou le rattachement de ces derniers à des ensembles issus des nécessités de l'activité bancaire. Le cadre de la collecte étant ainsi défini, il a fallu mettre en place le dispositif nécessaire à sa réalisation ainsi qu'à l'évaluation des résultats obtenus.

Dans la mesure où les dénominations rassemblées dans les deux classes possèdent des fréquences plus ou moins élevées, plusieurs dispositifs ont été mis en place afin d'identifier les principales unités apparaissant dans leur voisinage. Dans tous les cas, la carte des sections de Lexico a été utilisée afin de localiser une dénomination et ses variantes dans les paragraphes du corpus. Pour les unités de fréquence faible, c'est un repérage manuel qui a été utilisé. Pour les unités dont la fréquence se situe entre dix et vingt occurrences, on a eu recours au calcul des co-occurents. Ces derniers sont obtenus en calculant les formes spécifiques positives des paragraphes dans lesquels apparaissent la dénomination ou ses variantes. On a ensuite contrôlé la manière dont les principaux co-occurents sont reliés aux dénominations propres dans les textes. Enfin, pour les unités possédant une fréquence de plus d'une vingtaine d'occurrences, on a procédé en deux étapes. Le calcul des co-occurents décrit précédemment a été effectué. Mais au lieu de chercher à identifier immédiatement les principaux co-occurents, qui peuvent se révéler très nombreux, on les a rassemblés pour constituer un corpus de deuxième niveau. Celui-ci est composé de tous les stocks de co-occurents des dénominations propres utilisées pour la collecte. L'analyse factorielle des correspondances permet ensuite de comparer ces stocks entre eux. Le calcul des spécificités a permis d'isoler, parmi les co-occurents d'une dénomination propre, ceux qui lui sont particulièrement attachés par opposition aux unités figurant dans les autres stocks. On a ainsi pu réduire le nombre de co-occurents à prendre en compte pour l'établissement du contenu des dénominations propres pilotes.

C'est à partir de ces observations qu'il a ensuite été possible d'établir des critères capables de guider de manière opérationnelle la réalisation de la collecte. Ces critères consistent donc en la caractérisation des moyens syntaxiques et graphiques utilisés dans les textes afin de réaliser les catégorisations des référents des dénominations propres pilotes, ou de rattacher ceux-ci à des secteurs d'activité ou des formes de clientèles. Ils ont permis de fiabiliser le repérage d'unités candidates.

La collecte a été établie à partir de cinquante-deux dénominations d'objets et de cinquante-cinq dénominations de personnes morales. Elle a permis de collecter, à partir des seules relations de catégorisation et de rattachement, plus de huit cents unités, près d'un millier si l'on compte également les dénominations-pilotes. L'enregistrement des données a confirmé le rôle prédominant des dénominations propres de personnes morales qui établissent les connexions les plus nombreuses et les plus fournies avec les autres unités du vocabulaire du corpus. On observe ainsi une sorte d'équivalence entre le type d'entités privilégié dans les rapports d'activité et la tendance à restituer l'activité bancaire à l'aide de notions générales.

L'intégration de ces résultats dans un référentiel terminologique a rendu nécessaire la justification de l'utilisation de relations formelles destinées en terminographie à relier des concepts. On a alors pu déterminer que la transposition de la catégorisation observée dans les discours peut être réalisée dans un tel référentiel sous la forme d'une relation d'instance. En revanche, les rattachements à un secteur d'activité ou à une forme de clientèle ne semblent pouvoir être traduits que par des relations associatives. Celles-ci sont rendues possibles grâce à l'établissement de la précédente relation d'appartenance, elle-même contrôlée par les catégorisations enregistrées pour le référent des dénominations propres. Dans tous les cas, la spécificité des dénominations propres entraîne deux types de restrictions pour leur utilisation dans un référentiel terminologique : il doit s'agir d'unités terminales (sans niveaux subordonnés) ; par ailleurs, l'héritage de la relation d'instance doit être bloqué au niveau immédiatement supérieur à la dénomination propre (la question de l'héritage ne se pose pas pour les relations associatives qui ont généralement une valeur informative).

Ces restrictions admises, les données collectées dans un tableur ont ensuite été restituées sous la forme d'un réseau arborescent à deux niveaux principaux. Cette présentation des référentiels terminologiques adaptables au contexte permet une recherche par mots-clés et une navigation dans l'arborescence du réseau qui a été constitué.

L'étude de la relation de catégorisation a mis en évidence que les dénominations propres-pilotes sont insérées dans des paradigmes désignationnels, dont l'établissement fournit le moyen de capter des unités appartenant à différentes strates d'un parler d'entreprise. On dispose ainsi d'un bon observatoire de l'utilisation d'un tel parler dans une situation de communication type. Les dénominations propres et les unités captées peuvent ensuite servir de références pour conduire de nouvelles explorations destinées à illustrer des situations dans lesquelles d'autres strates d'un même parler sont mobilisées. Par ailleurs, la mise au jour des unités utilisées pour découper la restitution de l'activité a permis de collecter des indications tout aussi utiles en matière d'organisation de l'information. Outre l'identification de l'utilisation de strates différentes d'un même parler d'entreprise, la collecte fournit ici, pour une même entité, des manières différentes de découper la restitution de l'activité, mais aussi des découpages réalisés à des niveaux de précision variables dans le même type de contexte. Enfin, l'observation de variations diachroniques, rendue possible par l'utilisation d'un corpus organisé sous la forme d'une série textuelle chronologique, a permis de mettre en évidence des évolutions dans la catégorisation et le rattachement aux secteurs d'activité. Ce faisant, l'étude du fonctionnement discursif des dénominations propres permet d'observer la manière dont s'articulent l'évolution d'un référentiel et le renouvellement des manières de dire et de classer propres à un locuteur collectif.

Il serait utile, dans une autre étude, d'évaluer le résultat d'une telle collecte à partir de l'usage qui en est fait pour la documentation d'activités d'information. Quoi qu'il en soit, la méthode proposée repose sur des étapes relativement simples qui visent à produire un résultat fournissant des exemples en prise sur le référentiel d'une entreprise particulière, et par conséquent, susceptibles d'être utilisés par des agents dans l'exercice de leurs activités professionnelles. Si certaines étapes d'une telle collecte gagnent probablement à être automatisées – on pense, par exemple, au

repérage des dénominations ou à la conduite d'une étude d'implantation –, l'établissement des structures de contenus, compte tenu de sa nature discursive, doit être réalisée de préférence à l'aide des techniques textométriques.

D'une manière générale, l'approche textométrique présente l'avantage de ne pas introduire de catégories relevant de présupposés extérieurs aux données textuelles exploitées. Elle impose, en revanche, un travail d'explicitation des notions utilisées. Celles-ci se rapportent aussi bien aux dispositifs textuels, avec la nature des contextes, la division des textes en rubriques ou en paragraphes, qu'aux unités linguistiques collectées, avec les conceptions du terme et du nom propre associées à une démarche terminographique. Par ailleurs, une telle approche rend possible la prise en compte des différentes particularités des dispositifs textuels, là encore, en permettant un repérage explicite à l'aide de balises interprétables par les programmes informatiques. Elle permet alors d'explorer toutes les formes de contextes dans lesquels les unités étudiées sont présentes : de la comparaison d'ensembles de vocabulaires volumineux à l'extrait de phrases, elle se révèle robuste et adaptable à différentes sortes de corpus. Enfin, la quantification des données textuelles n'a pas que des vertus statistiques : elle permet des analyses qualitatives précises et surtout reproductibles.

En résumé, la mise en place de référentiels terminologiques adaptables au contexte telle qu'on l'a exposée met en avant l'existence de référentiels d'objets propres à un sociolecte caractéristique d'une organisation. Ces derniers constituent un point d'accès "par le bas", par opposition à la construction d'une terminologie à partir de la mise en évidence du système de concepts propre à un domaine de connaissance ou d'activité. Il apparaît qu'une telle approche rend possible la documentation d'activités d'information correspondant à des situations de communication précises dans un cadre professionnel. Dès lors que l'on accepte que de tels référentiels se laissent appréhender à travers les objets de discours construits dans et par les échanges verbaux, le recours à des sources textuelles semble s'imposer pour la construction de référentiels terminologiques.

En outre, l'approche retenue confirme la nécessité de la prise en compte des genres de discours pour la constitution de ressources textuelles dédiées à de tels

projets terminographiques. Contrairement aux approches qui favorisent l'utilisation de ressources électroniques indifférenciées, la constitution de séries textuelles chronologiques homogènes fournit un cadre contrôlable permettant de mettre au jour des phénomènes de variations linguistiques qui, sans cela, passeraient inaperçues. Elle permet également d'envisager une cartographie rationnelle des usages susceptible de favoriser leur prise en compte dans les problématiques d'organisation de l'information. Une documentation terminologique considérable reste à constituer à partir d'une littérature professionnelle souvent négligée.

L'approche que l'on a adoptée dans ce travail s'appuie en partie sur des unités – les noms propres – qui habituellement n'occupent pas une place centrale dans les travaux terminologiques. Néanmoins, la nécessité d'établir des données fiables, reposant sur des observations réalisées à partir des textes, nous a conduit à explorer une zone intermédiaire entre la terminologie conceptuelle et la terminologie textuelle.

L'utilisation des dénominations propres pour la construction, ainsi que la mise à jour de référentiels terminologiques adaptables au contexte, répond au souci d'ancrer la nomenclature de tels outils dans des contextes singuliers. Ce parti pris se justifie par une visée opérationnelle. En effet, la documentation d'activités d'information au sein des entreprises requiert la prise en compte de référentiels d'objets spécifiques verbalisés à l'aide des parlars d'entreprises. De ce point de vue, la collecte des seuls noms propres présente un intérêt incontestable, accru par le fait que ces unités permettent aussi de sélectionner un certain nombre d'éléments caractéristiques d'un vocabulaire. Comme on l'a constaté, cela nécessite d'identifier les classes de noms propres susceptibles de fournir la collecte la plus productive et la mieux adaptée au projet terminographique qui est poursuivi.

L'organisation du contenu des dénominations propres à partir d'une structure-type apporte une réponse concrète aux objections sur le caractère indéterminé du contenu des noms propres : ce contenu étant de nature discursive, il ne peut être défini qu'à partir des données fournies par les textes d'un corpus. C'est le fonctionnement discursif des dénominations propres qui permet alors de définir les relations sémantiques établies avec d'autres unités. On dispose ainsi d'une base

contrôlable qui permet de sélectionner les relations récurrentes attestées dans un corpus de textes. En outre, l'établissement de séries textuelles chronologiques homogènes permet de contrôler ce contenu tout en donnant accès à certaines évolutions. Par ailleurs, l'intégration des noms propres dans les référentiels terminologiques adaptables au contexte est réalisable en adoptant certaines limitations liées à la nature de ces unités. La collecte réalisée à partir des dénominations propres constitue une approche fiable et reproductible, permettant de fournir des unités candidates susceptibles d'enrichir des référentiels plus classiques ou des langages documentaires tels que les thésaurus.

Une étude reste à faire pour développer une telle approche sur une plus vaste échelle, dans le cadre d'une cartographie des usages linguistiques à l'intérieur des organisations. Même s'il reste encore des points à approfondir, cette cartographie semble de nature à améliorer la diffusion de l'information. Ce processus est encore mal connu à l'intérieur des entreprises, ce qui le rend difficile à évaluer. Son manque de visibilité est renforcé par la composante immatérielle des échanges d'information. Quoiqu'il en soit, les intranets constituent un observatoire favorable pour appréhender ce type de problématique. La prise en compte des usages linguistiques – dont les référentiels terminologiques adaptables au contexte se veulent les instruments –, possède avec les intranets un terrain opérationnel où la diffusion de l'information peut commencer à être évaluée et quantifiée.

CORPUS, LEXIQUES ET DICTIONNAIRES UTILISÉS

1. Corpus

Rapport annuel du Crédit agricole, 1995, Paris, CNCA, 1996, 106 p.

Rapport annuel du Crédit agricole, 1996, Paris, CNCA, 1997, 117 p.

Rapport annuel du Crédit agricole, 1997, Paris, CNCA, 1998, 122 p.

Rapport annuel du Crédit agricole, 1998, Paris, CNCA, 1999, 126 p.

Rapport annuel du Crédit agricole, 1999, Paris, CNCA, 2000, 134 p.

Rapport annuel du Crédit agricole, 2000, Paris, CNCA, 2001, 143 p.

Rapport annuel de Crédit Agricole S.A., 2001, Paris, Crédit Agricole s.a., 2002, 264 p.

Rapport annuel de Crédit Agricole S.A., 2002, Paris, Crédit Agricole s.a., 2003, 251 p.

Rapport d'activité de Crédit Agricole S.A., 2003, Paris, Crédit Agricole s.a., 2004, 98 p.

Extraits du *Code monétaire et financier*, actualisé en permanence et consultable à l'adresse : <http://www.legifrance.gouv.fr>.

Nomenclature du *Lexique de la banque et des marchés financiers*, B. Sousi-Roubi, Paris, Dalloz, 5^e éd., 2001, 259 p.

2. Lexiques et dictionnaires

Thésaurus RESAGRI, 4 vol. alphabétique, permuté, thématique et géographique, Paris, RESAGRI, 1997, 536 p., 540 p., 159 p. et 40 p.

Delphes thésaurus matière – liste hiérarchique, Chambre de commerce et d'industrie de Paris, édition 2004 (format électronique Pdf), 101 p.

Le dictionnaire impossible du Crédit agricole et de la banque, G. Decourselle, Paris, Skippers, 1988, 694 p.

Dictionnaire fondamental de la langue française, (éd. revue et augmentée), G. Gougenheim, Paris, Didier, [1961]1977, 283 p.

Trésor de la Langue Française informatisé, CNRS éditions, Paris, 2004, CD-ROM.

Nouveau Petit Robert, version électronique 2.1, Dictionnaires Le Robert / VUEF, Paris, 2001, CD-ROM.

Grand Robert de la langue française, version électronique 2.0, Paris, Le Robert / SEJER, 2005, CD-ROM.

Grand dictionnaire terminologique, actualisé en permanence par l'Office québécois de la langue française, consultable à l'adresse : <http://www.granddictionnaire.com> .

BIBLIOGRAPHIE

Cette bibliographie est organisée en fonction des principaux domaines traités dans ce travail.

La section « Contextes » regroupe les références traitant de l'intranet, de l'organisation du travail et du Crédit agricole.

« Discours et textes » rassemble les références traitant de l'analyse du discours, des genres de discours, de la linguistique textuelle, des objets de discours, ainsi que des questions relatives à la vulgarisation scientifique et technique.

« Linguistique générale » contient les références relatives aux aspects linguistiques dont il est question dans ce travail et en particulier, les travaux relatifs aux noms propres et à la lexicologie.

« Textométrie et traitement automatique du langage naturel » recense les références relatives à l'exploitation des données textuelles à l'aide de programmes informatiques.

« Sciences de l'information » rassemble les références relatives à la recherche d'information, aux systèmes d'organisation de l'information, aux langages documentaires et aux questions relatives usagers des services d'information.

« Terminologie et terminographie » regroupe les références traitant des aspects théoriques et pratiques de la terminologie.

Contextes

- ALIN Frédéric, LAFONT Denis, MACARY Jean-François, *Le Projet intranet – de l'analyse des besoins de l'entreprise à la mise en œuvre des solutions*, Paris, Eyrolles, 2^e édition, 1998, 329 p.
- ALIN Frédéric, AMOROS Xavier, SALIOU Marc, *L'Entreprise intranet – Guide de conduite de projet*, Paris, Eyrolles, 2002, 228 p.
- AMIS Stéphane, MILAN Simon, *Intranet – mode d'emploi*, Paris, Les Presses du management, 1997, 216 p.
- AZAR-EXBRAYAT Sophie, « Un intranet documentaire au service de l'utilisateur : conception et mise en place », in *Documentaliste – Sciences de l'information*, vol. 39, n° 4-5, Paris, ADBS, 2002, p. 190-201.
- BAYLON Christian, MIGNOT Xavier, *La Communication*, Paris, Nathan, 2003, 416 p.
- BÉNARD Jean-Louis (dir.), *Les Portails d'entreprise – conception et mise en œuvre*, Paris, Hermès Science Publications, 2002, 204 p.
- BENGHOZI Pierre-Jean, FLICHY Patrice, d'IRIBANE Alain, (dossier coord. par), *Réseaux – Internet en entreprise*, vol. 18, n° 104, Paris, Hermès Science, 2000, 364 p.
- BENGHOZI Pierre-Jean *et al.*, *Le Travail en rése@u*, Paris, L'Harmattan, 2002, 95 p.
- BOSSÉNO Christian, *Crédit Agricole – Un siècle au présent*, Paris, Éditions Hervas – Crédit Agricole, 1994, 189 p.
- BOUDON Raymond, BOURRICAUD François, *Dictionnaire critique de la sociologie*, Paris, PUF, 4^{ème} éd., 1994 (1^{ère} éd. : 1982), 714 p.
- BOURDIEU Pierre, *Raisons pratiques – Sur la théorie de l'action*, Paris, Seuil, 1994, 251 p.
- CHANAL Valérie, « Communautés de pratiques et management par projet : à propos de l'ouvrage de Wenger (1998) *Communities of Practice : Learning, Meaning and Identity* », in *M@n@gement*, vol. 3, n°1, 2000, p. 1-30.
- CHOINEL Alain, ROUYER Gérard, *Le Système bancaire français*, Paris, PUF, coll. « Que sais-je ? », n°1264, 1996, 5^{ème} éd. refondue, 127 p.
- CHOINEL Alain, *Le Système bancaire français : approches française et européenne*, Paris, Revue Banque, 2002, 277 p.
- Guide d'information – Le document de référence*, Paris, COB, 2002, 17 p.
- COIFFAIT Nicolas (responsable de l'étude), *Synthèse de l'enquête réalisée sur les rapports annuels [1998]*, Paris, ACTIFIN et HEC Junior Conseil, 1999, 29 p.
- CORNU Gérard, *Vocabulaire juridique*, Paris, PUF, 2005 (1^{ère} éd. : 1987), 970 p.
- COURBON Jean-Claude, TAJAN Sylvère, *Groupware et intranet – Vers le partage des connaissances*, Paris, Dunod, 1999, 2^{ème} éd., 356 p.
- DIAS Cláudia, « Corporate portals : a literature review of a new concept in Information Management », in *International Journal of Information Management*, vol. 21, n° 4, Amsterdam, Elsevier Science Ltd., 2001, p. 269-287.
- Les Échos*, Supplément spécial rapports annuels 2000, Paris, Les Échos, 2000, 31 p.
- Les Échos*, Supplément spécial rapports annuels 2001, Paris, Les Échos, 2001, 27 p.
- ÉMERY Éric (dir.), *Science, technique et valeurs*, Actes des colloques de Crêt-Bérard et de Paris en hommage à Ferdinand Gonseth, 1996, Lausanne, L'Âge d'Homme, 1998, 279 p.

- FOULARD Claude (dir.), *L'Entreprise communicante*, 5^e session de l'Université d'été organisée par le pôle productique Rhône-Alpes, 8-12 sept. 1997, Paris, Hermès, 1998, 375 p.
- GAUDIBERT Jean-Claude, *Le Dernier empire français – Le Crédit Agricole*, Paris, Seghers, 1977, 277 p.
- GERMAIN Michel, *L'Intranet*, Paris, Economica, coll. « Gestion poche », 1998, 112 p.
- GERMAIN Michel, MALAISON Claude (dir.), *L'intranet dans tous ses états – Une approche interculturelle de ses multiples dimensions*, Québec, Isabelle Quentin éditeur, 2004, 195 p.
- GONSETH Ferdinand, « Mon itinéraire philosophique », in *Revue internationale de philosophie*, n°93-94, Wetteren (Belgique), Universa, 1970, p.398-433.
- GONSETH Ferdinand, *Le Référentiel univers obligé de médiatisation*, Lausanne, L'Âge d'Homme, coll. « Dialectica », 1975, 201 p.
- GRAWITZ Madeleine, *Méthodes des sciences sociales*, Paris, Dalloz, 11^{ème} édition, 2001, 1019 p.
- GUESLIN André, *Le Crédit agricole*, Paris, La Découverte, coll. « Repères », 1985, 125 p.
- GUESLIN André, *Histoire des crédits agricoles*, Paris, Economica, 1984, 2 vol., 955 et 463 p.
- GUESLIN André, *Les Origines du Crédit agricole : 1840-1914*, thèse 3^e cycle, Université Nancy 2, 1978, 454 p.
- GUIMARD Anne, *La Communication financière*, Paris, Economica, 1998 (2^{ème} éd.) et 2001 (3^{ème} éd.), 112 p.
- GUIMELLI Christian, *La Pensée sociale*, Paris, PUF, coll. « Que sais-je ? », 127 p.
- HABERMAS Jürgen, *Théorie de l'agir communicationnel – Tome I – Rationalité de l'agir et rationalisation de la société*, traduit de l'allemand par Jean-Marc Ferry, Paris, Fayard, 1987 (1981 pour l'édition allemande), 448 p.
- HABERMAS Jürgen, *Morale et communication*, traduit de l'allemand par Christian Bouchindhomme, Paris, Éditions du Cerf / Flammarion, coll. « Champs Flammarion », 1986 (1983 pour l'édition allemande), 212 p.
- HELDERLE Rodolphe, ROMIEUX William, *Intranet : information et pouvoir partagé*, Rueil-Malmaison, Éditions Liaisons, 1998, 161 p.
- HENRY José-Pierre, RÉGULIER Marcel, *Le Crédit agricole*, Paris, PUF, coll. « Que sais-je ? », n° 2274, 1986, 127 p.
- HOUZELLE Vincent, *Webmaster, émergence d'une nouvelle profession ?*, Mémoire de Master « Gestion de l'information dans l'entreprise » préparé sous la dir. de P.-E. Tixier, IEP de Paris, 2005, 92 p.
- JACOUD Gilles, *Le Système bancaire français*, Paris, Armand Colin, coll. « Synthèse », 1999, 96 p.
- Le Knowledge management* (recueil d'articles publiés dans la *Harvard Business Review*), Paris, Éditions d'Organisation, 1999, 277 p.
- KREMPER François (travaux des cabinets de conseil et d'audit Cauvin Angleys Saint-Pierre, Deloitte Touche Tohmatsu, Ernst & Young Audit et Mazars & Guérard, coordonnés par), *Les Établissements de crédit. Doctrine et pratiques européennes 99*, Meylan, CPC, coll. « L'information financière », 1999, 562 p.
- LAPLANTINE François, *La Description ethnographique*, Paris, Nathan, coll. « 128 », 1996, 128 p.
- LEFEBVRE Alain, *Intranet – client-serveur universel*, Paris, Eyrolles – Informatiques Magazine, 1997, 214 p.

- LEHNISCH Jean-Pierre, *La Communication dans l'entreprise*, Paris, PUF, coll. « Que sais-je ? », n° 2229, 3^e éd., 1991 (1985), 127 p.
- LexiGuide Builder version 1.6 March 2000*, Lexiquest, 2000, 334 p.
- LexiGuide version 1.6 Description Contractuelle*, Montreuil, Lexiquest, 2000, 19 p.
- LIBAERT Thierry, *La Communication d'entreprise*, Paris, Economica, coll. « Gestion poche », 1998, 112 p.
- LOCHARD Guy, BOYER Henri, *La Communication médiatique*, Paris, Seuil, coll. « Mémo », 1998, 96 p.
- MAKUCH Nathalie, PEYRE Jacques, PRUNET Pierre, *Le Crédit agricole*, Paris, Berger-Levrault, 1978, 403 p.
- MIÈGE Bernard, *La Pensée communicationnelle*, Grenoble, Presses universitaires de Grenoble, 2005, 126 p.
- OSMONT Béatrice, *Dynamiques cognitives et stratégies d'utilisateurs*, Paris, Masson, 1995, 185 p.
- PAILLARÉ Joris, *Accompagner le webmestre dans la restructuration d'un site intranet. Quelle aide pour prendre en compte l'utilisateur ?*, mémoire de DESS en Sciences de l'information et de la documentation spécialisées préparé sous la dir. de Brigitte Guyot, Paris, CNAM – INTD, 2005, 85 p.
- Thesaurus RESAGRI 1997*, alphabétique, permuté, thématique et géographique, Paris, RESAGRI, 1997, 4 vol., 536 p., 540 p., 159 p. et 40 p.
- SAINSAULIEU Renaud, *Sociologie de l'entreprise – Organisation, culture et développement*, 2^{ème} éd. revue et mise à jour, Paris, Presses de Sciences Po et Dalloz, 1997, 476 p.
- SANDOVAL Victor, *Intranet – le réseau d'entreprise*, Paris, Hermès, 1996, 152 p.
- SPERBER Dan, WILSON Deirdre, *La Pertinence – Communication et cognition*, Paris, Les Éditions de Minuit, 1989 (1^{ère} éd. : 1986), 396 p.
- TEYSSIER François de, *Milieus financiers et communication*, Paris, PUF, coll. « Que sais-je ? », 1998, 128 p.
- TISSEYRE René-Charles, *Knowledge Management – Théorie et pratique de la gestion des connaissances*, Paris, Hermès, 1999, 185 p.
- VAAST Emmanuelle, BENGHOZI Pierre-Jean, « Intranets et entreprises : technologie, apprentissages et organisation de la cohérence », in V^e Conférence de l'AIM, Montpellier, 26 p.
- VAAST Emmanuelle, « Les intranets, occasions de renforcement et de transformation des communautés de pratique : quatre études de cas », in X^e Conférence de l'Association internationale de management stratégique, 13-15 juin 2001, Université de Laval, Québec, 26 p.
- VEYBEL Laurent, PRIEUR Patrick, *Le Knowledge management dans tous ses états – La gestion des connaissances au service de la performance*, Paris, Éditions d'Organisation, 2003, 129 p.
- La Vie française*, Deloitte Touche Tohmatsu, Ecocom, Montera & Associés, *Quel rapport annuel en l'an 2000 ?*, Paris, étude coéditée par les auteurs collectifs, 1998, 47 p.
- WEBER Max, *Essais sur la théorie de la science*, traduits de l'allemand par Julien Freund, Paris, Plon, 1965, (1^{ère} éd. : 1913), 537 p.
- WINKIN Yves, *Anthropologie de la communication – De la théorie au terrain*, Bruxelles, De Boeck, 1996, 239 p.
- WINKIN Yves, (textes recueillis et présentés par), *La Nouvelle Communication*, Paris, Seuil, coll. « Points Essais », n° 136, 2000, 390 p.

- ZARIFIAN Philippe, *La Nouvelle productivité*, Paris, L'Harmattan, 1990, 212 p.
- ZARIFIAN Philippe, *Objectif compétence*, Rueil-Malmaison, Liaisons, 1999, 229 p.
- ZARIFIAN Philippe, *Quels modèles d'organisation pour l'industrie européenne ? – L'émergence de la Firme coopératrice*, Paris, L'Harmattan, 1993, 287 p.
- ZARIFIAN Philippe, *Travail et communication – Essai sociologique sur le travail dans la grande entreprise industrielle*, Paris, PUF, 1996, 213 p.

Discours et textes

- ADAM Jean-Michel, *Éléments de linguistique textuelle – Théorie et pratique de l'analyse textuelle*, Liège, Mardaga, 1990, 265 p.
- ADAM Jean-Michel, HERMAN Thierry, LUGRIN Gilles (sous la resp. de), *Genres de la presse écrite et analyse de discours*, Besançon, Presses Universitaires franc-comtoises, coll. « SEMEN », 2001, 204 p.
- ADAM Jean-Michel, *Linguistique textuelle : des genres de discours aux textes*, Paris, Nathan / HER, coll. « Fac. linguistique », 1999, 208 p.
- ADAM Jean-Michel, PETITJEAN André, *Le Texte descriptif. Poétique historique et linguistique textuelle*, Paris, Nathan, 1989, 239 p.
- ADAM Jean-Michel, GRIZE Jean-Blaise, ALI BOUACHA Magid, *Textes et discours : catégories pour l'analyse*, Dijon, EUD, 2004, 272 p.
- ADAM Jean-Michel, *Les Textes – types et prototypes – récit, description, argumentation, explication, dialogue – Des réponses originales à l'analyse des textes : la reconnaissance des séquences*, Paris, Nathan/HER, 2001, 4^e édition (1997), 223 p.
- ADAM Jean-Michel, « Unités rédactionnelles et genres discursifs : cadre général pour une approche de la presse écrite », *Pratiques*, n° 94, Metz, p. 3-18.
- APOTHÉLOZ Denis, REICHLER-BÉGUELIN Marie-José, « Construction de la référence et stratégies de désignation », in *TRANEL*, n° 23, vol. édité par Alain Berrendonner et Marie-José Reichler-Béguelin, Neuchâtel, Institut de linguistique de l'Université de Neuchâtel, 1995, p. 227-271.
- APOTHÉLOZ Denis, « Logique naturelle des objets de discours : propriétés-relation d'appartenance », in Jean-Blaise Grize (dir.), *Sémiologie du raisonnement*, Berne – Frankfurt-sur-le-Main – New York, Peter Lang, 1984, p. 187-207.
- AUTHIER Jacqueline, « La mise en scène de la communication dans les discours de vulgarisation scientifique », *Langue française*, n° 53, Larousse, Paris, 1982, p. 34-47.
- BAILLY Florence *et al.*, *Pratiques professionnelles et usages des écrits électroniques*, Paris, L'Harmattan, 2002, 235 p.
- BAKHTINE Mikhaïl (V.N. VOLOCHINOV), *Le Marxisme et la philosophie du langage – Essai d'application de la méthode sociologique en linguistique*, préface de R. Jakobson, trad. du russe par M. Yaguello, Paris, Seuil, 1977, 233 p.
- BAKHTINE Mikhaïl, *Esthétique de la création verbale*, trad. du russe par A. Aucouturier, préface de T. Todorov, Paris, Gallimard, 1984 (1979 pour l'édition russe), 400 p.
- BEACCO Jean-Claude, « À propos de la structuration des communautés discursives : Beaux-arts et appréciatif », dans *Les carnets du Cediscor n°3 - Les enjeux des discours spécialisés*, Paris, Presses de la Sorbonne nouvelle, 1995, p. 135-153.
- BEACCO Jean-Claude, MOIRAND Sophie, « Autour des discours de transmission des connaissances », in *Langages*, n° 117, Paris, Larousse, 1995, p. 32-53.

- BEACCO Jean-Claude, MOIRAND Sophie (coordination éditoriale), *Les Carnets du CEDISCOR*, 3, « Les enjeux des discours spécialisés », Paris, Presses de la Sorbonne nouvelle, 1995, 207 p.
- BIBER Douglas, CONRAD Susan, REPPEN Randi, *Corpus linguistics : Investigating Language Structure and Use*, Cambridge, Cambridge University Press, 1998, X-300 p.
- BIBER Douglas, *Dimensions of register variation : a cross-linguistic comparison*, Cambridge – New York, Cambridge University Press, 1995, XVI-428 p.
- BIBER Douglas, FINEGAN Edward, “Intra-textual variation within medical research articles”, in Nelleke Oostdijk et Pieter de Haan (éd.) *Corpus-based research into language*, Amsterdam/Atlanta, Rodopi, 1994, p. 201-221.
- BIBER Douglas, “Representativeness in Corpus Design”, *Literary and Linguistic Computing*, vol.8, n°4, Oxford, Oxford University Press, 1993, p. 243-257.
- BIBER Douglas, FINEGAN Edward (éd.), *Sociolinguistic Perspectives on Register*, Oxford University Press, New York – Oxford, 1994, 385 p.
- BIBER Douglas, *Variation across speech and writing*, Cambridge, Cambridge University Press, 1988, 299 p.
- BOREL Marie-Jeanne, « Objets de discours et représentation », in *Langages*, n° 103, Paris, Larousse, 1991, p. 36-50.
- BOREL Marie-Jeanne, GRIZE Jean-Blaise, MIÉVILLE Denis, *Essai de logique naturelle*, 2^e édition, Bern, Peter Lang, 1992 (1^{ère} éd. : 1983), 241 p.
- BORZEIX Anni, FRAENKEL Béatrice (coord.), *Langage et travail – Communication, cognition, action*, Paris, CNRS, 2005 (1^{ère} éd. : 2001), 379 p.
- BOURDIEU Pierre, « L'économie des échanges linguistiques », in *Langue française*, n° 34, Paris, Larousse, 1977, p. 17-34.
- BOURDIEU Pierre, *Langage et pouvoir symbolique*, préface de John B. Thompson, Paris, Seuil, 2001, (1982 pour les textes tirés de *Ce que parler veut dire*), 423 p.
- BOURDIEU Pierre, *Questions de sociologie*, Paris, Éditions de Minuit, 1980, 268 p.
- BOURDIEU Pierre, *Le Sens pratique*, Paris, Les Éditions de Minuit, 1980, 475 p.
- BOUTET Josiane, *Construire le sens*, Peter Lang, Bern, Berlin, Frankfurt/M., New York, Paris, Wien, 1997, (1^{ère} éd. : 1994), 236 p.
- BOUTET Josiane, GARDIN Bernard, LACOSTE Michèle, « Discours en situation de travail », in *Langages*, n° 117, Paris, Larousse, 1995, 12-31.
- BOUTET Josiane, *Langage et société*, Paris, Seuil, 1997, coll. « Mémo », 62 p.
- BOUTET Josiane (dir.), *Paroles au travail*, Paris, L'Harmattan, 1995, 267 p.
- BRONCKART Jean-Paul, *Activité langagière, textes et discours – Pour un interactionisme socio-discursif*, Lausanne – Paris, Delachaux et Niestlé, 1996, 351 p.
- BRONCKART Jean-Paul, et al., *Le Fonctionnement des discours – Un modèle psychologique et une méthode d'analyse*, Lausanne – Paris, Delachaux et Niestlé, 1985, 175 p.
- CHARAUDEAU Patrick, « Une analyse sémiolinguistique du discours », in *Langages*, n° 117, Paris, Larousse, 1995, p. 96-111.
- CHARAUDEAU Patrick, « Catégories de langue, catégories de discours et contrat de communication », dans Sophie Moirand et al. (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 315-325.
- CHARAUDEAU Patrick, « De la compétence sociale de communication aux compétences de discours », in Luc Collès et al. (dir.), *Didactique des langues romanes – le*

- développement des compétences chez l'apprenant*, Actes du colloque de Louvain-la-Neuve, janvier 2000, Bruxelles, De Boeck – Duculot, 2001, p. 34-43.
- CHARAUDEAU Patrick, MAINGUENEAU Dominique (dir.), *Dictionnaire d'analyse du discours*, Paris, Seuil, 2002, 661 p.
- CHARAUDEAU Patrick, *Le Discours d'information médiatique – La construction du miroir social*, Paris, Nathan, 1997, 286 p.
- CHASTAIN Charles, « Reference and context », in Keith Gunderson (éd.), *Minnesota Studies in The Philosophy of science vol. VII - Language, Mind and Knowledge*, Minneapolis, University of Minnesota Press, 1975, p. 194-269.
- CHEMINÉE Pascale, *Désignation et signification : paradigmes désignationnels et réseaux analogiques*, thèse sous la dir. de Marie-Françoise Mortureux, Université Paris 10 – Nanterre, 1996, 379 p.
- COLLÈS Luc *et al.* (dir.), *Didactique des langues romanes – le développement des compétences chez l'apprenant – Actes du colloque de Louvain-la-Neuve, janvier 2000*, Bruxelles, De Boeck – Duculot, 2001, 621 p.
- COMBETTES Bernard, TOMASSONE Roberte, *Le Texte informatif : aspects linguistiques*, Bruxelles, De Boeck, 1988, 140 p.
- CORBLIN Francis, *Les Formes de reprise dans le discours – Anaphores et chaînes de référence*, Presses universitaires de Rennes, coll. « Langue / Discours », Rennes, 1995, 246 p.
- CURTIUS Ernst Robert, *La Littérature européenne et le Moyen Âge latin*, Paris, Presses Pocket, 1991 (1^{ère} éd. : 1956), 960 p.
- CUSIN-BERCHE Fabienne, « L'agent peut-il verbaliser l'acteur : actualisation discursive et potentialités linguistiques », in *Langue française* n° 103, Paris, Larousse, 1994, p. 80-90.
- CUSIN-BERCHE Fabienne, *Le Management par les mots – Étude sociolinguistique de la néologie*, Paris, L'Harmattan, 1998, 218 p.
- DUBOIS Danièle (dir.), *Catégorisation et cognition : de la perception au discours*, Paris, Kimé, 1997, 316 p.
- DUBOIS Danièle, « Lexique et catégories naturelles : représentations ou connaissances ? », in *Cahiers de Praxématique*, n° 21, Montpellier, Université Paul Valéry, 1993, p. 105-124.
- DUBOIS Jean, « Lexicologie et analyse d'énoncé », in *Cahiers de lexicologie*, vol. XV – 2, Paris, Didier – Larousse, 1969, p. 115-126.
- FAITA D., « Mondes du travail et pratiques langagières », *Langages*, n° 93, Paris, Larousse, 1989, p. 110-123.
- FOUCAULT Michel, *L'Archéologie du savoir*, Paris, Gallimard, 1969, 257 p.
- FRAENKEL Béatrice, LEGRIS-DESSPORTES Christiane (dir.), *Entreprise et Sémiologie*, Paris, Dunod, 1999, 243 p.
- GARDIN Jean-Claude, *Les Analyses de discours*, Neuchâtel, Delachaux et Niestlé, coll. « Zethos », 1974, 178 p.
- GHIGLIONE Rodolphe, KEKENBOSCH Christiane, LANDRÉ Agnès, *L'Analyse cognitivo-discursive*, Grenoble, Presses universitaires de Grenoble, 1995, 142 p.
- GIRIN Jacques, « La théorie des organisations et la question du langage », in A. Borzeix, B. Fraenkel (coordinatrices), *Langage et travail – Communication, cognition, action*, Paris, CNRS, 2005 (1^{ère} éd. : 2001), p. 167-185.
- GIVÓN T., « The grammar of referential coherence as mental processing instructions », in *Linguistics*, vol. 30-1, Berlin-New York, Mouton de Gruyter, 1992, p. 5-55.

- GOLLUT Jean-Daniel, ZUFFEREY Joël, *Construire un monde – Les phrases initiales de La Comédie humaine*, Lausanne, Delachaux et Niestlé, 2000, 149 p.
- GOLLUT Jean-Daniel, ZUFFEREY Joël, « La référence en début de texte : *Salambô* », in *Textes et discours : catégories pour l'analyse*, Dijon, Éditions universitaires de Dijon, 2004, p. 43-50.
- GRICE H. Paul, « Logique et conversation », traduit de l'américain par F. Berthet et M. Bozon, in *Communications*, n° 30, Paris, Seuil, 1979, p. 57-72.
- GRIZE Jean-Blaise, « Argumentation, schématisation et logique naturelle », in *Revue européenne des sciences sociales*, 12/32, 1974, p. 183-200.
- GRIZE Jean-Blaise, *De la logique à l'argumentation*, Genève, 1982, 266 p.
- GRIZE Jean-Blaise, « Discours, connaissance et actions », in *Dialectica – International review of philosophy of knowledge (...)*, vol. 31, fasc. 1-2 (hommage à F. Gonseth), Bienne (Suisse), 1977, p. 129-139.
- GRIZE Jean-Blaise, *Logique et langage*, Paris, Ophrys, 1990, 153 p.
- GRIZE Jean-Blaise, « Logique naturelle, activité de schématisation et concept de représentation », in *Cahiers de praxématique*, n° 31, Montpellier, PUM3, 1998, p. 115-125.
- GRIZE Jean-Blaise, *Logique naturelle et communications*, Paris, PUF, coll. « Psychologie sociale », 1996, 161 p.
- GRIZE Jean-Blaise, VERGÈS Pierre, SILEM Ahmed, *Salariés face aux nouvelles technologies – Vers une approche socio-logique des représentations sociales*, Paris, Éditions du CNRS, 1987, 223 p.
- GRIZE Jean-Blaise (éd.), *Sémiologie du raisonnement*, Berne – Francfort-sur-le-Main – New York, Peter Lang, 1984, 261 p.
- GUNNARSSON Britt-Louise, « Une analyse de la « structure cognitive » de textes de spécialité », in Sophie Moirand *et al.* (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 171-185.
- HALLIDAY M.A.K., HASAN Ruqaiya, *Cohesion in English*, London, Longman, 1976, 374 p.
- HALLIDAY M.A.K., HASAN Ruqaiya, *Language, context, and text : aspects of language in a social-semiotic perspective*, Oxford, Oxford University Press, 1991 (1^{ère} éd.: 1985), 126 p.
- HARRIS Zellig Sabbetai, « Analyse du discours », traduit de l'américain par F. Dubois-Charlier, in *Langages*, n° 13, Paris, Didier - Larousse, 1969, p. 8-45, (1^{ère} éd. : *Language*, vol. 28, 1952, p. 1-30).
- HARRIS Zellig Sabbetai *et al.*, *The Form of Information in Science – Analysis of an Immunology Sublanguage*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1989, 590 p.
- HARRIS Zellig Sabbetai, *Structures mathématiques du langage*, trad. de l'américain par C. Fuchs, Paris, Dunod, 1971, 248 p., (1^{ère} édition : *Mathematical structures of language*, New York, John Wiley & Sons, 1968).
- HYMES Dell H., *Vers la compétence de communication*, (trad. de France Mugler), Paris, Hatier / Didier, 1991 (1^{ère} éd.: 1982), 219 p.
- JACOBI Daniel, *La Communication scientifique. Discours, figures, modèles*, Grenoble, Presses universitaires de Grenoble, 1999, 277 p.
- JACOBI Daniel, « Du discours scientifique, de sa reformulation et de quelques usages sociaux de la science », *Langue française*, n° 64, Paris, Larousse, 1984, p. 37-51.

- JACOBI Daniel, « L'homme : une nouvelle à caractère scientifique dans la presse et les magazines », in Sophie Moirand *et al.* (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 11-22.
- JACOBI Daniel, « Reformulation et socialisation des connaissances dans des discours de vulgarisation scientifique », *Études de Lettres*, Lausanne, Université de Lausanne, oct.-déc. 1986, p. 23-44.
- JACOBI Daniel, « Les séries superordonnées dans les discours de vulgarisation scientifique », in *Langages*, n° 98, Paris, Larousse, 1990, p. 103-114.
- JACOBI Daniel, *Textes et images de la vulgarisation scientifique*, Bern, Peter Lang, 1987, 166 p.
- JODELET Denise, « Réflexions sur le traitement de la notion de représentation sociale en psychologie sociale », in *Communication-information*, vol. IV, n°2-3, p. 15-42.
- JODELET Denise (dir.), *Les Représentations sociales*, Paris, PUF, 1997, 5^e édition (1^{ère} éd. 1989), 447 p.
- LACOSTE Michèle, « Parole, action, situation », in Josiane Boutet (dir.), *Paroles au travail*, Paris, L'Harmattan, 1995, p. 23-44.
- Langages n° 117, Les Analyses du discours en France*, par D. Maingueneau *et alii*, Paris, Larousse, 1995, 127 p.
- LECOMTE Alain, « Espace des séquences – approche topologique et informatique de la séquence », in *Langages*, n° 81, Paris, Larousse, 1986, p. 91-109.
- LECOMTE M, LÉON J, MARANDIN Jean-Marie, « Analyse de discours : stratégie de description textuelle », *Mots n° 9*, Paris, Presses de la FNSP, 1984, p. 143-167.
- LE NY Jean-François, *Science cognitive et compréhension du langage*, Paris, PUF, 1989, 249 p.
- MAINGUENEAU Dominique, *L'Analyse du discours : introduction aux lectures de l'archive*, Paris, Hachette, coll. « Sup », 1991, 268 p.
- MAINGUENEAU Dominique, *Analyser les textes de communication*, Paris, Nathan, coll. « Lettres sup. », 2000, 211 p.
- MAINGUENEAU Dominique, *Initiation aux méthodes de l'analyse de discours*, Paris, Hachette, 1976, 192 p.
- MAINGUENEAU Dominique, *Les Termes clés de l'analyse de discours*, Paris, Seuil, coll. « Mémo », 1996, 93 p.
- MARANDIN Jean-Marie, « À propos de la notion de thème de discours. Éléments d'analyse dans le récit », *Langue française*, n° 78, Paris, Larousse, 1988, p. 67-87.
- MARANDIN Jean-Marie, « Ce est un autre. L'interprétation anaphorique du syntagme démonstratif », in *Langages*, n° 81, Paris, Larousse, 1986, p. 75-89.
- MARCELLESI Jean-Baptiste, « Analyse de discours à entrée lexicale (application à un corpus de 1924-1925) », in *Langages*, n° 41, Paris, Didier – Larousse, 1976, p. 79-123.
- MARCHAND Pascal, *L'Analyse du discours assistée par ordinateur – Concepts, méthodes, outils*, Paris, Armand Colin, 1998, 222 p.
- MARTIN Évelyne, *Reconnaissance de contextes thématiques dans un corpus textuel – éléments de lexico-sémantique*, Paris, Didier érudition, 1993, 283 p.
- MAZIÈRE Francine, *L'Analyse du discours*, Paris, PUF, coll. « Que sais-je ? », n° 3735, 127 p.
- MIÉVILLE Denis, APOTHÉLOZ Denis, Pierre-Yves BRANDT, « Les organisations raisonnées – Analyse de l'articulation de séquences descriptives », *Travaux du Centre de recherches sémiologiques*, n° 60, Neuchâtel (Suisse), Centre de Recherches Sémiologiques (Université de Neuchâtel), 1992, 124 p.

- MIKHAILOVA Miléna, « Le sujet et le nom dans un discours fondé sur les topoi », in Christian Plantin (dir.), *Lieux communs, topoi, stéréotypes, clichés*, Paris, Kimé, 1993, p. 70-79.
- MOESCHLER Jacques, REBOUL Anne, *Dictionnaire encyclopédique de pragmatique*, Paris, Seuil, 1994, 562 p.
- MOIRAND Sophie, « Des choix méthodologiques pour une linguistique de discours comparative », in *Langages*, n° 105, Paris, Larousse, 1992, p. 28-41.
- MOIRAND Sophie, *Une histoire de discours... - une analyse des discours de la revue Le Français dans le monde 1961-1981*, Paris, Hachette, 1988 (thèse publiée), 802 p.
- MONDADA Lorenza, DUBOIS Danièle, « Construction des objets de discours et catégorisation : une approche des processus de référénciation », in *Tranel*, n° 23, Alain Berrendonner et Marie-José Reichler-Béguelin (éd.), Neuchâtel, Institut de linguistique de l'Université de Neuchâtel, 1995, p. 273-302.
- MONDADA Lorenza, *Décrire la ville – La construction des savoirs urbains dans l'interaction et dans le texte*, Paris, Anthropos – Economica, 2000, 284 p.
- MONDADA Lorenza, « Processus de catégorisation et construction discursive des catégories », in *Catégorisation et cognition : de la perception au discours*, Paris, KIMÉ, 1997, p. 291-313.
- MONDADA Lorenza, *Verbalisation de l'espace et fabrication du savoir – Approche linguistique de la construction des objets de discours*, thèse publiée, Université de Lausanne, Lausanne, 1994, 671 p.
- MOSCOVICI Serge (dir.), *Psychologie sociale*, Paris, PUF, 1998, 7^e édition mise à jour, (1^{ère} éd. : 1984), 620 p.
- OGER Claire, OLLIVIER-YANIV Caroline, « Analyse du discours institutionnel et sociologie compréhensive : vers une anthropologie des discours institutionnels », in *Mots*, n° 71, Lyon, ENS éditions, p. 125-145.
- PÊCHEUX Michel, *L'Inquiétude du discours*, textes de M. Pêcheux choisis et présentés par D. Maldidier, Paris, Éditions des Cendres, 1990, 332 p.
- PÈNE Sophie, BORZEIX Anni, FRAENKEL Béatrice (textes réunis par), *Le Langage dans les organisations – Une nouvelle donne*, Paris, L'Harmattan, 2001, 244 p.
- PETIOT Geneviève, « Voile, tchador ou foulard ? Problèmes de dénomination dans les discours des médias », *Les Carnets du Cediscor 3, Les enjeux des discours spécialisés*, Paris, Presses de la Sorbonne nouvelle, 1995, p. 43-62.
- PÉTROFF André Jean, « Sémiologie de la reformulation dans le discours scientifique et technique », in *Langue française*, n° 64, Paris, Larousse, 1984, p. 53-67.
- PLANTIN Christian (dir.), *Lieux communs, topoi, stéréotypes, clichés*, Paris, Kimé, 1993, 522 p.
- PRINCE E.F., « Toward a Taxonomy of Given-New Information », in P. Cole (éd.), *Radical Pragmatics*, New York, Academic Press, 1981, p. 223-256.
- Rapport annuel... Document de référence 2001 – CRÉDIT AGRICOLE SA*, Paris, Crédit Agricole SA, 2002, 364 p.
- RASTIER François (dir.), *L'Analyse thématique des données textuelles*, Paris, Didier Érudition, 1995, 282 p.
- RASTIER François, « Le problème épistémologique du contexte et le statut de l'interprétation dans les sciences du langage », in *Langage*, n° 129, Paris, Larousse, 1998, p. 97-111.
- REBOUL Anne, MOESCHLER Jacques, *Pragmatique du discours – De l'interprétation de l'énoncé à l'interprétation du discours*, Paris, Armand Colin, 1998, 220 p.

- RICOEUR Paul, *Du Texte à l'action – Essais d'herméneutique II*, Paris, Seuil, 1986, 452 p.
- ROBIN Régine, *Histoire et linguistique*, Paris, Armand Colin, 1973, 306 p.
- SARFATI Georges-Elia, *Éléments d'analyse du discours*, Paris, Nathan, coll. « Nathan Université », 1997, 128 p.
- SCHEGLOFF Emanuel A., « Notes on a Conversational Practice : Formulating Place », in David Sudnow (éd.), *Studies in Social Interaction*, New York – London, The Free Press, 1972, p. 75-119.
- SINCLAIR John, *Preliminary recommendations on Corpus Typology*, Rap. Tech., EAGLES (Expert Advisory Group On Language Engineering Standards), CEE, 1996.
- SITRI Frédérique, *Un modèle d'objet de discours dialogique, entre thématization et reprise*, thèse préparée sous la dir. de Sophie Moirand, Université Paris 3, 1998, 2 vol., 419 et 152 p.
- SITRI Frédérique, *L'Objet du débat – La construction des objets de discours dans des situations argumentatives orales*, Paris, Presses de la Sorbonne nouvelle, 2003, 229 p.
- SUOMELA-SALMI Eija, « Le mot du PDG : un leurre franco-finlandais », in Sophie Moirand et al. (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 219-230.
- TODOROV Tzvetan, *Mikhaïl Bakhtine – le principe dialogique* suivi de *Écrits du Cercle de Bakhtine*, Paris, Seuil, 1981, 315 p.
- VAN HOOLAND Michèle, *Analyse critique du travail langagier. Du langage taylorisé à la compétence langagière*, Paris, L'Harmattan, coll. « Espaces discursifs », 2000, 251 p.
- VERGÈS Pierre, APOTHÉLOZ Denis, MIÉVILLE Denis, « Cet obscur objet du discours : opérations discursives et représentations sociales », in *Revue européenne des sciences sociales*, XXV, 77, Genève, Droz, 1987, p. 209-224.
- VERGÈS Pierre, « Représentations sociales de l'économie : une forme de connaissance », in Denise Jodelet (dir.), *Les Représentations sociales*, Paris, PUF, 1997, p. 407-428.
- VÉRONIQUE Daniel, « Pierre Bourdieu et les échanges linguistiques : quelques réflexions en guise d'hommage », in *Marges linguistiques*, n° 3, Saint-Chamas, MLMS, 2002, p. 1-9.
- WILSON Deirdre, SPERBER Dan, « Remarques sur l'interprétation des énoncés selon Paul Grice », in *Communications*, n° 30, Paris, Seuil, 1979, p. 80-94.

Linguistique générale

- APOTHÉLOZ Denis, *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle*, Genève-Paris, Droz, 1995, 349 p.
- ARIEL Mira, « Referring and accessibility », in *Journal of Linguistics*, vol. 24, n° 1, Cambridge, Cambridge University Press, 1988, p. 65-87.
- ARISTOTE, *Rhétorique*, Livres I et II, texte établi et traduit par M. Dufour, Livre III, texte établi et traduit par M. Dufour et André Wartelle, Paris, Gallimard, 2007 (1^{ère} éd. 1991), 297 p.
- ARISTOTE, *Topiques – Organon V*, trad. par J. Tricot, Paris, Vrin, 2004, 362 p.
- ARNAULD Antoine, LANCELOT Claude, *Grammaire générale et raisonnée*, présentation de Jean-Marc Mandosio, Paris, Allia, 1997 (1^{ère} éd. : 1660), 114 p.
- AUROUX Sylvain, *La Révolution technologique de la grammatisation*, Liège, Mardaga, 1994, 216 p.

- BATIME Christine, « Le langage vu comme un 'référentiel de représentations négociées' constitutif d'un système d'information au sein d'une organisation », in T. Bulot, D. Dubois, C. Batime (éd.), *Cahiers de linguistique sociale. Recherches documentaires*, Actes des deuxièmes Journées Linguistique et ergonomie, Université de Rouen, 17-18 nov, 1994, p 19-25.
- BAUDELLE Yves, « Contribution à une sémantique des noms propres : le cas de l'onomastique romanesque », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 169-180.
- BEAULIEU Françoise, « Dis-moi ton nom, je te dirai qui tu es : réflexion sur les pouvoirs de synthèse de la dénomination », *RSSI*, vol. 15, n° 1-2, Montréal (Québec), Association canadienne de sémiotique, 1995, p. 35-48.
- BENVENISTE Émile, *Problèmes de linguistique générale. 1*, Paris, Gallimard, 1966, (2000, coll. « Tel »), 356 p.
- BENVENISTE Émile, *Problèmes de linguistique générale. 2*, Paris, Gallimard, 1974, (1998, coll. « Tel »), 286 p.
- BERTIN Jacques, *La Graphique et le traitement graphique de l'information*, Paris, Flammarion, 1977, 277 p.
- BILLY Pierre-Henry, « Pour une redéfinition du nom propre », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 137-144.
- BOSREDON Bernard, « « Le N NPR » entre langue et discours », in *Par monts et par vaux – Itinéraires linguistiques et grammaticaux – Mélanges de linguistique générale et française offerts au professeur Martin Riegel pour son soixantième anniversaire par ses collègues et amis*, Louvain – Paris, Peeters, 2001, p. 55-66.
- BOSREDON Bernard, « Prédicativité et lexicalisation », dans *Studia Romanica Posnaniensia*, vol. 16, Poznan, Pologne, Adam Mickiewicz University Press, 1991, p. 85-100.
- BOSREDON Bernard, TAMBA Irène, « Rôle des relations d'implication et d'inclusion dans les descriptions sémantiques », in Martin Riegel et Irène Tamba (éd.), *Actes du Colloque sur l'implication dans les langues naturelles et dans les langages artificiels* organisé par le Groupe de linguistique de l'Université de sciences humaines de Strasbourg, 5-7 déc. 1985, Paris, Klincksieck, 1987, 216 p.
- BOSREDON Bernard, TAMBA Irène, « Titres de tableaux et noms propres », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 123-135.
- BOSREDON Bernard, *Les Titres de tableaux – Une pragmatique de l'identification*, Paris, PUF, 1997, 274 p.
- BRÉAL Michel, *Essai de sémantique : science des significations*, Paris, Hachette, reprint de la 6^e éd. de 1913, Gérard Monfort éditeur, 2001, (1^{ère} éd. : 1897), 349 p.
- BRUNET Étienne, « Les noms propres dans le vocabulaire de Chateaubriand », in *Annales de la Faculté des Lettres et Sciences humaines de Nice*, n° 38, Nice, Les Belles Lettres, 1979, p. 83-94.
- BUYSENS Éric, « Les Noms singuliers », in *Cahiers Ferdinand de Saussure*, n° 28, Genève, Droz, 1973, p. 25-34.
- CATACH Nina, *La Ponctuation*, Paris, PUF, coll. « Que sais-je ? », n° 2818, 1996 (2^{ème} éd.), 127 p.
- CÉDITEC, Le Nom propre en discours, Journée d'étude co-organisée par le Céditec, EA de l'Université Paris 12, le Syled-Cediscor, EA de l'université Paris 3, et le Celted, EA

- de l'Université Paul Verlaine – Metz , 12 oct. 2006, actes à paraître dans Les Carnets du Cediscor.
- CHAMPOLLION Jean-François, *Lettre à M. Dacier (...) relative à l'alphabet des hiéroglyphes phonétiques employés par les Égyptiens pour inscrire sur leurs monuments les titres, les noms et les surnoms des souverains Grecs et Romains*, (suivi de *La bataille des hiéroglyphes* par Jean-Claude Goyon), Fontfroide, Fata Morgana, 1988, [1822], 82 p.
- CHARAUDEAU Patrick, *Grammaire du sens et de l'expression*, Paris, Hachette, 1992, 927 p.
- CHAROLLES Michel, *La Référence et les expressions référentielles en français*, Paris, Ophrys, 2002, 258 p.
- CHAURAND Jacques, « Deux livres sur le nom propre : Marie-Noëlle Gary-Prieur, *Grammaire du nom propre* et Kerstin Jonasson, *Le nom propre. Constructions et interprétations* », in *Nouvelle revue d'onomastique*, n° 23-24, Paris, Société française d'onomastique, 1994, p. 3-12.
- CISLARU Georgeta, *Étude sémantique et discursive du nom de pays dans la presse française avec référence à l'anglais, au roumain et au russe*, Thèse de doctorat en Sciences du langage, Université Paris 3 – Sorbonne nouvelle, 2005, 2 vol., 701 p.
- COLLINOT André, MAZIÈRE Francine, *Un prêt à parler : le dictionnaire*, Paris, PUF, 1997, 226 p.
- COMBETTES Bernard, « Hiérarchie des référents et connaissances partagée : les degrés dans l'opposition connu/nouveau », in *L'Information grammaticale*, n° 52, Paris, Éditions Baillièrre, 1992, p. 11-14.
- CRUSE Alan, « La signification des noms propres de pays en anglais », in S. Rémy-Giraud et P. Rétat (dir.), *Les mots de la nation*, Lyon, PUL, 1996, p. 93-102.
- DARMESTER Arsène, *De la Création actuelle des mots nouveaux dans la langue française et les lois qui la régissent*, Paris (réimpression de l'édition de 1877), Slatkine reprints, 1972, 307 p.
- DELEUZE Gilles, GUATTARI Félix, *Rhizome – Introduction*, Paris, Les Éditions de Minuit, 1976, 74 p.
- DESCLÉS Jean-Pierre, « Langages quasi-naturels articulés avec une base de connaissances : présentation et problèmes », in *Condenser* (suppl. n°1) – *Actes du colloque Traitement automatique des langues naturelles et systèmes documentaires*, Clermont-Ferrand, CILN, 1984, p. 45-94.
- DESCLÉS Jean-Pierre, « Réseaux sémantiques : la nature logique et linguistique des relateurs », *Langages*, n° 87, Paris, Larousse, 1987, p. 55-78.
- DESCAMPS Jean-Luc *et al.*, *Sémantique et concordances*, Paris, INALF / Klincksieck, coll. « Saint-Cloud », 1992, 395 p.
- DOPPAGNE Albert, *Majuscules, abréviations, symboles et signes*, Paris – Louvain-la-Neuve, Duculot, 1991, 111 p.
- DUBOIS Danièle (dir.), *Sémantique et cognition – Catégories, prototypes, typicalité*, Paris, CNRS, 1991, 342 p.
- DUBOIS Jean *et al.*, *Dictionnaire de linguistique et des sciences du langage*, Paris, Larousse, 1994, 514 p.
- DUBOIS Jean, *Grammaire structurale du français : nom et pronom*, Paris, Larousse, 1974 [1965], 189 p.

- DUBOIS Jean, *Le Vocabulaire politique et social en France – de 1869 à 1872 – À travers les œuvres des écrivains, les revues et les journaux* (Thèse publiée), Paris, Larousse, 1962, 460 p.
- DUCROT Oswald, *Le Dire et le dit*, Paris, Éditions de Minuit, 1984, 237 p.
- DUCROT Oswald, SCHAEFFER Jean-Marie, *Nouveau dictionnaire encyclopédique des sciences du langage*, Paris, Seuil, 1972, 1995, 821 p.
- DUMARSAIS César Chesneau, *Œuvres de Dumarsais – Mélanges de grammaire, de philosophie, etc. tirés de l'Encyclopédie*, éditées par Marie-Emile-Guillaume Duchosal et Charles Million, Tome IV, Paris, Pougin, an V [1797], 396 p.
- ECO Umberto, *Sémiotique et philosophie du langage*, trad. de l'italien par Myriem Bouzaher, Paris, PUF, coll. « Quadrige », 2001 (1988 et 1984 pour l'éd. italienne), 285 p.
- ECO Umberto, *La Structure absente – Introduction à la recherche sémiotique*, traduit de l'italien par Uccio Esposito-Torrigiani, Paris, Mercure de France, 1972 (1968 pour l'édition italienne), 447 p.
- ELUERD Roland, *La Lexicologie*, Paris, PUF, coll. « Que sais-je ? », 2000, 128 p.
- ELUERD Roland, *La Pragmatique linguistique*, Paris, Nathan, 1985, 222 p.
- ERMINE Jean-Louis, *Les Systèmes de connaissances*, Paris, Hermès, 1996, 160 p.
- EVENS WALTON Martha (éd.), *Relational models of lexicon*, Cambridge, Cambridge University Press, 1988, 390 p.
- FIALA Pierre, LAFON Pierre, PIGUET Marie-France, (textes réunis par), *La Locution : entre lexique, syntaxe et pragmatique*, Paris, INALF – Klincksieck, coll. « Saint-Cloud », 1997, 346 p.
- FLAUX Nelly, « La catégorisation du nom propre », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 63-73.
- FLANDERS Julia, BAUMAN Syd, CATON Paul, COURNANE Mavis, « Names Proper and Improper : Applying the TEI to the classification of Proper Nouns », in Ide Nancy and Greenstein Daniel (éd.), *Computers and the Humanities*, vol. 31 n° 4, Dordrecht, the Netherlands, Kluwer Academic Publishers, 1998, p. 285-300.
- FONTENELLE Thierry, « Dictionnaires électroniques et relations lexicales : une comparaison entre quelques programmes européens », *Revue française de linguistique appliquée*, vol II-1, Amsterdam, Éditions 'De Werelt', 1997, p. 65-77.
- FORSGRÉN Mats, « Nom propre, référence prédication et fonction grammaticale », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 95-105.
- FRATH Pierre, OUESLATI Rochdi, ROUSSELOT François, « Identification des relations sémantiques par repérage et analyse des cooccurrences de signes linguistiques », in Jean Charlet, Manuel Zacklad, Gilles Kassel (éd.), *Ingénierie des connaissances - Évolution récentes et nouveaux défis*, Paris, Eyrolles, « Collection technique et scientifique des télécommunications », 2000, p. 291-304.
- FREGE Gottlob, « Sens et dénotation », dans *Écrits logiques et philosophiques*, traduction et introduction de Claude Imbert, Paris, Seuil, 1971, (1^{ère} éd. : 1892), p. 102-126.
- FUCHS Catherine, LE GOFFIC Pierre, *Initiation aux problèmes des linguistiques contemporaines*, Paris, Hachette, 1985, 135 p.
- FUCHS Catherine (dir.), *La linguistique cognitive*, Paris, Ophrys – MSH, 2004, 261 p.
- FUCHS Catherine, *La Paraphrase*, Paris, PUF, 1982, 184 p.
- FUCHS Catherine, *Paraphrase et énonciation*, Paris, Ophrys, 1994, 185 p.

- GADET Françoise, *La Variation sociale en français*, Paris, Ophrys, 2003, 135 p.
- GALISSON Robert, ANDRÉ Jean-Claude, *Dictionnaire de noms de marques courants – Essai de lexiculture ordinaire*, Paris, Didier, 1998, 342 p.
- GARY-PRIEUR Marie-Noëlle, *La Grammaire du nom propre*, Paris, PUF, 1994, 262 p.
- GARY-PRIEUR Marie-Noëlle, « La modalisation du nom propre », in *Langue française* n° 92, Paris, Larousse, 1991, p. 46-63.
- GARY-PRIEUR Marie-Noëlle, « Le nom propre constitue-t-il une catégorie linguistique ? », in *Langue française* n° 92, Paris, Larousse, 1991, p. 4-25.
- GARY-PRIEUR Marie-Noëlle, *Les Termes clés de la linguistique*, Paris, Seuil, coll. « Mémo », 1999, 61 p.
- GARY-PRIEUR Marie-Noëlle, *L'Individu pluriel*
- GENETTE Gérard, *Figures II*, Paris, Seuil, 1969, 293 p.
- GENETTE Gérard, *Seuils*, Paris, Seuil, 1987, 388 p.
- GILLIÉRON Jules, MONGIN Jean, *Scier dans la Gaule romane du Sud et de l'Est*, Paris, Honoré Champion, 1905, 30 p.
- GOODY Jack, *Entre l'oralité et l'écriture*, traduit de l'anglais par Denise Paulme et révisé par Pascal Ferroli, Paris, PUF, 1994, 323 p.
- GOODY Jack, *La Logique de l'écriture – Aux origines des sociétés humaines*, Paris, Armand Colin, 1986, 195 p.
- GOODY Jack, *La Raison graphique – La Domestication de la pensée sauvage*, traduit de l'anglais par J. Bazin et A. Bensa, Paris, Les Éditions de Minuit, 1979, 274 p.
- GOUGENHEIM Georges, *Dictionnaire fondamental de la langue française*, (éd. revue et augmentée), Paris, Didier, 1977 (1^{ère} éd. : 1961), 283 p.
- GOUGENHEIM Georges, MICHÉA R., SAUVAGEOT A., RIVENC P., *L'Élaboration du français fondamental (1^{er} degré) – Étude sur l'établissement d'un vocabulaire et d'une grammaire de base*, nouvelle édition refondue et augmentée, Paris, Didier, 1964, 302 p.
- GRASS Thierry, « Typologie et traductibilité des noms propres de l'allemand vers le français », in *TAL*, vol. 41, n° 3, dirigé par Denis Maurel et Franz Guenther, ATALA / Hermès Science Publications, Paris, 2000, p. 643-669.
- GREIMAS Algirdas Julien, *La Mode en 1830*, Paris, PUF, 2000 (1^{ère} éd. : 1948), 419 p.
- GREIMAS Algirdas Julien, *Sémantique structurale*, Paris, Larousse, 1966, 262 p.
- GREIMAS Algirdas Julien, COURTÉS Joseph, *Sémiotique – Dictionnaire raisonné de la théorie du langage*, Paris, Hachette, 1993, 454 p.
- GRIZE Jean-Blaise, « Logique et Langue », in *Dialectica – International review of philosophy of knowledge (...)*, vol. 24, fasc. 1-3 (hommage à F. Gonseth), Bienne (Suisse), Dialectica, 1970, p. 67-77.
- GROSS Gaston, « Degré de figement des noms composés », *Langue française*, n° 87, Paris, Larousse, 1988, p. 57-72.
- GROSS Gaston, *Les Expressions figées en français, noms composés et autres locutions*, Paris, Ophrys, coll. « L'essentiel français », 1996, 161 p.
- GUENTHNER Franz, SABATIER Paul, « Sémantique formelle et représentation des connaissances », in *Langages*, n° 87, Paris, Larousse, 1987, p. 103-122.
- GUILLAUME Gustave, *Leçons de linguistique de Gustave Guillaume – 1944-1945 séries A et B*, (publiées sous la dir. de Roch Valin et al.), Lille, Presses universitaires de Lille, 1992, 295 p.

- ISAAC Luc, « Nom propre et publicité », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest 21-24 avril 1994, Paris, Klincksieck, 1995, p. 161-168.
- JAKOBSON Roman, *Essais de linguistique générale – 1. Les Fondations du langage*, traduit de l'anglais et préfacé par Nicolas Ruwet, Paris, Les Éditions de Minuit, 1963, 260 p.
- JAKOBSON Roman, *Essais de linguistique générale – 2. Rapports internes et externes du langage*, Paris, Les Éditions de Minuit, 1973, 317 p.
- JONASSON Kestin, *Le Nom propre – Constructions et interprétations*, Louvain-la-Neuve (Belgique), Duculot, 1994, 255 p.
- KERBRAT-ORECCHIONI Catherine, *L'Énonciation. De la subjectivité dans le langage*, Paris, Armand Colin, 4^e éd. 1999, 267p. (1^{ère} éd. : 1980, 290 p.).
- KERBRAT ORECCHIONI Catherine, *L'Implicite*, Paris, Armand Colin, 1986 (2^e éd.: 1998), 404 p.
- KERBRAT ORECCHIONI Catherine, *Les Interactions verbales – 1. Approche interactionnelle et structure des conversations*, Paris, Armand Colin, 1990 (3^e éd.1998), 318 p.
- KLEIBER Georges, *L'Anaphore associative*, Paris, PUF, 2001, 386 p.
- KLEIBER Georges, *Anaphores et pronoms*, Louvain-la-Neuve, Duculot, 1994, 229 p.
- KLEIBER Georges, « Cap sur les topiques avec le pronom *IL* », in *L'information grammaticale*, n°52, Paris, Éditions Baillière, 1992, p. 11-15-25.
- KLEIBER Georges, « Contexte, interprétation et mémoire : approche standard vs approche cognitive », in *Langue française* n° 103, Paris, Larousse, 1994, p. 9-22.
- KLEIBER Georges, « Dénomination et relations dénominatives », *Langages*, n° 76, Paris, Larousse, 1984, p. 77-94.
- KLEIBER Georges, *Nominales – Essais de sémantique référentielle*, Paris, Armand Colin, 1994, 247 p.
- KLEIBER Georges, « Noms propres et noms communs : un problème de dénomination », in *META*, vol. 41, n°4, Montréal, Presses de l'Université de Montréal, 1996, p. 567-589.
- KLEIBER Georges, « Peut-on sauver un sens de dénomination pour les noms propres ? », in *Grounding and Headedness in the noun phrase, Functions of Language*, 11:1, Amsterdam, John Benjamins Publishing Company, 2004, p. 115-145.
- KLEIBER Georges, *Problèmes de référence : descriptions définies et noms propres*, Université de Metz, Centre d'analyse syntaxique, Paris [diffusion] Klincksieck, 1981, 538 p.
- KLEIBER Georges, *La Sémantique du prototype*, Paris, PUF, 1990, 199 p.
- KLEIBER Georges, « Sens, référence et existence : que faire de l'extra-linguistique ? », in *Langages*, n° 127, Paris, Larousse, 1997, p. 9-37.
- KLEIBER Georges, « Sur la définition des noms propres : une dizaine d'années après », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest 21-24 avril 1994, Paris, Klincksieck, 1995, p. 11-36.
- KLEIBER Georges, « Sur la sémantique et pragmatique des SN – Le projet Delors et la camarade Catherine », in *L'Information grammaticale*, n° 26, Paris, Baillière, 1985, p. 3-9.
- KRIEG Alice, *Émergence et emplois de la formule « purification ethnique » dans la presse française (1980 – 1994). Une analyse de discours*, thèse préparée sous la direction de Patrick Charaudeau, Université Paris 13, 2000, 3 vol., 840 f.

- KRIPKE Saul, *La Logique des noms propres (Naming and Necessity)*, traduit de l'américain par Pierre Jacob et François Recanati, Paris, Les Éditions de Minuit, 1982 (1972, 1980), 173 p.
- LABOV William, *Sociolinguistique*, trad. de l'américain par A. Kihm, Paris, Minuit, 1976 (1^{ère} éd. : 1972), 457 p.
- LAKOFF George, *Women, Fire, and Dangerous Things – What Categories Reveal about the Mind*, Chicago, London, The University of Chicago Press, 1987, 614 p.
- LAKS Bernard, *Langage et cognition – l'approche connexionniste*, Paris, Hermès, 1996, 188 p.
- Langue française*, n° 64, « Français technique et scientifique : reformulation et enseignement », numéro coordonné par Jean Peytard, Daniel Jacobi et André Pétrouff, Paris, Larousse, 1984.
- Langue française*, n° 92, « Syntaxe et sémantique des noms propres », numéro coordonné par M.-N. Gary-Prieur *et al.*, Paris, Larousse, 1991, 127 p.
- LAUFER Roger (textes réunis par), *La Notion de paragraphe*, Paris, Éditions du CNRS, 1985, 150 p.
- LE GUERN Michel, *Les Deux logiques du langage*, Paris, Honoré Champion, 2003, 181 p.
- LE PESANT Denis, MATHIEU-COLAS Michel, « Introduction aux classes d'objets », in *Langages*, n° 131, Paris, Larousse, 1998, p. 6-33.
- LE PESANT Denis, « Utilisation des propriétés des anaphores dans la définition des relations lexicales », in *Langages* n° 131, Paris, Larousse, 1998, p. 115-124.
- LERAT Pierre, *Sémantique descriptive*, Hachette, coll. « Langue Linguistique Communication », Paris, 1983, 128 p.
- LEROY Sarah, « Quand un nom propre en glose un autre », in *Langues et langages – Le mot et sa glose*, n° 9, Aix-en-Provence, Publications de l'Université de Provence, 2003, p. 169-187.
- LEROY Sarah, *Le Nom propre en français*, Paris, Ophrys, 2004, 137 p.
- LEVRAT Bernard, « 'Sorte de', une façon de rendre compte de la relation d'Hyponymie / hyperonymie dans les réseaux sémantiques », in *Langages*, n° 98, Paris, Larousse, 1990, p. 87-102.
- LÉVI-STRAUSS Claude (dir.), *L'Identité*, Paris, PUF, coll. « Quadrige », (1977 et 1983), 2000, 344 p.
- Lexique des règles typographiques en usage à l'Imprimerie nationale*, Paris, Imprimerie nationale, 3^e édition, 1990, 197 p.
- LYONS John, *Éléments de sémantique*, traduction de J. Durand, Paris, Librairie Larousse, 1978, 295 p.
- LYONS John, *Linguistique générale – introduction à la linguistique théorique*, traduit par Françoise Dubois-Charlier et David Robinson, Paris, Larousse, 1970 (1^{ère} éd. : 1968), 382 p.
- LYONS John, *Sémantique linguistique*, (traduction de J. Durand et D. Boulonnais), Paris, Larousse, 1990, (1^{ère} éd. : 1978), 495 p.
- MARTIN Robert, *Comprendre la linguistique*, Paris, PUF, 2002, 190 p.
- MARTIN Robert, *Inférence, antonymie et paraphrase – Éléments pour une théorie sémantique*, Strasbourg, Librairie C. Klincksieck, 1976, 174 p.
- MARTIN Robert, *Langage et croyance – Les Univers de croyance dans la théorie sémantique*, Bruxelles, Pierre Mardaga, 1987, 189 p.
- MARTIN Robert, *Pour une logique du sens*, Paris, PUF, 1992 (2^{ème} éd.), 319 p.

- MATORÉ Georges, *L'Espace humain – L'Expression de l'espace dans la vie, la pensée et l'art contemporains*, Paris, La Colombe, 1962, 299 p.
- MATORÉ Georges, *La Méthode en lexicologie*, Paris, Didier, 1953, 126 p.
- MATORÉ Georges, *Le Vocabulaire de la société du XVI^e siècle*, Paris, PUF, 1988, 379 p.
- MATORÉ Georges, *Le Vocabulaire et la société sous Louis-Philippe*, Genève, Slatkine reprints, 1967, 369 p.
- MATHIEU-COLAS Michel, « Illustration d'une classe d'objets : les voies de communication », in *Langages* n° 131, Paris, Larousse, 1998, p. 77-90.
- MATHIEU-COLAS Michel, « Typologie de la composition nominale », *Cahiers de lexicologie*, n° 69, Paris, Didier, 1996, p. 65-118.
- MAURO Tullio de, *Une introduction à la sémantique*, trad. de l'italien par L.-J. Calvet, Paris, Payot, 1969, 220 p.
- MELBY A. K., « Des causes et des effets de l'asymétrie partielle des réseaux sémantiques liés aux langues naturelles », *Cahiers de lexicologie*, n° 58, Paris, Didier érudition, 1991, p. 5-43.
- MILL John Stuart, *Système de la logique déductive et inductive : exposé des principes de la preuve et des méthodes de recherche scientifique*, trad. de l'anglais par L. Peisse sur la 6^{ème} éd., Paris, Alcan, 1889 (1865 pour la 1^{ère} éd. en anglais), 2 vol., XXIV-514, XII-561 p.
- MORTUREUX Marie-Françoise, « Comment peut-on définir la propriété d'un mot » in Sophie Moirand *et al.* (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 3-10.
- MORTUREUX Marie-Françoise, « La dénomination : approche socio-linguistique », *Langages*, n° 76, Paris, Larousse, 1984, p. 95-112.
- MORTUREUX Marie-Françoise, « Hypo-/hyperonyme et stratégies discursives », *Langages*, n°98, Paris, Larousse, 1990, p. 115-121.
- MORTUREUX Marie-Françoise, *La Lexicologie entre langue et discours*, Paris, SEDES, coll. « Campus », 1997, 192 p.
- MORTUREUX Marie-Françoise, « Paradigmes désignationnels », *SEMEN*, 8, Paris, Université de Franche-Comté / Les Belles Lettres, 1993, p.123-141.
- MORTUREUX Marie-Françoise, « Paraphrase et métalangage dans le dialogue de vulgarisation », *Langue française*, n° 53, Paris, Larousse, 1982, p. 48-61.
- Mots – Les Langages du politique – Noms propres*, n° 63, numéro coordonné par Jean-Paul Honoré, Marie-Anne Paveau, Gabriel Périès, Lyon, ENS Éditions, 2000, 159 p.
- MOUNIN Georges (dir.), *Dictionnaire de la linguistique*, Paris, PUF, 1974, 1995 2^e édition, coll. « Quadrige », 340 p.
- MULDER Walter de, « Nom propre et essence psychologique. Vers une analyse cognitive des noms propres ? » in *Lexique*, n° 15, Lille, Presses universitaires du Septentrion, 2000, p. 47-62.
- NOAILLY Michèle (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, 377 p.
- NOAILLY Michèle, « Noms propres hors champ », in *Nouvelle revue d'onomastique*, n° 25-26, Société française d'onomastique, Paris, 1995, p. 11-18.
- NOAILLY Michèle, *Le Substantif épithète*, Paris, PUF, 1990, 221 p.
- OGDEN C.K., RICHARDS I.A., *The Meaning of Meaning – A study of the influence of language upon thought and of the science of symbolism*, San Diego, New York, London, Harcourt, Brace, Jovanovich, 1989 (1^{ère} éd.: 1923), 363 p.

- PARIENTE Jean-Claude, « Le nom propre et la prédication dans les langues naturelles », in *Langages*, n° 66, Paris, Larousse, 1982, p. 37-65.
- PEIRCE Charles S., *Écrits sur le signe*, rassemblés, traduits et commentés par G. Deledalle, Paris, Seuil, 1978, 262 p.
- PICOCHÉ Jacqueline, *Précis de lexicologie française – L'Étude et l'enseignement du vocabulaire*, Paris, Nathan, 1992, 191 p.
- PORHIEL Sylvie, « Les introducteurs de cadres thématiques », in *Cahiers de lexicologie*, vol. 85, n° 2, Paris, Didier, 2004, p. 9-45.
- POTTIER Bernard, *Théorie et analyse en linguistique*, Paris, Hachette, 1987, 223 p.
- PRUVOST Jean, *Dictionnaires et nouvelles technologiques*, Paris, PUF, 2000, 177 p.
- RASTIER François, *Arts et sciences du texte*, Paris, PUF, coll. « Formes sémiotiques », 2001, 303 p.
- RASTIER François, « Catégorisation, typicalité et lexicologie », in Danièle Dubois (dir.), *Sémantique et cognition – Catégories, prototypes, typicalité*, Paris, CNRS, 1991, p. 259-277.
- RASTIER François, « Représentation du contenu lexical et formalismes de l'intelligence artificielle », *Langages*, n° 87, Paris, Larousse, 1987, p. 79-102.
- RASTIER François, *Sémantique et recherches cognitives*, Paris, PUF, 2001, 2^e éd. revue et augmentée (1^{ère} éd. : 1991), 272 p.
- RASTIER François, *Sémantique interprétative*, Paris, PUF, 1996, 2^{ème} éd. corrigée et augmentée (1^{ère} éd. : 1987), 284 p.
- RASTIER François, « La triade sémiotique, le trivium et la sémantique linguistique », in *Nouveaux actes sémiotiques*, n° 9, Limoges, PULIM, 1990, p. 5-39.
- REY Alain, *Encyclopédies et dictionnaires*, Paris, PUF, coll. « Que sais-je ? », 1982, 127 p.
- REY Alain, REY-DEBOVE Josette (rédaction générale), *Le Grand Robert des noms propres*, Dictionnaires Le Robert, Paris, 1991, (1^{ère} éd. : 1974), Présentation, p. XI-XXIX.
- REY Alain, *Théories du signe et du sens*, tome 1, Paris, Klincksieck, 1973, 299 p.
- REY Alain, *Théories du signe et du sens*, tome 2, Paris, Klincksieck, 1976, 408 p.
- REY Alain, *Le Lexique : images et modèles – Du dictionnaire à la lexicologie*, Paris, Armand Colin, 1977, 307 p.
- REY-DEBOVE Josette, *Le Métalangage : étude du discours sur le langage*, Paris, Le Robert, nouvelle éd., Armand Colin, coll. « U », 1997 (1^{ère} éd. : 1978), 410 p.
- REY-DEBOVE Josette, *La Linguistique du signe. Une approche sémiotique du langage*, Paris, Armand Colin, coll. « U », 1998, 303 p.
- REY-DEBOVE Josette, « Nom propre, lexique et dictionnaire de langue », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 107-122.
- RIEGEL Martin, « Ces noms dits "composés". Arguments et critères », dans *Studia Romanica Posnaniensia*, vol. 16, Poznan, Pologne, Adam Mickiewicz University Press, 1991, p. 149-161.
- RIEGEL Martin, PELLAT Jean-Christophe, RIOUL René, *Grammaire méthodique du français*, Paris, PUF, 1996 (2^{ème} éd.), 646 p.
- SARFATI Georges-Elia, « Le statut lexicographique du nom propre : remarques méthodologiques et linguistiques », in *Mots* n° 63, Lyon, ENS Éditions, 2000, p. 105-124.

- SAUSSURE Ferdinand de, *Cours de linguistique générale*, édition critique préparée par Tullio de Mauro, postface de Louis-Jean Calvet, Paris, Éditions Payot & Rivages, 1995, (1^{ère} éd. : 1916), 520 p.
- SCHNEDECKER Catherine, *Nom propre et chaînes de référence*, coll. « Recherches linguistiques », n° 21, Université de Metz - Klincksieck, Paris, 1997, 231 p.
- SEARLE John R., « Proper names », in *MIND*, vol. VXXVII, Edinburgh, T. Nelson & Sons Ltd., 1958, p. 166-173.
- SEARLE John R., *Les Actes de langage – Essai de philosophie du langage*, préf. d'Oswald Ducrot, Paris, Hermann, 1972, 260 p.
- SEARLE John R., *Sens et expression – Étude de théorie des actes de langage*, traduction et préface de Joëlle Proust, Paris, Les Éditions de Minuit, 1982 (1^{ère} éd. : 1979), 242 p.
- SEARLE John R., *L'Intentionnalité – Essai de philosophie des états mentaux*, trad. de l'américain par Claude Pichevin, Paris, Les Éditions de Minuit, 1985 (1^{ère} éd. : 1983), 340 p.
- SIBLOT Paul, « De l'actualisation du nom propre », in *Modèles linguistiques*, XV, n° 2, Lille, ENSAM-URA 1030 CNRS, 1994, p. 69-84.
- SIBLOT Paul, « Choisir un nom », in Béatrice Fraenkel et Christiane Legris-Desportes (dir.), *Entreprise et Sémiologie*, Paris, DUNOD, 1999, p. 39-46.
- SIBLOT Paul, « De la dénomination à la nomination – Les dynamiques de la signifiante nominale et le propre du nom », *Cahiers de praxématique*, n° 36, Montpellier, PUM3, 2001, p. 189-214.
- SIBLOT Paul, « D'un nom l'autre », in *Nouvelle revue d'onomastique*, n° 29-30, Paris, Société française d'onomastique, 1997, p. 3-18.
- SIBLOT Paul, « De la signifiante du nom propre », in *Cahiers de praxématique*, n° 8, Montpellier, Université Paul Valéry, 1987, p. 97-114.
- SIBLOT Paul, « Noms et images de marque : de la construction du sens dans les noms propres », in M. Noailly (éd.), *Nom propre et nomination*, Actes du colloque de Brest, 21-24 avril 1994, Paris, Klincksieck, 1995, p. 147-160.
- SIMONS Peter, *Parts – A Study in Ontology*, Oxford, Clarendon Press, 1987, XIII – 390 p.
- SOWA John F., « Logical Structures in the Lexicon » in J. Pustejovsky et S. Bergler (éd.), *Lexical Semantics and Knowledge Representation*, Berlin Heidelberg New York, Springer-Verlag, First SIGLEX Workshop Berkley, CA, USA, June 17, 1991, Proceedings, 1992, p. 39-60.
- STRAWSON Peter Frederick, *Les Individus*, trad. de l'anglais par A. Shalom et Paul Drong, Paris, Éditions du Seuil, 1973, (1959 pour l'édition anglaise), 282 p.
- STRAWSON Peter Frederick, *Subject and Predicate in Logic and Grammar*, London, Methuen & Co Ltd, 1974, 144 p.
- SUMPF Joseph, « Noms – Noms propres », *Langages*, n° 76, Paris, Larousse, 1984, p. 113-128.
- TAMBA Irène, « Un puzzle sémantique : le couplage des relations de tout à partie et de partie à tout », in *Le Gré des langues*, n° 7, Paris, L'Harmattan, 1994, p.
- TAMBA-MECZ Irène, *La Sémantique*, Paris, PUF, coll. « Que sais-je ? », 1998, 4^e édition corrigée, (1^{ère} éd. : 1988), 128 p.
- VAN LANGENDONCK W., « Neurolinguistic and syntactic evidence for basic level meaning in proper names », in *Functions of Language*, 6.1, Amsterdam, John Benjamins B.V., 1999, p. 95-138.
- VAN DE VELDE Danièle, « Existe-t-il des noms propres de lieux ? », in *Lexique* n° 15, Lille, Presses universitaires du Septentrion, 2000, p. 35-45.

- VAN LANGEDONCK Willy, « La théorie du nom propre et la neurolinguistique », in *Nouvelle Revue d'onomastique*, n° 35-36, Société française d'onomastique, Paris, 2000, p. 13-24.
- VAXELAIRE Jean-Louis, « Le nom propre en contexte – une approche lexicologique », in *Mots, termes et contextes – Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005*, CPI – Éditions des archives contemporaines, Paris, 2007, p. 591-597.
- VAXELAIRE Jean-Louis, *Pour une lexicologie du nom propre*, thèse préparée sous la dir. de John Humbley, Université Paris 7, 2001, 601 p.
- VÉDÉNINA L. G., *Pertinence linguistique de la représentation typographique*, Paris, Peeters / Selaf, 1989, 153 p.
- VENDRYÈS Joseph, *Le Langage – Introduction linguistique à l'histoire*, Paris, Albin Michel, 1968 (1^{ère} éd. : 1923), 444 p.
- VENDRYÈS Joseph, « Marcel Proust et les noms propres », in *Choix d'études linguistiques et celtiques*, Paris, Klincksieck, 1952, p. 80-88.
- WAGNER Robert Léon, *Les Vocabulaires français I Définitions – Les Dictionnaires*, Paris, Bruxelles, Montréal, Didier, 1967, 189 p.
- WAGNER Robert Léon, *Les Vocabulaires français II Les Tâches de la lexicologie synchronique – Glossaires et dépouillements – analyse lexicale*, Paris, Bruxelles, Montréal, Didier, 1970, 183 p.
- WARTBURG Walter von, ULLMANN Stephen, *Problèmes et méthodes de la linguistique*, trad. de l'allemand par P. Maillard, Paris, PUF, 1969, 350 p.
- WILMET Marc, « Nom propre et ambiguïté », in *Langue française* n° 92, Paris, Larousse, 1991, p. 113-124.
- WILMET Marc, « Pour en finir avec le nom propre ? », in *L'Information grammaticale*, n°65, Paris, Editions Baillière, 1995, p. 3-11.
- WILMET Marc, *Grammaire critique du français*, Paris, Hachette – Duculot, 2^{ème} éd. (1^{ère} éd. : 1997), 1998, 704 p.
- WITTGENSTEIN Ludwig, *Tractatus logico-philosophicus – suivi de Investigations philosophiques*, traduit de l'allemand par Pierre Klossowski, Paris, Gallimard, 1961 (1986 coll. « Tel »), (1^{ères} éd. : 1918 pour le *Tractatus* et 1945 pour les *Investigations*), 364 p.
- WITTGENSTEIN Ludwig, *Remarques philosophiques*, édition posthume de Rush Rhees, traduit de l'allemand par Jacques Fauve, Paris, Gallimard, 1975 (1996 coll. « Tel »), (1^{ère} éd.: 1964), 330 p.
- ZABEEH Farhang, *What is in a name ? – An inquiry into the semantics and pragmatics of proper names*, The Hague (Netherlands), Martinus Nijhoff, 1968, 78 p.
- ZONABEND Françoise, « Pourquoi nommer (Les noms de personnes dans un village français : Minot-en-Châtillonnais) », in *L'Identité – Séminaire interdisciplinaire dirigé par C. Lévi-Strauss, professeur au Collège de France 1974-1975*, Paris, PUF, 2000 (1^{ère} éd. : 1977), p. 256-286.

Textométrie et traitement automatique du langage naturel

- ALLARD Michel, ELZIERE May, GARDIN Jean-Claude, HOURS Francis, *Analyse conceptuelle du Coran sur cartes perforées*, Paris La Haye, Mouton & CO, 1963, vol.1 (code), 110 p., vol.2 (commentaire), 187 p.

- ASSADI Houssem, BOURIGAULT Didier, « Analyses syntaxique et statistique pour la construction d'ontologies à partir de textes », in *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, « Collection Technique et scientifique des télécommunications », 2000, p. 243-255.
- AUSSENAC-GILLES N., CONDAMINES Anne, « Base de connaissances terminologiques : enjeux pour la consultation documentaire », in J. Maniez et W. Mustafa El Hadi (éd.), *Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information*, Université Charles-de-Gaulle, Lille III, 1999.
- BACHIMONT Bruno, « Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances », in *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, « Collection technique et scientifique des télécommunications », 2000, p. 305-323.
- BELLEIL Claude, *Reconnaissance, typage et traitement des coréférences des toponymes français et de leurs gentils par dictionnaire électronique relationnel*, thèse préparée sous la dir. de Denis Maurel, Université de Nantes, 1997, 135 p.
- BELLOT Patrice, EL-BÈZE Marc, « Classification locale non supervisée pour la recherche documentaire », in *TAL*, vol. 41, n°2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 335-365.
- BIBER Douglas, « Co-occurrence Patterns among Collocations : a Tool for Corpus-Based Lexical Knowledge Acquisition », *Computational Linguistics*, 1993, vol. 19, n° 3.
- BIBER Douglas, « Using Register-Diversified Corpora for General Language Studies », in *Computational Linguistics*, vol. 19, n° 2, ACL, 1993, p. 220-241.
- BIKEL Daniel M., MILLER Scott, SCHWARTZ Richard, WEISCHEDEL Ralph, « Nymble : a High-Performance Learning Name-finder », in *Proceedings, 5th Conference on Applied Natural Language Processing (ANLP 97)*, Washington, ACL, 1997, p. 194-201.
- BILGER Mireille (coord. par), *Linguistique sur corpus – Études et réflexions, Cahiers de l'Université de Perpignan n° 31*, Perpignan, Presses universitaires de Perpignan, 2000, 238 p.
- BODENREIDER Olivier, ZWEIGENBAUM Pierre, « Stratégies d'identification des noms propres à partir de nomenclatures médicales parallèles », in *TAL*, vol. 41, n° 3, sous la dir. de Denis Maurel et Franz Guenther, ATALA / Hermès Science Publications, Paris, 2000, p. 727-757.
- BOGURAEV Branimir, PUSTEJOVSKY James (éd.), *Corpus Processing for Lexical Acquisition*, Cambridge (Massachusetts) – London, The MIT Press, 1996, 245 p.
- BOMMIER-PINCEMIN Bénédicte, *Diffusion ciblée automatique d'information : conception et mise en œuvre d'une linguistique textuelle pour la caractérisation des destinataires et des documents*, thèse préparée sous la direction de François Rastier, Université Paris 4, 1999, 806 p.
- BONNAFOUS Simone, TOURNIER Maurice, « Analyse de discours, lexicométrie, communication et politique », in *Langages* n° 117, Paris, Larousse, 1995, p. 67-81.
- BORGMAN Christine L., SIEGFRIED Susan L., « Getty's Synoname and its Cousins : A Survey of Applications of Personal Name-Matching Algorithms », in *Journal of the American Society for Information Science*, vol. 43 n° 7, New York, John Wiley & Sons, Inc., 1992, p. 459-476.
- BOURIGAULT Didier, « LEXTER, a Natural Language Processing tool for terminology extraction », *Proceedings of the 7th EURALEX International Congress*, Goteborg, 1996.

- BOUAUD Jacques *et al.*, « Regroupements issus de dépendances syntaxiques sur un corpus de spécialité : catégorisation et confrontation à deux conceptualisations du domaine », in *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, « Collection technique et scientifique des télécommunications », 2000, p. 275-290.
- BOUILLON Pierrette, FABRE Cécile, SÉBILLOT Pascale, JACQMIN Laurence, « Apprentissage de ressources lexicales pour l'extension des requêtes », in *TAL*, vol. 41, n° 2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 367-393.
- BOURIGAULT Didier, JACQUEMIN Christian, « Construction de ressources terminologiques », in Jean-Marie Pierrel (dir.), *Ingénierie des langues*, Paris, Hermès Science Publications, 2000, p. 215-230.
- BRUNET Étienne, *Hyperbase. Logiciel documentaire et statistique pour la création et l'exploitation de bases hypertextuelles, Manuel de référence, version 5.2*, CNRS, Université de Nice, 1999, 2001 (pour la mise à jour).
- BRUNET Étienne, « Le lemme comme on l'aime », in A. Morin, P. Sébillot (éd.), *Actes des 6^{es} Journées internationales d'analyse des données textuelles, Saint-Malo, 13-15 mars 2002*, 2 vol., p. 221-232.
- BRUNET Étienne, *Le Vocabulaire français de 1789 à nos jours – d'après les données du Trésor de la langue française*, tome I, Genève, Paris, Slatkine, Champion, 1981, 852 p.
- BRUNET Étienne, *Le Vocabulaire de Proust*, Genève, Slatkine, 1983, 3 vol. VI-261 + 1644 p.
- BRUNET Étienne, *Le Vocabulaire de Victor Hugo*, Genève, Slatkine, Paris, Champion, 1988, 3 vol. (V-484, 637, 531 p.).
- BRUNET Étienne, *Le Vocabulaire de Zola*, Genève, Slatkine, Paris, Champion, 1985, 3 vol. (472, 646, 357 p.).
- CAPPONI Nicolas, TOUSSAINT Yannick, « Interprétation de classes de termes par généralisation de structures prédicats-arguments », in *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, « Collection technique et scientifique des télécommunications », 2000, p. 337-353.
- CHARTRON Ghislaine, *Analyse des corpus de données textuelles, sondage de flux d'information*, thèse préparée sous la dir. de Daniel Laurent, Université Paris 7, 1988, 273 p.
- CHARLET Jean, ZACKLAD Manuel, KASSEL Gilles, BOURIGAULT Didier, *Ingénierie des connaissances – Évolutions récentes et nouveaux défis*, Paris, Eyrolles, 2000, 610 p.
- CHETOUANI Lamria, HEIDEN Serge, « Sémantique des noms propres – Méthode des cooccurrences », in M. Rajman, M. Decrauzat, J.-C. Chappelier, *Actes des 5^{es} Journées internationales d'analyse des données textuelles, 9-11 mars 2000*, Lausanne, 2000.
- CHURCH K. W., HANKS P., « Word association norms, mutual information, and lexicography », *Computational Linguistics*, 16, 1, 1990, p. 22-29.
- COATES – STEPHENS Sam, « The Analysis and Acquisition of Proper Names for The Understanding of Free Text », in *Computers and the Humanities*, vol. 26, n° 5 – 6, Dordrecht – Boston – London, Kluwer Academic Publishers, 1992, p. 441-456.
- COATES-STEPHENS Sam, *The Analysis and Acquisition of Proper Names for Robust Text Understanding*, Thèse, City University, London, 1992, 290 p.

- CONDAMINES Anne et REBEYROLLE Josette, « Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode », dans *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, « Collection technique et scientifique des télécommunications », 2000, p. 225-241.
- DAILLE Béatrice, *Approche mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistiques*, thèse préparée sous la direction de Laurence Danlos, Université Paris 7, 1994, 228 p.
- DAILLE Béatrice, « Reconnaissance automatique des noms propres de la langue écrite : les récentes réalisations », in *TAL*, vol. 41, n° 3, sous la dir. de Denis Maurel et Franz Guenther, ATALA / Hermès Science Publications, Paris, 2000, p. 602-621.
- DAILLE Béatrice, « Repérage et extraction de terminologie par une approche mixte statistique et linguistique », *T.A.L.*, n° 36, n° 1-2, Paris, Association pour le traitement automatique des langues, 1995, p. 101-118.
- DAL Georgette, NAMER Fiammetta, « Génération et analyse automatique de ressources lexicales construites utilisables en recherche d'informations », in *TAL*, vol. 41, n° 2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 423-446.
- DAOUST François, *SATO (système d'analyse de texte par ordinateur) version 4.0 Manuel de référence*, ATO, Université du Québec à Montréal, 1996, 212 p.
- DAVID Sophie, PLANTE Pierre, « De la nécessité d'une approche morpho-syntaxique dans l'analyse de textes », Montréal, *ICO*, vol. 2, n° 3, 1990, p. 140-155.
- DAVID Sophie, SOUCHARD Maryse, « Analyse de discours et traitement automatique de données textuelles : le logiciel TERMINO », in T. Bulot, D. Dubois, C. Batime (éd.), *Cahiers de linguistique sociale. Recherches documentaires*, Actes des deuxièmes Journées Linguistique et ergonomie, Université de Rouen, 17-18 nov. 1994, p 61-73.
- DESGRAUPES Bernard, *Introduction aux expressions régulières*, Paris, Vuibert, 2001, 248 p.
- DIAS Gaël, *et al.*, « Extraction automatique d'unités lexicales complexes : un enjeu fondamental pour la recherche documentaire », in *TAL*, vol. 41, n°2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 447-472.
- DROUIN Patrick, *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*, Thèse, Université de Montréal, 2002, 289 p.
- DROUIN Patrick, « Spécificités lexicales et acquisition de la terminologie », in *Actes des 7^{es} Journées internationales d'Analyse des Données Textuelles*, Louvain (Belgique), UCL Presses universitaires de Louvain, 2004, p. 345-352.
- ENGUEHARD C., PANTERA L., « Automatic natural acquisition of a terminology », *Journal of Quantitative Linguistics*, vol. 2, n° 1, 1995, p. 27-32.
- ERLOS Frédéric, « Entités nommées et domaines d'activité dans les discours de communication institutionnelle », in *Actes des 8^{èmes} Journées internationales d'analyse des données textuelles*, Besançon, Presses universitaires de Franche-Comté, 2006, p. 401-414.
- FIALA Pierre, HABERT Benoît, LAFON Pierre, PINEIRA Carmen, « Des mots aux syntagmes. Figements et variations dans la résolution générale du congrès 1978 de la CGT », in *MOTS*, n°14, Paris, Presses de la FNSP, 1987, p. 47-87.
- FIALA Pierre, « L'interprétation en lexicométrie. Une approche quantitative des données lexicales. », in *Langue française*, n° 103, Paris, Larousse, 1994, p. 113-122.

- FRIBURGER Nathalie, MAUREL Denis, «Élaboration d'une cascade de transducteurs pour l'extraction des noms de personnes dans les textes », in *8^e Conférence annuelle sur le traitement automatique des langues naturelles*, 2-5 juill. 2001, sous l'égide de l'ATALA, org. par le Laboratoire d'informatique et l'axe Langues et technologies de l'information de l'Université de Tours, 2001, 508 p.
- GREFENSTETTE G., «Automatic thesaurus generation from raw text using knowledge-poor techniques», in *Proceedings of the 9th Conference on Oxford English dictionary*, Oxford, 1993.
- GREFENSTETTE G., «Evaluation techniques for automatic semantic extraction : Comparing syntactic and window based approaches », in Boguraev B., Pustejovsky J. (éd), *Corpus processing for Lexical Acquisition*, The MIT Press, Cambridge, Massachusetts, 1996, p. 205-216.
- GREFENSTETTE G., *Explorations in Automatic Thesaurus Discovery*, Kluwer, Dordrecht, 1994, XII-305 p.
- GREFENSTETTE Gregory, GAUSSIÉ Eric, HULL David, ROUX Claude, « Recherche d'information en français et traitement automatique des langues », in *TAL*, vol. 41, n°2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 473-493.
- GUÉRY Louis, *Dictionnaire des règles typographiques*, Paris, CFPJ, 2000, 2^{ème} éd., 282 p.
- GUIRAUD Pierre, *Les Caractères statistiques du vocabulaire – Essai de méthodologie*, Paris, PUF, 1954, 116 p.
- GUIRAUD Pierre, *Index du vocabulaire du symbolisme – I – Index des mots d'Alcools de Guillaume Apollinaire*, avec un avant-propos de R.-L. Wagner, Paris, Klincksieck, 1953, VIII-29 p.
- GUIRAUD Pierre, *Problèmes et méthodes de la statistique linguistique*, Paris, PUF, 1960, 145 p.
- HABERT Benoît, « L'analyse des formes 'spécifiques', bilan critique et propositions », *Mots*, n°11, Paris, Presses de la FNSP, 1985, p. 127-154.
- HABERT Benoît, SALEM André, « L'utilisation de catégorisations multiples pour l'analyse quantitative de données textuelles », dans *T.A.L.*, vol. 36, n° 1-2, Paris, Association pour le traitement automatique des langues, 1995, p. 249-275.
- HABERT Benoît, NAZARENKO Adeline, SALEM André, *Les linguistiques de corpus*, Paris, Armand Colin / Masson, coll. « U », 1997, 240 p.
- HABERT Benoît, « Des corpus représentatifs: de quoi, pour quoi, comment ? », *Linguistiques sur corpus. Études et réflexions, Cahiers de l'Université de Perpignan*, n°31, Presses universitaires de Perpignan, Perpignan, 2000, p. 11-58.
- HEIDEN Serge, LAFON Pierre, « Cooccurrences – La CFDT de 1973 à 1992 », in P. Fiala, P. Lafon (éd.), *Des mots en liberté – Mélanges Maurice Tournier*, Fontenay-aux-Roses, ENS éditions, 1998, Tome 1, p. 65-83.
- HEIDEN Serge, *Weblex version 4.1, Manuel Utilisateur*, Lyon, UMR 8503, CNRS, 2002, 180 p.
- ILLOUZ G. et al., « Maîtriser les déluges de données hétérogènes », in P. Amsili (éd.), *6^e Conférence annuelle sur le traitement automatique des langues naturelles*, Cargèse 12-17 juillet 1999, p. 37-46.
- JACQUEMIN Christian, TZOUKERMANN Évelyne, « NLP for term variant extraction : synergy between morphology, lexicon, and syntax », in *Natural Language Information Retrieval*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1999, p. 25-74.

- JACQUEMIN Christian, *Variation terminologique : reconnaissance et acquisition automatique de termes et de leurs variantes en corpus*, Mémoire d'habilitation à diriger des recherches en informatique fondamentale, Université de Nantes, 1997.
- JUILLARD Michel, LUONG Xuan, « De nouveaux arbres pour un nouveau corpus », in *Revue informatique et statistique dans les sciences humaines*, (30) n° 1 à 4, Université de Liège, Liège, 1994, p. 37-51.
- KIETZ Jörg-Uwe, VOLZ Raphael, MAEDCHE Alexander, « Extracting a Domain-Specific Ontology from a Corporate Intranet », in *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, 2000, p. 167-175.
- LABBÉ Dominique, « La France chez de Gaulle et Mitterrand », in P. Fiala, P. Lafon (éd.), *Des mots en liberté – Mélanges Maurice Tournier*, Fontenay-aux-Roses, ENS éditions, 1998, Tome 1, p. 183-193.
- LABBÉ Dominique, HUBERT Pierre, « Un modèle de partition du vocabulaire », in Dominique Labbé, Philippe Thoiron, Daniel Serant (éd.), *Études sur la richesse et la structure lexicales*, Paris-Genève, Slatkine-Champion, 1988, p. 92-114.
- LABBÉ Dominique, *Normes de saisie et de dépouillement des textes politiques*, Cahiers du C.E.R.A.T. n° 7, Saint Martin d'Hères, CERAT, 1990, 135 p.
- LAFON Pierre, « Sur la variabilité des formes dans un corpus », in *Mots* n° 1, Paris, Presses de la FNSP, 1980, p. 127-165.
- LAFON Pierre, « Analyse lexicométrique et recherche des cooccurrences », in *MOTS*, n° 3, Paris, Presses de la FNSP, 1981, p. 95-148.
- LAFON Pierre, *Dépouillements et statistiques en lexicométrie*, Préf. de Charles Muller, Genève - Paris, Slatkine - Champion, 1984, 217 p.
- LAFON Pierre, SALEM André, « L'inventaire des segments répétés d'un texte », *Mots*, n° 6, Paris, Presses de la FNSP, 1983, p. 161-177.
- LAFON Pierre, LEFÈVRE Josette, SALEM André, TOURNIER Maurice, *Le Machinal – Principes d'enregistrement informatique des textes*, Paris, INaLF - Klincksieck, coll. « Saint-Cloud », 1985, 61 p.
- LAFON Pierre, LEFÈVRE Josette, « Le figement : prise en compte discursive, incidences sur les statistiques textuelles et sur l'interprétation », in *La Locution : entre lexique, syntaxe et pragmatique – identification en corpus, traitement, apprentissage*, Paris, Publication de l'Inalf, coll. « Saint-Cloud », Klincksieck, 1997, p. 295-306.
- LAMALLE Cédric, SALEM André, « Types généralisés et topographie textuelle dans l'analyse quantitative des corpus textuels », in A. Morin, P. Sébillot (éd.), *Actes des 6^{es} Journées internationales d'analyse des données textuelles, Saint-Malo, 13-15 mars 2002*, 2 vol., p. 403-411.
- LEBART Ludovic et SALEM André, *Statistique textuelle*, préface de Christian Baudelot, Paris, Dunod, 1994, 342 p.
- LEBART Ludovic, SALEM André, BERRY Lisette, *Exploring textual data*, Dordrecht (The Netherlands), Kluwer Academic Publishers, 1998, 245 p.
- LOUPY Claude de, « L'apport de connaissances linguistiques en recherche documentaire », in *8^e Conférence annuelle sur le traitement automatique des langues naturelles, 2-5 juill. 2001*, sous l'égide de l'ATALA, org. par le Laboratoire d'informatique et l'axe Langues et technologies de l'information de l'Université de Tours, 2001.
- MAC DONALD David D., « Internal and External Evidence in the Identification and Semantic Categorization of Proper Names », in Branimir Boguraev and James Pustejovsky (éd.), *Corpus Processing for Lexical Acquisition*, Cambridge – London, MIT Press, 1996, p. 21-39.

- MAISONDIEU Aude, KUNCOVA Andréa, *Manuel d'utilisation de LEXICO 3*, ILPGA, Université Paris 3 – Sorbonne nouvelle, 2000, 49 p.
- MANI Inderjeet, MAC MILLAN T. Richard, « Identifying Unknown Proper Names in Newswire Text », in Branimir Boguraev and James Pustejovsky (éd.), *Corpus Processing for Lexical Acquisition*, (Edited by), Cambridge – London, MIT Press, 1996, p. 41-59.
- MARTINEZ William, *Contribution à une méthodologie de l'analyse des cooccurrences lexicales multiples dans les corpus textuels*, Thèse préparée sous la dir. d'André Salem, Université Paris 3 – Sorbonne nouvelle, 2003, 468 p.
- MARTINEZ William, « Mise en évidence de rapports synonymiques par la méthode des cooccurrences », in M. Rajman, M. Decrauzat, J.-C. Chappelier (éd.), *Actes des 5^{es} Journées internationales d'analyse des données textuelles, 9-11 mars 2000*, Lausanne, 2000.
- MATHIEU Yvette Yannick, GROSS Gaston, FOUQUERÉ Christophe, « Vers une extraction automatique des néologismes », in *Cahiers de lexicologie*, vol. LXXII, n° II, 1998, p. 199-208.
- MATHIEU-COLAS Michel, « Les noms de divinités : web, contextes et classes d'objet », in *Mots, termes et contextes – Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005*, CPI – Éditions des archives contemporaines, Paris, 2007, p. 391-408.
- MAUREL Denis, BELLEIL Claude, PITON Odile, EGGERT Elmar, « Un dictionnaire électronique relationnel de noms propres », *Revue française de linguistique appliquée*, vol. II-1, Amsterdam, Editions 'De Werelt', 1997, p. 101-111.
- MAUREL Denis, GUENTHNER Franz (dir.), *Traitement automatique des langues – Traitement automatique des noms propres*, vol. 41, n°3, Paris, ATALA / Hermès Science Publications, 2001, p.600-836.
- MAUREL Denis, PITON Odile, EGGERT Elmar, « Les relations entre noms propres : lieux et habitants dans le projet Prolex », in *TAL*, vol. 41, n° 3, sous la dir. de Denis Maurel et Franz Guentner, ATALA / Hermès Science Publications, Paris, 2000, p. 623-641.
- MAUREL Denis, « Les mots inconnus sont-ils des noms propres? », in Luc Collès *et alii* (dir.), *Actes des 7^{es} Journées internationales d'analyse statistique des données textuelles*, Louvain-la-Neuve (Belgique), UCL Presses universitaires de Louvain, vol. 2, p. 776-784.
- MAYAFFRE Damon, *Le Poids des mots – Le Discours de gauche et de droite dans l'entre-deux-guerres – Maurice Thorez, Léon Blum, Pierre-Etienne Flandin et André Tardieu (1928-1939)*, Paris, Honoré Champion, 2000, 798 p.
- MIKHEEV Andrei, MOENS Marc, GROVER Claire, « Named Entity Recognition without Gazetteers », in *Proceedings, 9th Conference of the European Chapter of the Association for Computational Linguistics (EACL 99)*, Bergen, ACL, 1999, p. 1-8.
- MIKHEEV Andrei, « Periods, Capitalized Words, etc. », in *Computational Linguistics*, vol. 28, n° 3, MIT Press, 2002, p. 289-318.
- MORTUREUX Marie-Françoise, « Lexique, vocabulaire, comptage », in P. Fiala, P. Lafon (éd.), *Des mots en liberté – Mélanges Maurice Tournier*, Fontenay-aux-Roses, ENS éditions, 1998, Tome 1, p. 257-266.
- MULLER Charles, *Essai de statistique lexicale – L'illusion comique de Pierre Corneille*, Paris, Klincksieck, 1964, 202 p.
- MULLER Charles, *Étude de statistique lexicale – Le Vocabulaire du théâtre de Pierre Corneille*, (thèse publiée), Paris, Larousse, 1967, 379 p.
- MULLER Charles, *Initiation à la statistique linguistique*, Paris, Larousse, 1968, 247 p.

- MULLER Charles, *Initiation aux méthodes de la statistique linguistique*, Paris, Champion, 1992 (1^{ère} éd. : 1973), 185 p.
- MULLER Charles, *Principes et méthodes de statistique lexicale*, Paris, Champion, 1992 (1^{ère} éd. : 1977), 205 p.
- Outils linguistiques et nouvelles technologies*, Journée d'étude organisée par l'ADBS le 19 décembre 1996 à la Vidéothèque de Paris, ADBS, 1996, 14 p.
- PAIK Woojin, LIDDY Elizabeth D., YU Edmund, MAC KENNA Mary, « Categorizing and Standardizing Proper Nouns for Efficient Information Retrieval », in Branimir Boguraev and James Pustejovsky (éd.), *Corpus Processing for Lexical Acquisition*, Cambridge – London, MIT Press, 1996, p. 61-73.
- PÉRY-WOODLEY Marie-Paule, « Quels corpus pour quels traitements automatiques ? », *T.A.L.*, vol. 36, n° 1-2, Paris, Association pour le traitement automatique des langues, 1995, p. 213-232.
- PIERREL Jean-Marie (éd.), *Ingénierie des langues*, Paris, Hermès, 2000, 354 p.
- POIBEAU Thierry, *Extraction automatique d'information – du texte brut au web sémantique*, Paris, Hermès - Lavoisier, 2003, 239 p.
- POIBEAU Thierry, « Repérage des entités nommées : un enjeu pour les systèmes de veille », in *Actes des rencontres TIA'99*, Nantes, Terminologies nouvelles, 1999, p. 43-51.
- PROST Antoine, *Vocabulaire des proclamations électorales de 1881, 1885, et 1889*, Paris, PUF, 1974, 196 p.
- RAJMAN Martin, BESANÇON Romaric, CHAPPELIER Jean-Cédric, « Le modèle DSIR : une approche à base de sémantique distributionnelle pour la recherche documentaire », in *TAL*, vol. 41, n° 2, sous la dir. de Christian Jacquemin, ATALA / Hermès Science Publications, Paris, 2000, p. 549-578.
- RASTIER François, CAVAZZA Marc, ABEILLÉ Anne, *Sémantique pour l'analyse – De la linguistique à l'informatique*, Paris, Masson, coll. « Sciences cognitives », 1994, 240 p.
- RAVIN Yael, WACHOLDER Nina, *IBM Research Report – Extracting Names from Natural-Language Text*, Almaden – T.J. Watson – Tokyo – Zurich, IBM Research Division, 1997, VIII – 31 p.
- ROUAULT Jacques, *Linguistique automatique – Applications documentaires*, Berne, Francfort-sur-le-Main, New York, Paris, 1987, 309 p.
- SALEM André, « Analyse factorielle et lexicométrie – synthèse de quelques expériences », *Mots*, n°4, Paris, Presses de la FNSP, 1982, p. 147-168.
- SALEM André, « Approches du temps lexical – Statistique textuelle et séries chronologiques », *Mots*, n° 17, Paris, Presses de la FNSP, 1988, p. 105-143.
- SALEM André, « De *travailleurs* à *salariés* – Repères pour une étude de l'évolution du vocabulaire syndical (1971 – 1990) », *Mots*, n° 36, Paris, Presses de la FNSP, 1993, p. 74-83.
- SALEM André, *Méthodes de la statistique textuelle*, Thèse pour le doctorat d'État, sous la direction de M. Tournier, Université Paris 3 – Sorbonne nouvelle, 1993, vol. 1 et 2, 819 p. et vol. 3, annexes, 179 p.
- SALEM André, *Pratique des segments répétés – Essai de statistique textuelle*, Klincksieck, Publications de l'INALF, coll. « Saint-Cloud », Paris, 1987, 333 p.
- SILBERZTEIN Max, *Dictionnaires électroniques et analyse automatique de textes, Le système INTEX*, Paris, Masson, coll. Informatique linguistique, 1993, 233 p.

- SINCLAIR John, *Corpus – Concordance – Collocation*, Oxford, Oxford University Press, 1991, 179 p.
- SMADJA F., « Retrieving collocations from text : Xtract », *Computational Linguistics*, vol. 19, n° 1, 1993, p. 143-177.
- THOMPSON Paul, DOZIER Christopher, « Name recognition retrieval performance », in *Natural Language Information Retrieval*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1999, p. 261-272.
- TOURNIER Maurice, « Cooccurrence autour de ‘travail’ », *Mots*, n°14, Paris, Presses de la FNSP, Paris, 1987, p. 89-123.
- TOURNIER Maurice, « Les discours sociopolitiques et l’analyse lexicométrique », in *Sociolinguistique – Territoires et objets*, Lausanne – Paris, Delachaux et Niestlé, 1996, p. 179-213.
- TOURNIER Maurice, « D’où viennent les fréquences de vocabulaire ? – La lexicométrie et ses modèles », *Mots*, n°1, Paris, 1980, p. 189-209.
- TOURNIER Maurice, MOUILLAUD Maurice, LAFON Pierre *et al.*, *Des Tracts en mai 68*, Paris, Armand Colin, 1975, 487 p.
- WACHOLDER Nina, RAVIN Yael, CHOI Misook, « Disambiguation of Proper Names in Text », in *Proceedings, 5th Conference on Applied Natural Language Processing (ANLP 97)*, Washington, ACL, 1997, p. 202-208.
- ZIMINA-POIROT Maria, « Topographie bitextuelle et approches quantitatives de l’indexation de ressources traductionnelles à partir de corpus parallèles », in *Mots, termes et contextes – Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005*, CPI – Éditions des archives contemporaines, Paris, 2007, p. 175-186.
- ZINGLÉ Henri, *Conception et réalisation d’un environnement de génie linguistique*, thèse préparée sous la dir. de Christophe Roche, Université de Chambéry, 1998, 137 p.
- ZWEIGENBAUM P., « Encoder l’information médicale : des terminologies aux systèmes de représentation des connaissances », *ISIS*, n° 2, 1999.

Sciences de l’information

- Association Française de Normalisation, *Règles d’établissement des thesaurus en langue française. NF Z 47-100*, Paris, AFNOR, 1981, 20 p.
- AITCHISON Jean, GILCHRIST Alan, *Construire un thesaurus. Manuel pratique*, traduit de l’anglais par Dominique HERVIEU, Paris, ADBS Éditions, 1992, 217 p.
- AMAR Esther, *Internet-intranet – Les Concepts de base*, Paris, AFNOR, 1997, vol. 1, 47 p.
- AMAR Esther, *Internet-intranet pour l’entreprise*, Paris, AFNOR, 1997, vol. 2, 47 p.
- AMAR Muriel, « « Indexation discursive » versus « indexation lexicale », in *Terminologies nouvelles*, n°21, Agence de la francophonie – Communauté française de Belgique, Bruxelles, 2000, p. 71-79.
- AMAR Muriel, *Les Fondements théoriques de l’indexation. Une approche linguistique*, préface de Michel Le Guern, Paris, ADBS Éditions, 2000, 355 p.
- AMAR Muriel, « Quelques remarques sur : Une indexation classique est-elle viable en philosophie ? », in *Documentation et philosophie II – Á propos de l’indexation discursive – Autour des travaux de Muriel Amar*, textes réunis et présentés par Benoît Hufschmitt *et al.*, Besançon, Presses universitaires franc-comtoises, 2003, p. 71-83.

- ANDRÉ J., « Balises, structures et tei », *Cahiers Gutenberg*, n°24, [], 1996, p. 11-22.
- BELKIN Nicholas J., « Anomalous states of knowledge as a basis for information retrieval », in *Canadian Journal of Information Science*, n°5, 1980
- BELKIN Nicholas J., ODDY R.N., BROOKS H.M., « ASK for information retrieval : Part 1. Background and theory », in *Journal of Documentation*, vol. 38, n°2, Aslib, London, 1982, p. 61-71.
- BELKIN Nicholas J., ODDY R.N., BROOKS H.M., « ASK for information retrieval : Part 2. Results of a design Study », in *Journal of Documentation*, vol. 38, n°2, Aslib, London, 1982, p. 145-164.
- BELKIN Nicholas J. *et al.*, « Cases, Scripts, and Information-Seeking Strategies : On the Design of interactive Information Retrieval Systems », in *Expert Systems with Applications*, vol. 9 n°3, 1995, p. 379-395.
- BELKIN Nicholas J., « Interaction with Texts : Information Retrieval as Information-Seeking Behavior », in *Information retrieval – 10 – von der Modelierung zur Andwertung*, Konstanz, Germany, Universitaetsverlag, 1993, p. 55-66.
- BENNET J.L., « The user interface in interactive systems », in *Annual review of Information Science and Technology*, vol. 7, 1972, p. 159-196.
- BERTRAND-GASTALDY Suzanne, PAGOLA Gracia, « L'élaboration et la gestion d'un vocabulaire de domaine dans le contexte des bases de données textuelles : remises en question et méthodologies », in *Actes du Colloque 'Repérage de l'information textuelle'*, Hydro-Québec et le Ministère des communications du Québec, Montréal, 18 sept. 1991, p. 51-71.
- BERTRAND-GASTALDY Suzanne, PAGOLA Gracia, « L'analyse du contenu textuel en vue de la construction de thésaurus et de l'indexation assistées par ordinateur ; applications possibles avec SATO (système d'analyse de textes par ordinateur) », in *Documentation et bibliothèques*, 38/2, 1992, p. 75-89.
- BITOUZET Christine, FOURNIER Philippe, TEZENAS DU MONTCEL Benoît, *Management et intranet*, Paris, Hermès, 1997, 155 p.
- BLAIR David C., *Language and Representation in Information Retrieval*, Amsterdam, Elsevier Science, 1990, 335 p.
- BLAIR David, *Wittgenstein, Language and Information : « Back to the Rough Ground ! »*, Dordrecht, Springer, 2006, 358 p.
- BLANQUET Marie-France, *Science de l'information et philosophie*, Paris, ADBS Éditions, 1997, 149 p.
- BOMBARDE Sophie, *Intranet – Pour une communication réussie*, Issy-les-Moulineaux, Éditions Stratégies, 2000, 217 p.
- BROOKES Bertram C., « The foundations of information science – Part I. Philosophical aspects », in *Journal of Information Science*, vol.2, n°3-4, Amsterdam, North Holland Publishing Company, 1980, p. 125-133.
- BROOKES Bertram C., « The foundations of information science – Part III. Quantitative aspects: objective map and subjective landscapes », in *Journal of Information Science*, vol. 2, n° 6, Amsterdam, North Holland Publishing Company, 1980, p. 269-275.
- BROOKES Bertram C., « The foundations of information science – Part IV. Information science : the changing paradigm », in *Journal of Information Science*, vol. 3, n° 1, Amsterdam, North Holland Publishing Company, 1981, p. 3-12.
- BUCKLAND Michael K., « Forme, signification et structure des systèmes de sélection du savoir », in *Deuxième colloque du chapitre français de l'International Society for Knowledge Organization ISKO99*, Lyon, 21-22 oct.1999, 7 p.

- BUCKLAND Michael K., *Information and information systems*, New York, Praeger, 1991, 225 p.
- BUCKLAND Michael K., « Information as thing », in *Journal of the American Society of Information Science*, vol. 42, n° 5, New York, John Wiley & Sons Inc., 1991, p. 351-360.
- BUCKLAND Michael K., « Vocabulary as a central concept in library and information science », in T. Arpanac *et al.* (éd.), *Digital libraries : Interdisciplinary Concepts, Challenges, and Opportunities, Proceedings of the 3rd International Conference on Conceptions of Library and Information Science CoLIS3*, Dubrovnik, Croatie, 23-26 mai 1999, Lokve, Zagreb, p. 3-12.
- BUCKLAND Michael K., « What is a « document » ? », in *Journal of the American Society of Information Science*, vol. 48, n° 9, New York, John Wiley & Sons Inc., 1997, p. 804-809.
- BULOT Thierry, DUBOS Danièle, BATIME Christine (éd.), *Recherches documentaires, Actes des deuxièmes Journées Linguistique et ergonomie*, Université de Rouen, 17-18 novembre 1994, Rouen, ADBS, CNRS, 1995, 144 p.
- CACALY Serge (dir.), *Dictionnaire encyclopédique de l'information et de la documentation*, Paris, Nathan, 1997, 634 p.
- CACALY Serge *et alii* (dir.), *Dictionnaire de l'information*, Paris, Armand Colin, 2004, 274 p.
- CALLON Michel, COURTIAL Jean-Pierre, PENAN Hervé, *La Scientométrie*, Paris, PUF, coll. « Que sais-je ? », n° 2727, 1993, 126 p.
- CAPURRO R., HJØRLAND B., « The concept of Information », in *Annual Review of Information Science and Technology (ARIST)*, vol. 37, Medford, New Jersey (USA), Information Today Inc., 2003, p. 343-411.
- CARPENTER Michael, SVENONIUS Elaine (éd.), *Foundations of Cataloging – A Sourcebook*, Littleton, Colorado, Libraries Unlimited, Inc., 1985, 276 p.
- CHAUDIRON Stéphane (dir.), *Évaluation des systèmes de traitement de l'information*, Hermès, Paris, 2004, 375 p.
- CHAUDIRON Stéphane et IHADJADENE Madjid, « Évaluer des systèmes de recherche d'information : vers de nouveaux modèles informationnels ? », in J. Perriault et V. Paul (éd.), *Critique de la raison numérique*, Hermès, n° 39, CNRS, Paris, 2004, 272 p.
- CHAUDIRON Stéphane, « La place de l'utilisateur dans l'évaluation des systèmes de recherche d'information », in S. Chaudiron (dir.), *Évaluation des systèmes de traitement de l'information*, Paris, Hermès Science – Lavoisier, 2004, p. 287-310.
- CHAUDIRON Stéphane et IHADJADENE Majid, « Quelle place pour l'utilisateur dans l'évaluation des SRI ? », in Viviane Couzinet et Gérard Régimbeau (dir.), *Recherches récentes en sciences de l'information – Convergences et dynamiques*, Actes du colloque MICS LERASS, 21-22 mars 2002, Toulouse, Paris, ADBS, 2002, p. 211-231.
- CHAUDIRON Stéphane, « Terminologie, ingénierie linguistique et gestion de l'information », in Loïc Depecker (dir.), *La Terminologie : nature et enjeux, Langages*, n° 157, 2005, Larousse, Paris, p. 25-35.
- CHAUMIER Jacques, *Le Traitement linguistique de l'information*. Entreprise moderne d'édition, Paris, 1988, 186 p.
- CHAUMIER Jacques, « L'indexation documentaire : de l'analyse conceptuelle humaine à l'analyse automatique morphosyntaxique », *Documentaliste Sciences de l'information*, vol.27, n°6, Paris, ADBS, 1990, p. 275-279.

- CHAUMIER Jacques, *Les Techniques documentaires*, 9^e édition, Paris, PUF, coll. « Que sais-je ? », 2002, (1^{ère} éd. : 1971), 127 p.
- CHAUMIER Jacques, GICQUEL Florence, *Les Techniques documentaires au fil de l'histoire – 1950-2000*, Paris, ADBS Éditions, 2003, 179 p.
- CHU Heting, *Information representation and retrieval in the digital age*, Medford – New Jersey, Information Today for ASIST, Inc., 2003, 248 p.
- COOL Colleen, « The concept of situation in Information Science », *Annual Review of Information Science and Technology (ARIST)*, Martha E. Williams (éd.), vol. 35, Medford, New Jersey (USA), Information Today Inc., 2001, p. 5-42.
- COURTIAL Jean-Pierre, *Introduction à la scientométrie – De la bibliométrie à la veille technologique*, Paris, Anthropos – Economica, 1990, 137 p.
- COURTIAL Jean-Pierre (dir.), *Science cognitive et sociologie des sciences*, Paris, PUF, 1994, 221 p.
- COURBIÈRES Caroline, « Une approche communicationnelle de l'analyse documentaire », in Viviane Couzinet et Gérard Régimbeau (dir.), *Recherches récentes en sciences de l'information – Convergences et dynamiques – actes du colloque MICS LERASS 21-22 mars 2002, Toulouse*, Paris, ADBS, 2002, p. 105-125.
- COUZINET Viviane, RÉGIMBEAU Gérard (dir.), *Recherches récentes en sciences de l'information – Convergences et dynamiques*, Actes du colloque MICS LERASS 21-22 mars 2002, Toulouse, Paris, ADBS, 2002, 418 p.
- COYAUD Maurice, *Introduction à l'étude des langages documentaires*, thèse de doctorat de 3^e cycle, Librairie C. Klincksieck, Paris, 1966, 148 p.
- COYAUD Maurice, *Linguistique et documentation*, Paris, Larousse, 1972, 173 p.
- CUTTER Charles Ammi, "Library catalogues", in *Public libraries in the United States of America, their history, condition and management – special report*, Department of the interior, bureau of education, Part I, Washington, Government printing office, 1876, p. 526-622.
- CUTTER Charles Ammi, *Rules for a dictionary catalog*, Washington, Government Printing Office, 1904, 4^e édition mise à jour, 173 p.
- DAHLBERG Ingetraut, « Knowledge Organization and Terminology : Philosophical and Linguistic Bases », in *International Classification*, vol. 19, n° 2, Frankfurt, Indeks Verlag, 1992, p. 65-71.
- DALBIN Sylvie, SALLERAS Bruno, « Une expérience d'utilisation d'un système d'information documentaire en langage naturel », *Documentaliste - Sciences de l'information*, vol. 37, n° 5, Paris, ADBS, 2000, p. 312-324.
- DALBIN Sylvie, « Thésaurus et informatique documentaires : partenaires de toujours ? », in *Documentaliste – Sciences de l'information*, vol. 43, n°1, Paris, ADBS, 2007, p. 42-55.
- DÉGEZ Danièle, MÉNILLET Dominique, *Thésauroglossaire des langages documentaires – Un outil de contrôle sémantique*, Paris, ADBS, 2001, 181 p.
- Documentaliste – Sciences de l'information*, vol. 43, n°1, numéro spécial langages documentaires et outils linguistiques, Paris, ADBS, 2007.
- DUBOIS Danièle, « Interrogation documentaire : recherche d'information ou gestion de connaissances ? », in T. Bulot, D. Dubois, C. Batime (éd.), *Cahiers de linguistique sociale. Recherches documentaires*, Actes des deuxièmes Journées Linguistique et ergonomie, Université de Rouen, 17-18 nov, 1994, p. 87-96.
- DUFOUR Arnaud, GHERNAOUTI-HÉLIE Solange, *Internet*, Paris, PUF, 2006, 127 p.

- ESCARPIT Robert, *Théorie générale de l'information et de la communication*, Paris, Hachette, 1976, 218 p.
- FARRADANE Jason, « Relational indexing – Part I », in *Journal of Information Science*, vol.1, Amsterdam, North Holland Publishing Company, 1980, p. 267-276.
- FARRADANE Jason, « Relational indexing – Part II », in *Journal of Information Science*, vol.1, Amsterdam, North Holland Publishing Company, 1980, p. 313-324.
- FAVIER Laurence et MARTIN-JUCHAT Fabienne, « La science de l'information face à de nouveaux paradigmes : prise en compte de la dimension sociale de la recherche d'information et remise en cause de la figure de l'utilisateur », in Viviane Couzinet et Gérard Régimbeau (dir.), *Recherches récentes en sciences de l'information – Convergences et dynamiques*, Actes du colloque MICS LERASS, 21-22 mars 2002, Toulouse, Paris, ADBS, 2002, p 255-268.
- FEATHER John, STURGES Paul (éd.), *International Encyclopedia of Information and Library Science*, 2^e éd., London – New York, Routledge, 2003, 688 p.
- FLUHR Christian, « Indexation et recherche d'information textuelle », in Jean-Marie Pierrel (dir.), *Ingénierie des langues*, Paris, Hermès Science Publications, 2000, p. 235-251.
- FOENIX-RIOU Béatrice, *Guide de recherche sur Internet – Outils et méthodes*, Paris, ADBS – Nathan, 2002, 127 p.
- GARDIN Jean-Claude *et al.*, *Le SYNTOL, étude d'un système général de documentation automatique*, Bruxelles, Presses académiques européennes, 1964, 4 vol., 589 p.
- GAUSSIÉR Éric, STÉPHANINI Marie-Hélène, *Assistance intelligente à la recherche d'informations*, Paris, Lavoisier Hermès, 2003, 319 p.
- GUIMIER-SORBETS Anne-Marie, « Des textes aux images : accès aux informations multimédias par le langage naturel », *Documentaliste - Sciences de l'information*, vol. 30 n° 3, Paris, ADBS, 1993, p. 127-134.
- GUYOT Brigitte, *Dynamiques informationnelles dans les organisations*, Paris, Hermès Science – Lavoisier, 2006, 236 p.
- HARTLEY R.V.L., « Transmission of Information », in *The Bell System Technical Journal – A journal devoted to the scientific and engineering aspects of electrical communication*, New York, America Telephone and Telegraph Company, 1928, p. 535-563.
- HJØRLAND Birger, « The concept of « subject » in information science », in *The Journal of Documentation*, vol.48, n°2, London, ASLIB, 1992, p. 172-200.
- HJØRLAND Birger, « Domain analysis in information science – Eleven approaches, traditional as well as innovative », in *Journal of documentation*, vol. 58, n° 4, Bradford (England), Emerald, 2002, p. 422-462.
- HJØRLAND Birger, « Information Retrieval, Text Composition, and Semantics » in *Knowledge Organization*, vol. 25, n° 1-2, Frankfurt, Index Verlag, 1998, p. 16-31.
- HJØRLAND Birger, *Information Seeking and Subject Representation – An Activity-Theoretical Approach to Information Science*, Westport, Connecticut, Greenwood Press, 1997, 213 p.
- HJØRLAND Birger, ALBRECHTSEN Hanne « Toward a New Horizon in Information Science: Domain-Analysis » in *Journal of the American Society of Information Science*, vol.46, n°6, New York, John Wiley & Sons Inc., 1995, p. 400-425.
- HOLZEM Maryvonne, *Terminologie et documentation. Pour une meilleure circulation des savoirs*, Paris, ADBS Éditions, 1999, 292 p.
- HUFSCHMITT Benoît, COTTEN Jean-Pierre, VARET Marie-Madeleine (textes réunis et présentés par), *Documentation et philosophie II – À propos de l'indexation discursive*

- *Autour des travaux de Muriel Amar*, (Philex vol. 7 – Annales Littéraires de l'Université de Franche-Comté, 754), Besançon, Presses universitaires franc-comtoises, 2003, 130 p.
- IHADJADENE Madjid (dir.), *Méthodes avancées pour les systèmes de recherche d'information*, Paris, Lavoisier Hermès, 2004, 247 p.
- IHADJADENE Madjid (dir.), *Les Systèmes de recherche d'informations – Modèles conceptuels*, Paris, Lavoisier Hermès, 2004, 216 p.
- IHADJADENE Majid, « Usages des moteurs de recherche », in S. Chaudiron (dir.), *Évaluation des systèmes de traitement de l'information*, Paris, Hermès Science – Lavoisier, 2004, p. 337-351.
- Information, interaction, intelligence*, numéro hors-série « Le Web sémantique », coordonné par J. Charlet, P. Laublet, C. Reynaud, Toulouse, Cépaduès Éditions, 2004, 187 p.
- INGWERSEN Peter, « Cognitive Information Retrieval », in *ARIST*, vol. 34, Martha E. Williams Editor, Medford, ASIS, 1999, p. 3-52.
- INGWERSEN Peter, « Cognitive perspectives of information retrieval interaction : elements of a cognitive IR theory », in *Journal of Documentation*, vol. 52, n° 1, London, Aslib, 1996, p. 3-50.
- INGWERSEN Peter, *Information Retrieval Interaction*, London, Taylor Graham Publishing, 1992, 246 p.
- INGWERSEN Peter, « Search procedures in the library – Analysed from the cognitive point of view », in *Journal of Documentation*, vol. 38, n° 3, Aslib, London, 1982, p. 165-191.
- La Recherche d'information sur les réseaux – Cours INRIA, 30 septembre – 4 octobre 2002, Le Bono*, Paris, ADBS, 2002, 322 p.
- International Organization for Standardization, ISO 5127 /1-1983 (E/F), *Documentation et information – Vocabulaire – Partie 1 : Notions fondamentales*, Genève, ISO, 1983, 20 p.
- International Organization for Standardization, ISO 2788 : *Guidelines for the establishment and development of monolingual thesauri*, 2^e éd. Genève, ISO, 1986, 32 p.
- International Organization for Standardization, ISO 5127:2001, *Information and documentation – Vocabulary*, Genève, ISO, 2001, 152 p.
- JACOB Elin K., SHAW Debora, « Sociocognitive Perspectives on Representation », in *ARIST*, vol. 33, Medford, ASIS, 1998, p. 131-185.
- JACOBI Daniel et AUGER Nathalie, « Remarques sur quelques convergences entre sciences de l'information et de la communication et sciences du langage », in Viviane Couzinet et Gérard Régimbeau (dir.), *Recherches récentes en sciences de l'information – Convergences et dynamiques – actes du colloque MICS LERASS 21-22 mars 2002, Toulouse*, Paris, ADBS, 2002, p. 91-104.
- KOLMAYER Elisabeth. « Représentations de la situation d'interrogation : quelques approches en sciences de l'information », in T. Bulot, D. Dubois, C. Batime (éd.), *Cahiers de linguistique sociale. Recherches documentaires*, Actes des deuxièmes Journées Linguistique et ergonomie, Université de Rouen, 17-18 nov, 1994, p 27-34.
- LAFOUGE Thierry, LE COADIC Yves-François, MICHEL Christine, préface de Leo Egghe, *Éléments de statistique et de mathématique de l'information – Infométrie, bibliométrie, médiométrie, scientométrie, muséométrie, webométrie*, Villeurbanne, Presses de l'ENSSIB, 2002, 319 p.
- LAINE-CRUZEL Sylvie, « Vers de nouveaux systèmes d'information prenant en compte le profil des utilisateurs », *Documentaliste – Sciences de l'information*, vol. 31, n° 3, 1994, p. 143-147.

- LAMIZET Bernard, SILEM Ahmed (dir.), *Dictionnaire encyclopédique des sciences de l'information et de la communication*, Paris, Ellipses, 1997, 590 p.
- LANCASTER F. W., *Indexing and Abstracting in Theory and Practice*, Champaign, Illinois, University of Illinois, 3^e éd., 2003, 451 p.
- LE COADIC Yves-François, *Le Besoin d'information – Formulation, négociation, diagnostic*, Paris, ADBS Éditions, 2^e éd., 2007, 204 p.
- LE COADIC Yves-François, *La Science de l'information*, Paris, PUF, coll. « Que sais-je ? », 1997, 2^e édition corrigée (1^{ère} éd. : 1994), 127 p.
- LE COADIC Yves-François, *Usages et usagers de l'information*, Paris, ADBS – Nathan, 2001, 127 p.
- LEFEVRE Philippe. *La Recherche d'informations - Du texte intégral au thésaurus*, Hermès Science Publications, Paris, 2000, 253 p.
- LE GUERN Michel, « Les descripteurs d'un système documentaire. Essai de définition », in *Condenser* (suppl. n°1), Actes du colloque Traitement automatique des langues naturelles et systèmes documentaires, Clermont-Ferrand, CILN, 1984, p. 163-169.
- LE GUERN Michel, « Un analyseur morpho-syntaxique pour l'indexation automatique », *Le Français moderne*, 1 (59), Paris, CILF, 1991, p. 22-35.
- LE MAREC Joëlle, « Les études d'usage et leur prise en compte dans le champ culturel », in S. Chaudiron (dir.), *Évaluation des systèmes de traitement de l'information*, Paris, Hermès Science – Lavoisier, 2004, p. 353-372.
- MANIEZ Jacques, *Actualité des langages documentaires – Fondements théoriques de la recherche d'information*, Paris, ADBS, 2002, 395 p.
- MANIEZ Jacques, « Database merging and the compatibility of indexing languages », in *Knowledge Organization*, vol. 24, n°4, p. 213-224.
- MANIEZ Jacques, *Les Langages documentaires et classificatoires : conception, construction et utilisation dans les systèmes documentaires*, Paris, Les Éditions d'organisation, 1987, 291 p.
- MANIEZ Jacques, MUSTAFA EL HADI Widad (éd.), *Organisation des connaissances en vue de leur intégration dans les systèmes de représentation et de recherche d'information*, Villeneuve d'Ascq, Université Charles-de-Gaulle – Lille 3, 1999, 403 p.
- MARON M. E., « On Indexing, Retrieval and the Meaning of About », in *Journal of the American Society for Information Science*, vol. 28, n°1, Washington, John Wiley & Sons – ASIS, 1977, p. 38-43;
- MORIZIO Claude, *La Recherche d'information*, Paris, Nathan - ADBS, coll. 128, 2002, 126 p.
- MUCCHIELLI Alex, *Les Sciences de l'information et de la communication*, Paris, Hachette, coll. « Les Fondamentaux », 2001, 3^e édition, 158 p.
- MUSTAFA EL HADI Widad, « La contribution de la terminologie à la conception théorique des langages documentaires et à l'indexation des documents », *Meta*, vol. 37, n° 3, 1992, p. 465-473.
- MUSTAFA EL HADI Widad, « Indexation humaine et indexation automatisée : la place du terme et de son environnement », in *Mots, termes et contextes – Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005*, CPI – Éditions des archives contemporaines, Paris, 2007, p. 157-166.

- MUSTAFA EL HADI Widad, « Penser les nouvelles fonctionnalités du thésaurus », in *Terminologies nouvelles*, n°21, Agence de la francophonie – Communauté française de Belgique, Bruxelles, 2000, p. 65-70.
- MUSTAFA EL HADI Widad (dir.), *Terminologie et accès à l'information*, Paris, Hermès-Lavoisier, 2006, 261 p.
- NIELSEN Jakob, *Conception de sites Web : l'art de la simplicité*, (titre original : *Designing Web usability*), Paris, Campus Press, 2004, VII-387 p.
- NIELSEN Jakob, *Usability engineering*, Boston, San Diego, New York, 1993, Academic Press, XIV-358 p.
- PAGANELLI Céline, *La Recherche d'information dans une base de documents techniques en texte intégral. Étude de l'activité des utilisateurs*, thèse préparée sous la dir. de Jacques Rouault, Université Grenoble 3, 1997, 352 p.
- PETTIGREW Karen E., FIDEL Raya, BRUCE Harry, « Conceptual Frameworks in Information Behavior », in *ARIST*, vol. 35, 2001, Martha E. Williams (éd.), Medford, ASIST, 2001, p. 43-78.
- POIBEAU Thierry, *Extraction automatique d'information – Du texte brut au web sémantique*, Paris, Hermès – Lavoisier, 2003, 239 p.
- POLITY Yolla, « Évaluation des modes de recherche en langage naturel », *Documentaliste – Sciences de l'information*, vol. 31, n° 3, ADBS, Paris, 1994, p. 136-142.
- SALTON Gerard, *Automatic Text Processing, the Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley, Reading, 1989, 543 p.
- SALTON Gerard, MCGILL Michael J., *Introduction to Modern Information Retrieval*, New York, McGraw-Hill, 1983, 448 p.
- SALTON Gerard, *A Theory of Indexing*, Philadelphia (Pennsylvania), Society for Industrial and Applied Mathematics, 1975, 56 p.
- SAMIER Henry, SANDOVAL Victor, *La Recherche intelligente sur l'internet et l'intranet – Outils et méthodes*, Paris, Hermès Science Publications, 1999, 2^e éd., 190 p.
- SAMIER Henry, SANDOVAL Victor, *La Veille stratégique sur l'internet*, Paris, Hermès Science Publications, 2002, 191 p.
- SARACEVIC Tefko, « Extension and Application of the Stratified Model of Information Retrieval Interaction », in *Proceedings of the Annual Meeting of the American Society for Information Science*, vol. 34, 1997, p. 313-327.
- SARACEVIC Tefko, « Information Science », in *Journal of the American Society for Information Science*, vol. 50 n° 12, New York, John Wiley & Sons Inc., 1999, p. 1051-1063.
- SHANNON Claude Elwood, « A Mathematical Theory of Communication », in *The Bell System Technical Journal – A journal devoted to the scientific and engineering aspects of electrical communication*, New York, America Telephone and Telegraph Company, 1948, p. 379-423 et 623-656.
- Société française des sciences de l'information et de la communication (SFIC), *Les Recherches en information et communication et leurs perspectives*, Actes du XIII^e Congrès national des sciences de l'information et de la communication, Palais du Pharo, Marseille, 7-9 octobre 2002, Paris, SFIC, 2002, 488 p.
- SOLOMON Paul, « Discovering Information in Context », in *Annual review of information science and technology (ARIST)*, Blaise Cronin (Ed.), vol. 36, ASIS, Medford, New Jersey (USA), 2002, p. 229-264.
- SPARCK JONES Karen, *Automatic Keyword Classification for Information Retrieval*, London, Butterworth & Co, 1971, 253 p.

- SPARCK JONES Karen, « What is the role of NLP in text retrieval ? », in *Natural Language Information Retrieval*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1999, p. 1-24.
- STRZALKOWSKI Tomek, LIN Fang, WANG Jin, PEREZ-CARBALLO Jose, « Evaluating natural language processing techniques in information retrieval », in *Natural Language Information Retrieval*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1999, p. 113-145.
- STRZALKOWSKI Tomek (éd.), *Natural Language Information Retrieval*, Dordrecht – Boston – London, Kluwer Academic Publishers, 1999, 384 p.
- SVENONIUS Elaine, *The Intellectual Foundation of Information Organization*, Cambridge, Massachusetts, The MIT Press, 2000, 255 p.
- VAN SLYPE Georges, *Les Langages d'indexation : conception, construction et utilisation dans les systèmes documentaires*, Paris, Les Éditions d'organisation, 1987, 277 p.
- WALLER Suzanne, *L'Analyse documentaire – Une approche méthodologique*, Paris, ADBS, 1999, 319 p.
- WARREN S.R., CLARK S.N. (éd.), *Public libraries in the United States of America, their history, condition and management – Special report*, Department of the interior, bureau of education, Washington, Government Printing Office, 1876, part I, 1187 p., part II, 89 p.

Terminologie et terminographie

- Association française de normalisation, *Travaux terminologiques. Vocabulaire. Partie 1 : Théorie et application. NF ISO 1087-1*, Paris, AFNOR, 2001, 50 p.
- Association Française de normalisation, *Travail terminologique – Principes et méthodes. NF ISO 704*, Paris, AFNOR, 2001, 49 p.
- AUGER Pierre, DROUIN Patrick, L'HOMME Marie-Claude, « Automatisation des procédures de travail en terminographie », in *Meta*, vol. 36, n°1, Montréal, Presses de l'Université de Montréal, 1991, p. 119-127.
- BACHELARD Gaston, *La Formation de l'esprit scientifique – Contribution à une psychanalyse de la connaissance*, Paris, Vrin, 1938 (éd. de poche : 1993), 256 p.
- BÉJOINT Henri, MANIEZ François (dir.), *De la Mesure dans les termes*, Lyon, 2005, 445 p.
- BÉJOINT Henri, THOIRON Philippe, *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, 281 p.
- BESSÉ Bruno de, « Le domaine », in Henri Béjoint, Philippe Thoiron (dir.), *Le Sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, p. 182-197.
- BIÉBOW Brigitte, SZULMAN Sylvie, « Une approche terminologique pour catégoriser les concepts d'une ontologie », in *Ingénierie des connaissances - Évolutions récentes et nouveaux défis*, Paris, Eyrolles, Collection « Technique et scientifique des télécommunications », 2000, p. 325-336.
- BLAMPAIN Daniel, THOIRON Philippe, VAN CAMPENHOUDT Marc (dir.), *Mots, termes et contextes*, Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005, CPI – Éditions des archives contemporaines, Paris, 2007, 741 p.

- BOUDINOT A., CHARDONNEREAU J., FRABOT J.-C., *Dictionnaire : banque, bourse, commerce extérieur*, Paris, CLET Éditions Banque, 1981 et 1983 (supplément), 591 p. et 79 p.
- BOULANGER Jean-Claude., « L'aménagement des marques d'usage technolèctales dans les dictionnaires généraux bilingues », in Jean Pruvost (dir.), *Les Dictionnaires de la langue française*, Paris, Honoré Champion, 2001, p.247-263.
- BOULANGER Jean-Claude, L'HOMME Marie-Claude, « Les technolèctes dans la pratique dictionnaire générale. Quelques fragments d'une culture », in *Meta*, tome 36, vol. n° 1, Presses de l'Université de Montréal, Montréal, 1991, p. 23-40.
- CABRÉ Teresa, « Lexicographie versus terminographie : comment les technologies déplacent leurs affrontement », in H. Béjoint, F. Maniez (dir.), *De la mesure dans les termes*, Lyon, 2005, p. 189-211.
- CABRÉ Maria Teresa, « Terminologie et linguistique », in *Terminologies nouvelles*, n°21, Agence de la francophonie – Communauté française de Belgique, Bruxelles, 2000, p. 10-15
- CABRÉ Maria Teresa, *La Terminologie – Théorie, méthode et applications*, traduit du catalan et adapté par M. C. Cormier et J. Humbley, Armand Colin – Presses de l'Université d'Ottawa, Paris – Ottawa, 1998, 322 p.
- CAMPENHOUDT Marc Van, « De la lexicographie spécialisée à la terminographie : vers un 'métadictionnaire' ? », in *Le sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, p. 127-152.
- CHANSOU Michel, « Termes de la finance et de la bourse – Essai d'organisation conceptuelle d'une terminologie », in *La Banque des mots*, numéro spécial, Conseil international de la langue française, Paris, 1990, p. 35-48.
- CHANSOU Michel, « Circulation des techniques financières et des mots : cotation, fixing, fixage », in *Mots*, n° 41, Presses de la FNSP, Paris, 1994, p. 201-204.
- CHANSOU Michel, « Réalisation de produits terminographiques et observation de l'usage », in *Les Dictionnaires spécialisés et l'Analyse de la Valeur*, Louvain-la-Neuve, Peeters, 1997, p. 131-134.
- CICILE Jean-Michel, *Terminologie bancaire, économique et financière – français-anglais*, Paris, Revue Banque Édition, 2003, 206 p.
- COHEN Betty, « Méthode de repérage et de classement des cooccurrents lexicaux », in *Terminologie et traduction*, n°2/3 – 1992, Luxembourg, Commission des Communautés européennes, 1992, p. 505-511.
- Direction de la Communication de la CNCA, *Lexique financier Français – Anglais*, Paris, CNCA, 1994, 153 p.
- CONDAMINES Anne, AMSILI P., « Terminology Between Language and Knowledge : An Example of Terminological Knowledge Base », in *Proceedings of the Terminology and Knowledge Engineering Conference (TKE'93)*, Francfort, Allemagne, 1993, Springer Verlag, Heidelberg.
- CONCEIÇÃO Manuel Celio, *Concepts, termes et reformulations*, Lyon, PUL, 2005, 279 p.
- CROZET Yves et al., *Dictionnaire de banque et bourse*, Paris, Armand Colin, coll. « Cursus », 1993, 334 p.
- DECOURSELLE Gérard, *DICAB, Le Dictionnaire impossible du Crédit agricole et de la banque*, Paris, CNCA, 1988, 694 p.
- DEPECKER Loïc, *Entre signe et concept – Éléments de terminologie générale*, Paris, Presses de la Sorbonne nouvelle, 2002, 198 p.

- DEPECKER Loïc, MAMAVI Gina (éd.), *La Mesure des mots – Cinq études d’implantation terminologique*, Rouen, Publications de l’Université de Rouen, 1997, 528 p.
- DEPECKER Loïc, « Pour les jargons », in *Meta*, vol. XXXIX, n°4, Outremont, Les Presses de l’Université de Montréal, 1994, p. 736-740.
- DEPECKER Loïc, « Le signe entre signifié et concept », in *Le sens en terminologie*, sous la dir. de Béjoint Henri et Thoiron Philippe, Lyon, Presses universitaires de Lyon, 2000, p. 86-126.
- DEPECKER Loïc, DUBOIS Violette (actes réunis et mis en forme par), *Terminologie et sciences de l’information*, Actes du colloque du 2 décembre 2005, Société française de terminologie, Paris, 2006, 110 p.
- DEPECKER Loïc (dir), *Langages* n° 157, « La Terminologie : nature et enjeux », Paris, Larousse, 2005, 128 p.
- DESMET Isabel, « Variabilité et variation en terminologie et langues spécialisées : discours, textes et contextes », in *Mots, termes et contextes – Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005*, CPI – Éditions des archives contemporaines, Paris, 2007, p. 235-247.
- Dictionnaire Permanent Épargne et Produits Financiers*, Paris, Éditions législatives, 2001, feuillets 68.
- DUBUC Robert, *Manuel pratique de terminologie*, Montréal, Linguatéc éditeur inc., 2002, 194 p.
- ELUERD Roland, *Les Mots du fer et des lumières – Contribution à l’étude du vocabulaire de la sidérurgie française (1722-1812)*, Paris, Champion, 1993, 524 p.
- ERLOS Frédéric, « Référentiels terminologiques adaptables au contexte – L’exemple d’un système de recherche d’informations dans une grande entreprise », in *Actes des 7^{es} Journées internationales d’analyse des données textuelles*, Louvain (Belgique), UCL Presses universitaires de Louvain, 2004, p. 399-410.
- FELBER Helmut, *Manuel de terminologie*, Paris, UNESCO, INFOTERM, 1987, 375 p.
- GAUDIN François, *Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles*, avant-propos de Louis Guespin, préf. de Pierre Lerat, Rouen, Publications de l’Université de Rouen (n° 182), 1993, 231 p.
- GAUDIN François, *Socioterminologie – Une approche sociolinguistique de la terminologie*, Bruxelles, De Boeck – Duculot, 2003, 286 p.
- GOPNIK Myrna, *Linguistic Structure in Scientific Texts*, The Hague-Paris, Mouton, 1972, 145 p.
- GOUADEC Daniel (resp. de la publication), *Terminologie et terminotique – Outils, modèles et méthodes*, Actes de la première Université d’Automne en Terminologie, Rennes 2, 21-26 septembre 1992, Paris, La Maison du dictionnaire, 1993, 312 p.
- GUESPIN Louis, « Langue technique et langue de l’utilisateur », dans *Les Mots pour le faire*, Paris, Éditions Descartes, 1992, p. 210-225.
- GUILBERT Louis, *La Créativité lexicale*, Paris, Larousse, coll. « Larousse Université », 1975, 285 p.
- GUILBERT Louis, *La Formation du vocabulaire de l’aviation*, thèse publiée, Paris, Larousse, 1965, Tomes 1 et 2, 709 p.
- GUILBERT Louis, « La spécificité du terme scientifique et technique », *Langue française*, n° 17, Paris, Larousse, 1973, p. 5-17.
- GUILBERT Louis, *Le Vocabulaire de l’aéronautique*, Rouen, Publications de l’Université de Rouen, 1967, 361 p.

- HERMANS A. (éd.), *Les Dictionnaires spécialisés et l'analyse de la valeur*, Louvain-la-Neuve, Peeters, 1997, 286 p.
- HOUDE Olivier, KAISER Daniel, KOENIG Olivier *et al.*, *Vocabulaire de sciences cognitives*, Paris, PUF, 1998, 417 p.
- HUMBLEY John, « Terminologie et nom propre » in *Des arbres et des mots – Hommage à Daniel Blampain*, Bruxelles, Éditions du Hasard, 2006, p. 107-124.
- JACOBI Daniel, « Un produit à problèmes – Le diéthylstilboestrol (D.E.S.) », in *Travaux du Centre de recherches sémiologiques*, n° 47, Neuchâtel, Université de Neuchâtel, 1984, p. 113-145.
- JACKIEWICZ Agata, « L'expression lexicale de la relation d'ingrédience (partie – tout) », in *Faits de langues*, n° 7, Paris, Ophrys, 1996, p. 53-62.
- KOCOUREK Rostislav, *La Langue française de la technique et de la science – Vers une linguistique de la langue savante*, préface d'Alain Rey, Wiesbaden, Brandstetter Verlag, 1991, 2e éd. augmentée, refondue et mise à jour (1^{ère} éd. : 1982), 327 p.
- KOCOUREK Rostislav, « Termes et textes », in *Meta*, vol. 36, n° 1, Montréal, Presses de l'Université de Montréal, 1991, p. 71-76.
- LEBLANC Benoît, « Phraséologie et marques de fabrique », in *Terminologie et traduction*, n°2/3 – 1992, Luxembourg, Commission des Communautés européennes, 1992, p. 211-217.
- LERAT Pierre, « L'hyponymie dans la structuration des terminologies », in *Langages*, n° 98, Paris, Larousse, 1990, p. 79-86.
- LERAT Pierre, *Les Langues spécialisées*, Paris, PUF, 1995, 201 p.
- LERAT Pierre, « Lexicologie des institutions », in *Lexique* n° 3, Lille, Presses universitaires de Lille, 1985, p.
- LEROYER Patrick, SIMONSEN Henrik Køhler, « Contexte d'utilisation, contexte de communication – La double identité du dictionnaire d'entreprise », in *Mots, termes et contextes*, Actes des septièmes journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction, Bruxelles, 8-10 septembre 2005, CPI – Éditions des archives contemporaines, Paris, 2007, p. 283-298.
- LESELBAUM Jean, LABBÉ Dominique, « Lexicographie assistée par ordinateur. Signification de 'Banque' dans le vocabulaire économique », in A. Morin, P. Sébillot (éd.), *Actes des 6^{es} Journées internationales d'analyse des données textuelles, Saint-Malo, 13-15 mars 2002, Saint-Malo, 13-15 mars 2002*, 2 vol., p. 447-458.
- L'HOMME Marie-Claude, *La Terminologie : principes et techniques*, Montréal, Presses de l'Université de Montréal, 2004, 278 p.
- MORTUREUX Françoise, PETIT Gérard, « Fonctionnement du vocabulaire dans la vulgarisation et problèmes de lexique », *DRLAV* n° 40, 1989, p. 41-62.
- MORTUREUX Marie-Françoise, « Les vocabulaires scientifiques et techniques » dans *Les Carnets du Cediscor 3, Les enjeux des discours spécialisés*, Presses de la Sorbonne nouvelle, 1995, p.13-25.
- NAZARENKO Adeline, HAMON Thierry (dir.), *Structuration de terminologie*, TAL, vol. 43, n° 1, Paris, ATALA/Hermès Science Publications, 2002, 174 p.
- OTMAN Gabriel, *Les Représentations sémantiques en terminologie*, Paris, Masson, coll. « Sciences cognitives », 1996, 216 p.
- PETIT Gérard, « La désignation de 'timbre-poste' », *Les Carnets du Cediscor 3, Les enjeux des discours spécialisés*, Paris, Presses de la Sorbonne nouvelle, 1995, p. 27-42.
- PETIT Gérard, « Un hybride sémiotique : le nom déposé », *Linguisticae Investigationes*, vol. 23, n°1, p. 161-192.

- PHAL André, « Les groupes de mots et les problèmes qu'ils posent dans la prédiction de texte scientifiques destinés à l'analyse mécanographique », in *Cahiers de lexicologie*, vol. IV – 1, Paris, Didier – Larousse, 1964, p. 45-60.
- PHAL André, « La recherche en lexicologie au C.R.E.D.I.F. – La part du lexique commun dans les vocabulaires scientifiques et techniques », in *Langue française*, n° 2, Paris, Larousse, 1969, p. 73-81.
- PRINCE Violaine, *Vers une informatique cognitive dans les organisations – Le Rôle central du langage*, Paris, Masson, 1996, 190 p.
- QUEMADA Bernard, *Introduction à l'étude du vocabulaire médical (1600-1710)*, Paris, Les Belles Lettres, 1955, 198 p.
- QUEMADA Bernard, « Technique et langage – La Formation des vocabulaires français des techniques », in *Encyclopédie de la Pléiade – Histoire des techniques*, Paris, Gallimard, 1978 (1993), p. 1146-1240.
- RASTIER François, « Le terme : entre ontologie et linguistique », *La Banque des mots*, n° 7, 1995, p. 35-65.
- REBOUL Sandrine, « La terminologie télématique : problèmes de reformulations discursives autour du concept de *vidéographie* », dans Moirand Sophie *et al.* (dir.), *Parcours linguistiques de discours spécialisés*, Berne, Peter Lang, 1993, p. 22-32.
- REBOUL Sandrine, *Le Vocabulaire de la télématique. Du discours au lexique*, thèse préparée sous la dir. de M.-F. Mortureux, Université Paris 10, 1994.
- REY Alain, « Les terminologies : un défi pour le lexicologue », in *Études de lexicologie, lexicographie et stylistiques offertes en hommage à Georges Matoré*, Paris, Société pour l'Information grammaticale, 1987, p. 231-238.
- REY Alain, *La Terminologie. Noms et notions*, Paris, PUF, coll. « Que sais-je ? », 1992, (1^{ère} éd. : 1979), 128 p.
- RINT (Réseau international de néologie et de terminologie), *Terminologies nouvelles*, n° 13, « Terminologie en entreprise », Bruxelles, Agence de coopération culturelle et technique, 1995, 111 p.
- RINT, *Terminologies nouvelles*, n° 18, « Terminotique et documentation », Agence de la francophonie – Communauté française de Belgique, Bruxelles, 1998, 97 p.
- RINT, *Terminologies nouvelles*, n° 19, « Terminologie et intelligence artificielle », Agence de la francophonie – Communauté française de Belgique, Bruxelles, 1999, 136 p.
- RINT, *Terminologies nouvelles*, n° 20, « Nouveaux outils pour la néologie », Agence de la francophonie – Communauté française de Belgique, Bruxelles, 1999, 136 p.
- RONDEAU Guy, *Introduction à la terminologie*, Chicoutimi (Québec), Gaëtan Morin éditeur, 1983, (1^{ère} éd : 1981), 238 p.
- SAGER Juan C., « Pour une approche fonctionnelle de la terminologie », in Henri Béjoint Henri, Philippe Thoiron (dir.), *Les sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, p. 39-60.
- SAGER Juan C., *A Practical Course in Terminology Processing*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 1990, 258 p.
- SCHAETZEN Caroline de, « Corpus et terminologie », *Revue française de linguistique appliquée*, vol 1-2, Amsterdam, Editions 'De Werelt', 1996, p. 57-76.
- SILEM Ahmed, ALBERTINI Jean-Marie (dir.), *Lexique d'économie*, Paris, Dalloz, 2004, 713 p.
- SLODZIAN Monique, « L'émergence d'une terminologie textuelle et le retour du sens », in Henri Béjoint Henri, Philippe Thoiron (dir.), *Les sens en terminologie*, Lyon, Presses universitaires de Lyon, 2000, p. 61-85.

- SLODZIAN Monique, « Un obstacle à la conception de nouveaux outils : l'approche discontinue et fragmentaire de l'objet « dictionnaire », in *Les Dictionnaires spécialisés et l'Analyse de la Valeur*, Louvain-la-Neuve, Peeters, 1997, p. 23-31.
- SLODZIAN Monique, BOURIGAULT Didier, « Pour une terminologie textuelle », in *RNIT* n°19, Québec, Agence de la francophonie – Communauté française de Belgique, 1999, p. 29-32.
- SLODZIAN Monique, « Rationalisation des langues : l'état de la question entre les années 1930 et 1950 », in H. Béjoint, F. Maniez (dir.), *De la mesure dans les termes*, Lyon, 2005, p. 227-240.
- SOUSI-ROUBI Blanche, *Lexique de la banque et des marchés financiers*, Paris, Dalloz, 2001, 259 p.
- TEMMERMAN Rita, « Une théorie réaliste de la terminologie : le sociocognitivism », in *Terminologies nouvelles*, n°21, Agence de la francophonie – Communauté française de Belgique, Bruxelles, 2000, p. 58-64.
- TEMMERMAN Rita, *Towards New Ways of Terminology Description. The Sociocognitive Approach*, Amsterdam/Philadelphia, John Benjamins Publishing company, 2000, VX-258 p.
- TEMMERMAN Rita, KERREMANS Koen, VANDERVOORT Veerle, « La termontographie en contexte(s) », in *Mots, termes et contextes – Actes des septièmes Journées scientifiques du réseau de chercheurs Lexicologie, Terminologie, Traduction*, Bruxelles, 8-10 septembre 2005, CPI – Éditions des archives contemporaines, Paris, 2007, p. 429-439.
- TIBERGHIEEN Guy (dir.), *Dictionnaire des sciences cognitives*, Paris, Armand Colin, 2002, 336 p.
- VECCHI Dardo Mario de, « L'organisation, communauté linguistique et conceptuelle. Le cas des entreprises », in *La Revue des Sciences de Gestion*, n° 207, Épinay-sur-Orge, 2004, p. 61-71.
- VECCHI Dardo Mario de, « La terminographie dans l'innovation et l'apprentissage : perspective pragmatique », intervention présentée aux XI^{es} Journées de projectique – Innovation et apprentissage, Bidarte et Donostia, 20-21 octobre 2005, 12 p.
- VECCHI Dardo Mario de, *La Terminologie en entreprise – Formes d'une singularité lexicale*, hèse préparée sous la dir. de John Humbley, Université Paris 13, 1999, 5 vol., 1387 p.
- VECCHI Dardo de, *Vous avez dit jargon...*, Paris, Eyrolles, 2002, 159 p.
- WEXLER Peter J., *La Formation du vocabulaire des chemins de fer en France (1778-1842)*, Genève, Lille, Droz, 1955, 159 p.
- WIEGANDT Caroline, « La terminologie en entreprise : coûts de production élevés et enjeux mal perçus », in *Terminologies nouvelles* n° 13 – *Terminologie en entreprise*, Bruxelles, Agence de coopération culturelle et technique – Communauté française de Belgique, 1995, p. 17-19.
- WÜSTER Eugen, « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses », in G. Rondeau, Helmut Felber (réd.), sous la dir. de V. I. Siforov, *Textes choisis de terminologie : Vol I. Fondements théoriques de la terminologie*, GIRSTERM, Université de Laval, Québec, 1981, p. 55-114.
- WÜSTER Eugen, « La théorie générale de la terminologie – un domaine interdisciplinaire impliquant la linguistique, la logique, l'ontologie, l'informatique et les sciences des objets », *Essai de définition de la terminologie*, Actes du colloque international de

terminologie, Québec, Manoir du Lac Delage, 5-8 oct. 1975, L'Éditeur officiel du Québec, 1976, p. 49-57.

LISTE DES ANNEXES

Avertissement : les annexes sont rassemblées sous forme électronique dans le CD-ROM joint.

Annexes de la première partie

- Rapports annuels du Crédit agricole pour les années 1995 à 2003 au format Pdf
- Corpus de rapports d'activité du Crédit agricole 1995-2003 avec partition simple pour Lexico
- Corpus de rapports d'activité du Crédit agricole 1995-2003 avec partition CDN CDS pour Lexico
- Corpus de rapports d'activité du Crédit agricole 1995-2003 avec partition simple pour Hyperbase
- Intertitres présents dans le corpus RA9503

Annexes de la deuxième partie

- Liste des mots-clés du corpus obtenus par comparaison avec le *TLF*
- Nomenclature du *Lexique de la banque et des marchés financiers*
- Code monétaire et financier au 6 mai 2005 (format Pdf)
- Termes du *LBMF* attestés dans le corpus RA9503
- Vocables correspondant à des homonymes des termes ou à des emplois partiels de termes du *LBMF* dans le corpus RA9503
- Termes du *LBMF* récurrents dans le corpus RA9503 et marques de domaines dans le *TLFi* et le *NPRE*
- Vocabulaire commun du corpus RA9503 (fréquences égales ou supérieures à 9)
- Expressions entre guillemets dans le corpus RA9503
- Expressions entre parenthèses dans le corpus RA9503
- Exemple de contextes étendus pour la capture des expressions entre parenthèses dans le corpus RA9503
- Carte des sections des paragraphes du corpus RA9503 possédant une forte concentration de pronoms anaphoriques

- Antécédents apparaissant dans les contextes possédant une forte concentration de pronoms anaphoriques
- Corpus de rapports d'activité d'autres établissements bancaires (BNP Paribas, Société Générale, Crédit Lyonnais) pour la période 1997-2001

Annexes de la troisième partie

- Liste des formes graphiques avec majuscule initiale (Fréquence et récurrence)
- Comparaison des formes graphiques avec majuscule initiale et des noms propres calculés par Hyperbase (version 5.2)
- Liste des variantes du nom « Britney Spears » recensées par Google en 2003
- Liste des segments répétés du corpus RA9503 calculés avec et sans délimiteurs
- Liste complète des dénominations propres recensées dans le corpus RA9503
- Liste des toponymes par types et zones géographiques dans le corpus RA9503
- Liste des noms de marques déposés contenant Crédit agricole en 2006
- Liste complète des dénominations réutilisant « Crédit Agricole », « CAM » et « CA »
- Référentiel terminologique adaptable au contexte constitué à partir des dénominations propres d'objets et de personnes morales. Le format est celui de la collecte mise en forme pour la restitution sous la forme d'un réseau hypertexte.

INDEX DES NOMS PROPRES

Adam J.-M., 107, 698, 806, 813
 Albertini J.-M., 159, 291, 362
 Alin F., 64, 71
 Amar M., 16, 61, 86, 434, 823, 828
 Andersen J., 82
 Arnauld A., 124
 Auger P., 305
 Authier J., 155, 271
 Bailly F., 56
 Bakhtine M., 154, 414, 805
 Bates M., 16
 Belkin N., 16
 Benghozi P.-J., 68
 Bertrand-Gastaldy S., 16
 Bessé B. de, 696, 727
 Biber D., 518
 Blair D.C., 16, 82
 Boguraev B., 434, 505, 819, 820, 821,
 822
 Borzeix A., 56, 67, 71, 143, 145, 801
 Bosredon B., 447, 503, 504, 519
 Bosséno C., 56
 Bourigault D., 117, 128
 Boutet J., 48, 56, 563, 803
 Bréal M., 26
 Brunet E., 128, 194, 280, 446, 448,
 449, 453, 467, 523
 Buckland M.K., 16
 Cabré M.-T., 91, 279, 302, 303
 Cacaly S., 67, 84, 87, 93, 101, 118,
 142
 Campbell G., 82
 Charaudeau P., 75, 154, 202, 611,
 612, 810
 Chaudiron S., 16, 93, 825, 828, 829
 Choinel A., 56, 148, 729
 Cislaru G., 522
 Coiffait N., 158
 Collinot A., 606
 Conceição M.C., 389
 Condamines A., 710
 Crozet Y., 151, 351
 Culioli A., 595
 Cutter C.A., 16
 Dahlberg I., 16
 Daille B., 450, 453, 454
 Dalbin S., 16, 707
 De Mauro T., 98
 Decourselle G., 56, 292, 792
 Dégez D., 87
 Depecker L., 120, 305, 308, 825
 Desclés J.-P., 708, 709
 Desgraupes B., 191
 Drouin P., 304, 309
 Dubois D., 96, 290, 328, 802, 806,
 811, 813, 818, 826, 828
 Dubois J., 290, 719
 Ducrot O., 45, 600, 814
 Dumarsais C.C., 125, 808
 Eco U., 98, 595
 Eluerd R., 503, 719
 Erlos F., 355, 632, 752

Felber H., 21, 90, 94, 727, 836
 Fleury S., 516
 Foenix-Riou B., 40
 Fraenkel B., 56, 67, 71, 143, 145, 801, 814
 Frege G., 604, 707, 708
 Friedmann G., 156
 Gardin B., 56
 Gary-Prieur M.-N., 455, 456, 598, 599, 600, 807, 811
 Gaudin F., 46
 Genette G., 185, 701
 Germain M., 63, 64
 Gollut J.D., 319, 621, 684, 685, 687, 695
 Gonseth F., 106, 121, 796, 802, 809
 Goody J., 166
 Gouadec D., 589
 Gougenheim G., 279, 293, 296, 792
 Grice H.P., 683, 805
 Grize J.-B., 106, 107, 108, 594, 595, 596, 597, 598, 633, 684, 799
 Gross G., 338, 339
 Guéry L., 464
 Gueslin A., 56, 148, 170
 Guilbert L., 111, 352
 Guimard A., 150, 153, 159, 183
 Guiraud P., 277, 282, 329, 445
 Guyot B., 56, 66, 68, 70, 798
 Habermas J., 50, 52, 53, 55
 Habert B., 118, 166, 189, 516
 Harris Z., 116, 591
 Hartley R.V.L., 99, 100, 101
 Henry J.-P., 56, 170, 806, 830
 Hjørland B., 16
 Holzem M., 16, 90
 Humbley J., 453, 462, 463, 815, 832, 836
 Hymes D.H., 105
 Hyperbase, 30, 128, 192, 194, 280, 282, 285, 286, 287, 453, 467, 468, 570, 770, 817, 840, 841
 Ihadjadene M., 16, 93
 Imbs P., 315
 Ingwersen P., 16, 142
 Jacobi D., 389, 811
 Jacoud G., 56
 Jacquemin C., 128, 816, 817, 818, 819, 822
 Jakobson R., 101, 102, 103, 799
 Jonasson K., 598, 707, 807
 Kerbrat-Orecchioni C., 103, 104, 105, 108
 Kleiber G., 124, 454, 455, 456, 460, 490, 604, 614, 622, 652, 720
 Kocourek R., 21, 116, 272, 285, 302, 311, 339, 340, 460, 461, 462, 463, 507
 L'Homme M.-C., 309, 314
 Lacoste M., 56
 Lafon P., 278, 490, 819, 820, 821
 Lakoff G., 96
 Lancelot C., 124, 593, 598
 Le Coadic Y.-F., 69, 90, 93, 101, 440

Le Marec J., 93
 Lebart L., 30, 195, 209, 211, 231, 238,
 250, 278, 285, 329, 330, 447, 471,
 741
 Lerat P., 112, 279, 601, 833
 Lexico, 30, 128, 189, 192, 194, 215,
 216, 217, 225, 250, 282, 285, 287,
 358, 420, 468, 472, 475, 476, 481,
 484, 490, 491, 510, 517, 524, 531,
 534, 547, 556, 566, 570, 634, 667,
 669, 670, 672, 678, 679, 681, 688,
 701, 763, 768, 784, 840
 Libaert T., 144, 151
 Mac Donald D.D., 501
 Maingueneau D., 75, 145, 154, 202,
 591, 611, 612, 803
 Malaison C., 64
 Mamavi G., 305
 Marandin J-M., 593, 598, 633, 700
 Martin R., 456, 599, 600, 806, 813,
 820, 822
 Martinez W., 352, 406, 490
 Matoré G., 45, 727, 835
 Maurel D., 445, 809, 816, 818, 821
 Mayaffre D., 138, 293
 Mazière F., 606
 McDonald D.D., 452
 Miéville D., 597
 Mill J.S., 26
 Moeschler J., 45
 Mondada L., 523, 597
 Morin E., 450, 453, 454, 817, 820,
 834, 835
 Mortureux M.-F., 389, 390, 801, 835
 Muller Ch., 272, 277, 279, 445, 446,
 447, 448, 449, 820
 Mustafa El Hadi W., 16, 90, 816
 Nazarenko A., 118, 166
 Ollivier B., 49
 Otman G., 708, 709, 710
 Pagola G., 16
 Pêcheux M., 106
 Pellat J.-C., 393, 398, 698
 Phal A., 332
 Plantin C., 611, 612, 804
 Poibeau T., 434, 450, 451, 452, 477,
 632
 Prost A., 138, 330, 741
 Pustejovsky J., 434, 505, 814, 819,
 820, 821, 822
 Quemada B., 110, 279, 294
 Rastier F., 522, 594, 601, 633, 634,
 816
 Rebeyrolle J., 710
 Reboul A., 45
 Régulier M., 56, 170
 Rey A., 121, 315, 316, 395, 605, 606,
 607, 608, 834
 Rey-Debove J., 121, 315, 316, 395
 Riegel M., 215, 393, 398, 401, 464,
 596, 698, 699, 806
 Rioul R., 393, 398, 596, 698
 Rosch E., 651, 720

Rouyer G., 729
 Russel B., 604
 Sainsaulieu R., 47, 48, 49
 Salem A., 30, 118, 128, 147, 166, 194,
 195, 206, 209, 211, 231, 232, 238,
 250, 278, 285, 329, 330, 447, 471,
 741, 821
 Saracevic T., 16
 Saussure F. de, 99
 Schaeffer J.-M., 600
 Schnedecker C., 490, 491, 593, 634
 Searle J.R., 604, 605
 Shannon C.E., 98, 99, 100, 101, 102,
 103
 Siblot P., 458
 Silem A., 159, 291, 362
 Sinclair J., 118
 Sitri F., 597
 Slodzian M., 21, 117
 Sousi-Roubi B., 307, 351, 354, 373,
 792
 Sparck Jones K., 16
 Strawson P.F., 124, 125, 603, 604,
 640
 Strzalkowski T., 16
 Svenonius E., 82, 83, 84, 86
 Tamba I., 208, 447, 710, 806
 Teyssier F. de, 151
 Todorov T., 414, 799
 Tournier M., 202, 277, 741, 819, 820,
 821, 822
 Ullmann U., 279
 Vander Beke G., 282
 Vaxelaire J.-L., 451, 452, 454, 457,
 522, 601
 Vecchi D. de, 112, 113, 117, 126, 270,
 272, 462
 Vendryès J., 279
 Wagner R.-L., 741, 819
 Wartburg Von W., 279
 Weaver W., 100, 101
 Weber M., 52, 53, 54, 55
 Wexler P.J., 352, 719
 Wiegandt C., 589
 Wittgenstein L., 82, 824
 Wüster E., 21, 90, 98, 727
 Zarifian P., 47, 48, 49, 50, 51, 52, 53,
 54
 Zufferey J., 319, 621, 684, 685, 687,
 695

TABLE DES TABLEAUX

Tableau 1. Organigramme de direction simplifié du groupe Crédit Agricole S.A. en 2003.....	59
Tableau 2. Variations d'un parler d'entreprise autour de l'objet « livret jeune »	114
Tableau 3. Répartition des trois sortes de contextes dans les rapports d'activité du corpus en début et fin de période	164
Tableau 4. Rubriques les plus partagées par les sommaires des rapports d'activité du corpus (chaque rapport est signalé par une année)	176
Tableau 5. Rubriques apparues en 1997 et leur reprise dans les sommaires postérieurs	178
Tableau 6. Rubriques apparues dans les sommaires entre 1998 et 2000	179
Tableau 7. Rubriques apparues dans le sommaire de l'année 2001	180
Tableau 8. Rubriques apparues dans le sommaire 2002	182
Tableau 9. Rubriques apparues dans le sommaire 2003	184
Tableau 10. Principales caractéristiques textométriques du corpus de rapports d'activité « RA9503 » en fonction de sa partition principale	196
Tableau 11. Premières formes du corpus « Rapports d'activité » classées par ordre textométrique de fréquence	205
Tableau 12. Extrait du tableau lexical entier présentant les formes les plus fréquentes dans le corpus « Rapports d'activité », ventilées sur les neuf parties principales	206
Tableaux 13. Principales caractéristiques textométriques du corpus « Rapports d'activité » en fonction de la partition par type de contexte	207
Tableaux 14 et 15. Les dix premières formes les plus spécifiques des deux parties CDN et CDS du corpus de rapports d'activité	210
Tableau 16. Deux formes caractéristiques des CDN possédant un rapport fréquence CDN sur fréquence totale très différent	212
Tableaux 17 et 18. Les dix formes les plus typiques des deux contextes CDS et CDN du corpus des rapports d'activité	213
Tableaux 19 et 20. Principaux accroissements spécifiques positifs et négatifs de la partie RA2001.....	240
Tableaux 21 et 22. Principaux accroissements spécifiques positifs et négatifs de la partie RA2003.....	245
Tableau 23. Schéma du changement de statut de l'émetteur à partir de 2001	248
Tableau 24. Segments répétés de fréquence supérieur à 100 et comportant la particule « a »	251
Tableau 25. Quinze premières formes des accroissements spécifiques positifs de huit parties du corpus des rapports d'activité.....	253
Tableau 26. Cinquante premiers mots clés résultant de la comparaison entre les fréquences du TLF pour la période 1950-1990 et celles du corpus Rapports d'activité.....	283
Tableau 27. Principaux segments répétés correspondant à une dénomination et incluant l'une des formes du mot-clé « agricole ».....	287

Tableau 28. Extrait des quarante premiers co-occurents formes ou segments répétés de la forme « international » dans le corpus RA	290
Tableau 29. Répartition des mots-clés relevant de la langue commune en fonction des grandes catégories de composants de l'activité bancaire	295
Tableau 30. Comparaison des mots-clés relevant de la langue commune avec la nomenclature du Dictionnaire fondamental de la langue française.....	296
Tableau 31. Liste des cinquante-sept termes du LBMF récurrents dans le corpus RA	311
Tableau 32. Le champ du terme « banque » dans le LBMF et le corpus RA	312
Tableau 33. Le champ du terme « crédit » dans le LBMF et dans le corpus RA.....	313
Tableau 34. Cinquante premières formes du vocabulaire commun thématique du corpus RA	330
Tableau 35. Répartition des cinquante premières fréquences du vocabulaire commun thématique selon les grandes catégories de composants de l'activité bancaire.....	335
Tableau 36. « activité(s) », « produits » et « clientèle » et leurs principaux syntagmes incluants	341
Tableau 37. « banque », « crédit » et « épargne » et principaux syntagmes incluants	343
Tableau 38. « assurance », « entreprises » et « marché(s) » et leurs syntagmes incluants	345
Tableau 39. « gestion », « services » et « bancaire(s) » et leurs principaux syntagmes incluants	346
Tableau 40. Répartition des syntagmes utilisant le mot « banque » entre différentes variétés de terminologies.....	349
Tableaux 41 et 42. Principales formes co-occurentes de « banque à distance » et principaux syntagmes co-occurentes de ces formes	352
Tableau 43. Principaux coefficients de spécificité négative pour la partie 2003 du corpus RA	380
Tableau 44. Segments répétés fréquents incluant la forme « encours ».....	405
Tableaux 45 et 46. Principaux co-occurentes de la forme « encours » et sélection de co-occurentes des co-occurentes de cette forme.....	407
Tableau 47. Exemples d'appellations de concepts uniques dans la norme ISO 704 :2000	461
Tableau 48. Vingt premières formes du dictionnaire du corpus avec majuscule initiale	469
Tableau 49. Segments incluant une dénomination propres et ses délimiteurs	472
Tableau 50. Dix premiers segments incluant la forme « Indosuez » (827 occurrences).....	474
Tableau 51. Identification des variantes d'une dénomination propre	478
Tableau 52. Principales caractéristiques textométriques du corpus des rapports d'activité réduit aux formes graphiques entrant dans la composition de ses dénominations propres	481
Tableau 53. Formes dotées du plus fort coefficient de spécificité positif pour la partie 2003 dans le corpus RA9503 réduit aux formes correspondant à des dénominations ou parties de dénominations propres	484

Tableau 54. Principales dénominations propres appartenant à la catégorie « agents » avec leurs fréquences et leurs variantes	498
Tableau 55. Principales dénominations propres de personnes physiques appartenant à la catégorie « agents »	499
Tableau 56. Formes du corpus avec majuscule initiale et terminaison en « ca »	510
Tableau 57. Toponymes classés par fréquence décroissante supérieure à 20.....	526
Tableau 58. Classement des unités toponymiques en fonction des principales zones géographiques.....	527
Tableau 59. Classement des toponymes et apparentés présents dans le corpus.....	528
Tableau 60. Extrait de la liste des odonymes répertoriés dans le corpus « rapports d'activité »	530
Tableau 61. Liste des vingt premiers co-occurents des odonymes dans le corpus des rapports d'activité.....	532
Tableau 62. Vingt noms de villes les plus fréquents attestés dans le corpus RA9503	539
Tableaux 63 et 64. Principaux co-occurents de la forme « Paris » dans les contextes CDN (tableau 63) et CDS (tableau 64)	543
Tableau 65. Noms de pays les plus fréquents dans le corpus de rapports d'activité	546
Tableau 66. Principaux co-occurents du groupe des 20 noms de pays dans les contextes de type CDN.....	551
Tableau 67. Formes simples co-occurentes des noms de pays et principales dénominations d'entreprises auxquelles elles se rattachent	552
Tableau 68. Principaux co-occurents du groupe des 20 noms de pays dans les contextes de type CDS.....	555
Tableau 69. Quarante premiers co-occurents des graphies de « France » dans les contextes CDN.....	563
Tableau 70. Les quarante premiers co-occurents du toponyme « France » dans les CDS	565
Tableau 71. Premières fréquences de l'inventaire distributionnel droit de la forme « en ».....	566
Tableau 72. Dénominations de personnes morales (entreprises, organisations) les plus fréquentes.....	616
Tableau 73. Dénominations géographiques les plus fréquentes.....	616
Tableau 74. Dénominations de personnes physiques les plus fréquentes.....	617
Tableau 75. Dénominations d'objets les plus fréquentes.....	617
Tableau 76. Dénominations et syntagmes « délimiteurs ».....	619
Tableau 77. Structure-type des dénominations propres d'objets et informations collectées dans la partie 1995 pour « Compte service »	645
Tableau 78. Structure-type des dénominations propres d'objets et informations collectées dans la partie 2003 pour « Compte service »	647
Tableau 79. Richesse de la structure type de contenu et membres de la classe des dénominations propres de personnes morales	648

Tableau 80. Liste des cinquante-deux dénominations propres d'objets utilisées pour la collecte triées par ordre alphabétique.....	664
Tableau 81. Liste des cinquante-cinq dénominations propres de personnes morales utilisées pour la collecte.....	665
Tableau 82. Principaux co-occurents du groupe de formes « Banco Espirito Santo ».....	671
Tableau 83. Sélection de formes spécifiques du stock de co-occurents de CAES.....	678
Tableau 84. Extrait du tableau de collecte des données pour l'unité-pilote « Banque de Gestion Privée Indosuez » destiné à une restitution graphique	715
Tableau 85. Variétés de langue attestées dans le corpus Rapports d'activité.....	724
Tableaux 86, 87 et 88 Découpage théorique de la restitution de l'activité d'un groupe bancaire mutualiste.....	731
Tableau 89 Confrontation d'un découpage théorique de l'activité bancaire en secteurs avec les unités afférentes collectées.....	734
Tableau 90. Variations affectant certains des éléments composant le profil de l'entité « UI ».	752

TABLE DES FIGURES

Figure 1. Organigramme simplifié du groupe Crédit Agricole en 2004.....	57
Figure 2. L'indexation dans la chaîne documentaire.....	88
Figure 3. Schematic diagram of a general communication system (C.E. Shannon)	100
Figure 4. Reformulation du schéma de la communication (d'après C. Kerbrat-Orecchioni) ..	104
Figure 5. Schéma de la communication (selon J.-B. Grize).....	107
Figure 6. Référentiel et éléments caractéristiques susceptibles d'être mobilisés lors d'une activité d'information	122
Figure 7. Pour un intranete caractérisé par son métier et son entreprise de rattachement, situation des différents sites intranets disponibles	141
Figure 8. Découpage en zones de la page 18 du rapport d'activité du Crédit Agricole de 2002	165
Figure 9. Découpage en zones de la page 58 du rapport d'activité du Crédit Agricole 2002	167
Figure 10. Exemples d'éléments caractéristiques des rubriques non syntactiques : 1) organigramme financier ; 2) coordonnées de filiale ; 3) organigramme de direction ; 4) notice de caisse régionale	168
Figure 11. Reproduction du sommaire du rapport annuel du Crédit Agricole de 1991.....	172
Figure 12. Page extraite du rapport d'activité du Crédit Agricole pour l'année 2000 (côté gauche) et sa transposition en texte balisé exploitable par un logiciel de textométrie (côté droit).....	193
Figure 13. Spécificités de la préposition « de » en fonction de la partition du corpus entre CDN et CDS	208
Figure 14. Fréquences relatives de quatre formes des contextes CDS et CDN.....	211
Figure 15. Carte des sections (paragraphe) de l'année 1995 avec projection des formes des CDS (rouge) et des CDN (bleu)	217
Figure 16. Carte des sections (paragraphe) de l'année 1995 partitionnée en fonction des contextes CDN et CDS et sur laquelle les formes les plus typiques des deux contextes ont été projetées	219
Figure 18. Ventilation des deux groupes de formes typiques CDS et CDN sur l'ensemble du corpus partitionné en années.....	223
Figure 19. Carte des sections (paragraphe) pour l'année 2002, partitionnée en fonction des contextes CDN et CDS et sur laquelle les formes typiques des deux contextes ont été projetées	224
Figure 20. Page 28 du rapport d'activité de l'année 2002, correspondant à la zone délimitée dans la figure 19	226
Figure 21. Texte correspondant au premier paragraphe coloré en rouge de la zone isolée dans la figure 19	227
Figure 22. Texte correspondant au quatrième paragraphe non coloré de la zone isolée dans la figure 19	227

Figure 23. Répartition caractéristique d'une série chronologique homogène produite par l'AFC	232
Figure 24. AFC sur les parties 1995RA à 2000RA du corpus « rapports d'activité »	233
Figure 25. AFC sur les parties 1995RA à 2003RA du corpus « rapports d'activité »	234
Figures 26 et 27. Répartition des parties 1995cdn à 2003cdn – Figure 26 – et 1995cde à 2003cde – Figure 27 – du corpus « rapport d'activité », produite par une AFC.....	235
Figure 28. Accroissement du vocabulaire pour l'ensemble des parties dont se compose le corpus « rapports d'activité »	237
Figure 29. Ventilation en fréquences relatives des variantes graphiques du syntagme lexicalisé « caisses régionales » et de la dénomination « Crédit agricole »	242
Figures 30 et 31. Fiches signalétiques de la Caisse régionale de Crédit agricole de Paris et Ile- de-France pour l'année 2000 (figure 30) et l'année 2001 (figure 31)	243
Figure 32. Ventilation des fréquences absolues de « CNCA » et « Crédit Agricole S.A. » sur l'ensemble des parties du corpus	248
Figure 33. Ventilation de quatre segments répétés comportant la particule verbale « a ».....	252
Figure 34. Répartition des termes en grandes catégories selon leur contexte	324
Figure 35. Répartition des termes récurrents et des principales fréquences du vocabulaire commun thématique selon les composants de l'activité bancaire	336
Figure 36. Ventilation des occurrences de certains syntagmes lexicalisés ayant le mot « banque » pour régissant.....	350
Figure 37. Ventilation des syntagmes sémantiquement proches de « banque à distance » dans le corpus RA	353
Figure 38. Ventilation en fréquences relatives des dénominations de domaines d'activité bancaire dont la fréquence augmente à partir de 1996.....	355
Figure 39. Occurrence du syntagme « Banque de proximité » de la partie 1995 du corpus RA	358
Figure 40. Occurrences des syntagmes « Banque de proximité » (carrés orange) et « banque de proximité » (carrés verts) pour la partie 2000 du corpus RA.....	359
Figure 41. Occurrences des syntagmes « Banque de proximité » (carrés orange) et « banque de proximité » (carrés verts) pour la partie 2003 du corpus RA.....	361
Figure 42. En pourcentage (colonne des ordonnées), les principaux taux d'intérêt du marché pour la période 1995-2003	372
Figure 43. Ventilation des fréquences du segment « taux d'intérêt » sur le Corpus RA	372
Figure 44. Cours de clôture du CAC 40 sur la période 1995-2003	374
Figure 45. Ventilation des occurrences des groupes de formes « bourse », « introduction en bourse », « CAC 40 » et « actionnaires » sur les neuf parties du corpus RA	374
Figure 46. Ventilation des occurrences de « fonds propres » dans le corpus RA.....	376
Figure 47. Fréquences relatives de « francs », « euros », « milliards » et « millions » dans le corpus RA	377

Figure 48. Spécificités des quatre groupes de formes « langue commune », « terminologie LBMF », « noms propres » et « sens restreints » dans le corpus RA.....	378
Figure 49. Spécificités du groupe de forme « langue commune » et des deux sous ensembles « déficits » et « excédents » dans le corpus RA	381
Figure 50. Spécificités d'un sous-ensemble de formes du groupe « noms propres » dans le corpus RA	384
Figure 51. Spécificités d'un sous-ensemble de formes du groupe « noms propres » dans le corpus RA	385
Figure 52. Projection des groupes de formes « EST UN » et « TERMES TEST » sur la carte des sections correspondant aux phrases de la partie 2002 du corpus	396
Figure 53. Carte des sections représentant des paragraphes de la partie 1995 sur laquelle ont été projetées les formes du groupe « PRONOMS ANAPHORIQUES »	399
Figure 54. Exemples de dénominations propres emboîtées les unes dans les autres.....	470
Figure 55. Extrait des concordances du segment répété « Agricole Indosuez »	475
Figure 56. Concordances simultanées des cinq formes « Autonomie », « Automates », « Auvray », « AUBRY » et « Atlantico »	475
Figure 57. Accroissement du vocabulaire des rapports d'activité réduit à ses formes simples avec majuscules ou en capitales incluses dans des dénominations propres.....	481
Figure 58. AFC sur le corpus RA9503 réduit aux seules formes simples ayant une majuscule initiale ou en capitales, incluses dans des dénominations propres	483
Figure 59. Spécificités caractérisant la répartition des occurrences de la liste des dénominations propres sur l'ensemble du corpus.....	485
Figure 60. Ventilations des dénominations propres en fonction des contextes syntactiques et non syntactiques du corpus rapports d'activité.....	489
Figure 61. Cartes des paragraphes du corpus RA sur laquelle ont été projetées les formes simples appartenant à une dénomination propre et ayant une majuscule initiale ou figurant en capitales.....	493
Figure 62. Répartition des dénominations propres selon les principaux composants de l'activité bancaire et précédents classements.....	496
Figure 63. Ventilations des odonymes entre contextes syntactiques et non syntactiques	531
Figure 64. Sections (paragraphes) des CDN et CDS de l'année 2003 dans lesquelles les carrés colorés en rouge indiquent la présence de formes appartenant à des odonymes.....	534
Figure 65. Ventilation des fréquences de « Rue Impériale », « Rue la Boétie » et « rue Paul Claudel »	536
Figure 66. Ventilation des odonymes étrangers sur les neuf parties du corpus RA9503.....	537
Figure 67. Fréquences absolues de quatre groupes de noms de villes dans les deux sortes de contextes CDN et CDS	540
Figure 68. Reproduction de la carte des implantations du Crédit agricole donnée dans le rapport d'activité de 1997.....	547

Figure 69. Ventilation sur les neuf parties du corpus découpées en CDN et CDS des fréquences des odonymes, des villes et des pays étrangers (France non comprise)	548
Figure 70. Ventilation en fréquences absolues des noms de pays et des graphies de « France » entre contextes à dominante syntactique et non syntactique.....	550
Figure 71. Ventilation des noms de pays et des graphies de « France » entre contextes à dominante syntactique et non syntactique, en fréquences relatives.....	550
Figure 72. Présentation des entités du groupe bancaire localisées à l'étranger dans le rapport d'activité de 1997.....	553
Figure 73. Ventilation en fréquences relatives des groupes de formes « 20 pays » et « France » sur les parties du corpus découpées en fonction des contextes de type CDN et CDS	560
Figure 74. Spécificités des segments répétés « en France », « en Europe » et « en Asie »	568
Figure 75. Spécificités des segments répétés « en France et dans le monde », « en France et à l'international » et « en France et en Europe ».....	569
Figure 76. Analyse factorielle des correspondances des fréquences d'une liste de 27 toponymes ou apparentés du corpus Rapports d'activité (RA).....	570
Figure 77. Ventilation des formes et groupes de formes « mutualisme », « Caisses locales », « proximité », « Caisses Régionales » et « Caisses régionales ».....	579
Figure 78. Répartition entre CDN et CDS des occurrences des groupes de formes des dénominations propres de personnes morales les plus fréquentes.....	624
Figure 79. Répartition entre CDN et CDS des occurrences des noms de lieux les plus fréquents	626
Figure 80. Répartition entre CDN et CDS des occurrences des dénominations propres de personnes physiques.....	628
Figure 81. Répartition entre CDN et CDS des occurrences des dénominations propres d'objets	629
Figure 82. Structure-type du contenu de la classe des dénominations propres de personnes morales	637
Figure 83. Structure type du contenu de la classe des toponymes.....	642
Figure 84. Structure type du contenu de la classe des dénominations de personnes physiques.....	643
Figure 85. Structure-type de la classe des dénominations propres d'objets.....	644
Figure 86. Restrictions apportées à l'étude de la structure-type des dénominations propres des personnes morales	653
Figure 87. Restrictions apportées à l'étude de la structure-type des dénominations propres d'objets.....	653
Figure 88. Section sélectionnée par la présence d'une occurrence de l'unité « Agilor » dans la première partie du corpus, et texte du paragraphe correspondant.....	668
Figure 89. Groupe de formes « BES » projeté sur la carte des sections et contenu de l'un des paragraphes sélectionnés par une occurrence	669
Figure 90. AFC appliquée au corpus C ² composé de 55 stocks de co-occurents.....	673

Figure 91. Répartition des occurrences de huit formes-pilotes et de leurs variantes entre contextes à dominante syntactique et non syntactique dans le corpus RA.....	675
Figure 92. AFC appliquée au corpus C² composé de 47 stocks de co-occurents.....	676
Figure 93. Processus de repérage des unités à collecter pour une unité-pilote (ici, CAES), et projection d'une forme spécifique sélectionnée (« retraite ») sur les corpus C² et RA ouverts simultanément dans deux instances de Lexico	681
Figure 94. Recherche sur les segments répétés du corpus commençant par « Ce » à partir de l'interface de constitution de groupes de formes de Lexico.....	688
Figure 95 Carte des sections de Lexico représentant un extrait du corpus RA découpé en fonction des rubriques d'origine et occurrences des variantes de « Predica ».....	701
Figure 96. Extrait du graphe de « Bagheera est-un » d'après J.-P. Desclés	709
Figure 97. Représentation graphique des unités associées à « Banque de Gestion Privée Indosuez » dans un prototype de référentiel terminologique	713
Figure 98. Répartition des relations-types selon les classes d'unités-pilotes.....	717
Figure 99. Ventilation en fréquences absolues de trois groupes de formes sur les neuf parties du corpus de rapports d'activité	736
Figure 100. Ventilation des occurrences de « banque de proximité » et « banque de détail » sur les parties du corpus de rapports d'activité	739
Figure 101. Noms de secteurs d'activité rattachés à la dénomination propre « Predica » dans le corpus de rapports d'activité	740
Figure 103. Projection des positions de 8 stocks de co-occurents sur l'axe 1 de l'AFC réalisée sur les 47 stocks de co-occurents du corpus C²	750
Figure 104. Ventilation des dénominations « Segespar », « Indocam » et « CA AM » et de leurs variantes sur les neuf parties du corpus principal RA.....	750
Figure 105. Ventilation sur les neuf parties du corpus RA de « Calyon » et de trois noms de secteurs d'activité	756
Figure 106. Ventilation sur les neuf parties du corpus RA de « CLAM » et de « UAF » ainsi que de quatre dénominations de secteurs d'activité.....	757
Figure 107. Ventilation sur les neuf parties du corpus RA des noms de neuf sociétés ainsi que des noms des principaux secteurs d'activité auxquels elles sont associées	759

TABLE DES MATIÈRES

SOMMAIRE.....	7
Introduction générale	11
PREMIÈRE PARTIE RÉFÉRENTIELS ET DISCOURS D'ENTREPRISE	35
Introduction de la première partie.....	37
Chapitre 1 Contextes : terminologies, intranets et référentiels.....	43
I. Échanges langagiers et problématiques informationnelles dans les entreprises	47
A. Évolutions récentes de l'organisation du travail	47
1. Le modèle taylorien d'organisation du travail.....	47
2. L'émergence d'un modèle de la coopération.....	49
a. Caractéristiques générales de l'évolution récente de l'organisation du travail	49
b. Paradigme de la coopération et nouvelles évidences culturelles	50
c. Un horizon pour les échanges langagiers dans les organisations.....	54
B. Structures organisationnelles et échanges d'information sur les intranets... 56	
1. Structures d'un groupe bancaire : le Crédit agricole	56
2. Échanges d'information et intranet.....	60
a. L'intranet.....	60
b. Diffusion de l'information électronique au sein du Crédit Agricole	64
C. Activités d'information	66
1. Un agent producteur, utilisateur et gestionnaire d'informations	66
2. Activités d'information, intranet et langage	71
a. Un espace de publication décloisonné pour des informations compartimentées ?	71
b. L'artisanat de la publication numérique	73
D. Le statut paradoxal de l'intranet au regard des systèmes d'organisation de l'information	81
1. Intranet et système d'organisation de l'information	81
2. Une ontologie remise en cause	83
3. La question du recours à une normalisation de la langue pour l'intranet .	86
II. Cahier des charges pour un référentiel terminologique adaptable au contexte	92
A. Un observatoire des usages linguistiques propres à une entreprise	92

B. Prise en compte des situations de communication : une rupture en trois schémas	97
1. La transparence de la langue et le signal	97
2. Message et langage	101
3. Langage et situation de communication	103
C. Restitution des variations d'un parler d'entreprise	110
1. La notion de parler d'entreprise	110
2. Ressources textuelles des référentiels terminologiques	116
D. Référentiel et référentiels	120
1. Référentiel et objets de discours	120
2. Hypothèse 1 : un référentiel peut être saisi au moyen de ses noms propres	124
3. Hypothèse 2 : un vocabulaire peut être capté à partir de ses noms propres	126
4. Contraintes pesant sur la collecte	127
Chapitre 2 Un référentiel à l'épreuve des discours	137
I Corpus et référentiel	140
A Intranantes et offre d'information sur l'intranet d'un groupe bancaire	140
B. Ressources destinées à documenter les activités d'information	143
1. Propriétés remarquables des discours de communication d'entreprise ..	143
2. Référentiel d'entreprise et période couverte par un corpus de textes	147
II. Rapport annuel et rapport d'activité	149
A. Un « outil » de communication institutionnelle	149
1. Le rapport annuel est un document financier	149
2. Le rapport annuel est un support de communication	151
B. Le rapport d'activité : un sous-genre de discours économique ?	154
1. Finalités	154
2. Le propos	157
3. Statut des partenaires légitimes	157
4. Lieu et moment légitimes	159
5. Supports matériels	160
6. Organisation textuelle	161

III. Le dispositif textuel des rapports d'activité	162
A. Les rapports d'activité sont composés d'éléments hétéroclites	162
1. Identification de trois sortes de contextes	162
2. Contextes à dominante syntactique et non syntactique.....	164
B. Le découpage d'un monde	169
1. Les rubriques de sommaire : entre histoire du genre et histoire de l'institution.....	169
2. Le sommaire : une rubrique comme les autres	172
a. Les présentoirs de la vitrine	172
b. Effets de genre	173
c. Effets kaléidoscopiques.....	174
3. Les rubriques : un premier niveau d'interaction entre la restitution de l'activité, l'évolution d'un référentiel et les textes du corpus.....	175
a. Principales évolutions des rubriques de sommaire des textes du corpus	175
b. Définition de la notion de rubrique.....	185
IV. Constitution du corpus	186
A. Le passage au format « texte ».....	186
B. Traitements appliqués au texte brut	188
1. Balisages documentaires : rapports d'activité et rubriques.....	188
2. Balisage des contextes CDN et CDS et paragraphes	190
3. Préparation des textes en vue de leur exploitation textométrique.....	194
a. Traitements automatiques de base	194
b. Principales caractéristiques textométriques du corpus	195
Chapitre 3 Première approche diachronique.....	201
I Le contrôle des contextes ou la tectonique des dispositifs textuels.....	204
A. Mise en place d'un test textométrique	204
1. Rappel des traitements textométriques fondamentaux	204
2. Principales caractéristiques textométriques de la partition CDN - CDS	207
3. Caractéristiques des vocabulaires des CDN et des CDS	209
a. Formes spécifiques des deux sortes de contextes	209
b. Formes typiques des deux sortes de contextes.....	213

B. Mise en œuvre du test textométrique de contrôle des contextes	215
1. Contrôle du découpage sur la carte des sections de Lexico.....	215
2. Ventilation des formes typiques des deux sortes de contextes	220
3. Un contexte hybride pour un nouveau lectorat ?	224
II Série textuelle chronologique et évolution du vocabulaire : faire du neuf avec du vieux ?	230
A. Le renouvellement du vocabulaire des rapports d'activité	230
1. Le renouvellement théorique du vocabulaire.....	230
2. Une rupture introduite par l'année 2001	233
B. Changement de destinataire et vocabulaire du corpus	237
1. Vérification de l'accroissement du vocabulaire du corpus	237
2. Hypothèse d'une nouvelle configuration du vocabulaire : accroissements spécifiques des parties 2001 et 2003.....	238
a. Reconfiguration du groupe Crédit agricole.....	238
b. Intégration du groupe Crédit lyonnais	244
3. Le changement de statut de l'émetteur	247
III Accroissements spécifiques et noms propres.....	253
A. Spécificités évolutives	253
B. Part des noms propres dans les accroissements spécifiques	254
Conclusion de la première partie	261
DEUXIÈME PARTIE « PARLER D'ENTREPRISE » ET COMMUNICATION INSTITUTIONNELLE.....	267
Introduction de la deuxième partie	269
Chapitre 4 Langue commune et banque grand public	275
I. La langue commune, une question de norme	277
A. Valeurs approchées	277
B. Une comparaison avec l'usage du <i>TLF</i> : attentes et réserves.....	280
C. Résultats et méthode d'évaluation des « mots-clés » en contexte.....	282
II. Langue commune et vocabulaire de la banque	286
A. Documentation des « mots-clés » par les segments répétés et les co- occurents.....	286
1. Le cas de « agricole ».....	286

2. Le cas de « international »	289
B. Dénominations cachées, sens uniques et « mots-clés » relevant de la langue commune.....	291
C. Rôle de la langue commune dans les rapports d'activité et retour au « français fondamental »	294
1. Mots-clés et composants de l'activité bancaire	294
2. Mots-clés et français fondamental	296
Chapitre 5 Un usage « générique » de la terminologie bancaire et financière	301
I. Implantation de la terminologie du domaine.....	304
A. Adoption d'une référence externe en matière de terminologie bancaire et financière.....	304
B. Le <i>Lexique de la banque et des marchés financiers</i>	306
C. Usage de la terminologie bancaire et financière dans le corpus de rapports d'activité	308
1. Termes attestés et termes récurrents	308
2. Les termes « banque » et « crédit ».....	311
II. Banalisation du vocabulaire utilisé pour la restitution de l'activité bancaire	314
A. Le recours à de termes banalisés.....	314
B. Termes récurrents et composants de l'activité bancaire.....	318
1. Agents économiques	318
2. Objets économiques	320
3. Opérations économiques.....	321
4. Ensembles complexes et propriétés	323
Chapitre 6 Activité bancaire et phraséologie des rapports d'activité	327
I. Vocabulaire commun et phraséologie des rapports d'activité.....	329
A. Le vocabulaire commun thématique	329
B. Vocabulaire commun thématique, variétés de langue et composants de l'activité bancaire.....	333
C. La phraséologie des rapports d'activité : une abondance d'économismes.	338
II. Deuxième approche diachronique : fluctuation de l'activité bancaire et variétés de langue	347
A. Syntagmes lexicalisés ayant « banque » pour régissant.....	347

B. Banque centrale : un usage spécifique de la terminologie standard.....	351
C. Banque à distance : le « corps à corps » des néologismes	351
D. Banque d'affaires : la patine d'un terme obsolète.....	354
E. Banque privée : un parallèle possible entre variation de fréquence et importance d'un secteur d'activité.....	356
F. Banque de proximité : les apports d'un suivi topographique et diachronique	357
G. Banque universelle : un exemple de phraséologie en mutation	362
Chapitre 7 Un vocabulaire peut-il être transparent ?	367
I. Filtrages thématiques et variations de fréquences dans les rapports d'activité	370
A. Effet de discrétion, ou de l'économiquement correct	370
B. Effet de valorisation, ou l'actualité institutionnelle... en dépit du reste.....	373
C. Variations mécaniques	375
II. Variations de fréquence et variétés de langue.....	377
A. Répartition des usages du vocabulaire commun.....	377
B. Une restitution commerciale de l'activité bancaire accentuée à partir de 2001	379
C. Terminologie bancaire et valorisation de certains domaines d'activité	382
D. Les noms propres, une prise indirecte sur le référentiel du groupe bancaire	383
E. Usages sémantiques restreints	386
III. Les rapports d'activité constituent-ils un genre de discours second ?.....	388
A. Reformulation et discours de vulgarisation	388
B. Les indices typographiques de la reformulation, ou la banalité du LBO... ..	391
C. Indices métalinguistiques de la reformulation	394
D. Utilisations de l'anaphore comme procédé de reformulation	398
1. Les anaphores pronominales et la définition déguisée d'« organe central »	398
2. Les anaphores nominales et le cas des « encours », entre jargon et banalisation	403
Conclusion de la deuxième partie	413

TROISIÈME PARTIE DÉNOMINATIONS PROPRES ET RÉFÉRENTIEL TERMINOLOGIQUE	429
Introduction de la troisième partie	431
Chapitre 8 Noms propres en discours	439
I. Recensement des noms propres du corpus	444
A. Les approches antérieures et leurs apports.....	444
1. La statistique lexicale : noms propres ou dénominations propres ?.....	444
2. Le repérage automatique des entités nommées : catégorisation et majuscule	450
B. Statut des dénominations propres et périmètre de la collecte	454
1. Questions soulevées par la détermination du périmètre de la collecte ...	454
a Fonctionnement sémantique des noms propres prototypiques.....	454
b Fonctionnement sémantique des noms propres non prototypiques	456
c Une approche terminologique globale	460
2. Relation établie dans la terminographie entre une dénomination propre et ses variantes	465
II Méthode de collecte des dénominations propres et résultats obtenus.....	467
A Identification des dénominations propres et de leurs variantes	467
1. Établissement d'une liste brute de formes avec majuscule initiale.....	467
2. Segments répétés, concordances et identification des dénominations propres	470
3. Évaluation de la collecte	476
4. Le repérage des variantes et des sigles	477
B. Les dénominations propres dans le vocabulaire du corpus.....	479
1. Aspects statistiques relatifs aux dénominations propres.....	479
2. Les dénominations propres et le vocabulaire du corpus	484
3. Topographie des dénominations propres	488
III. Dénominations propres et composants de l'activité bancaire.....	494
A. Une catégorie atypique : l'agent « personne morale ».....	494
B. La catégorisation manuelle des dénominations propres.....	500
1. Les indices de type « internal evidence »	500
2. Fréquence ordinaire et patron type caractéristiques.....	502

3. Signalétiques et classes de dénominations propres.....	503
C. Les secteurs d'activité : existence d'un <i>continuum</i> homonymique.....	511
Chapitre 9 L'ancrage d'un référentiel : les toponymes.....	521
I. Les toponymes dans le corpus de rapports d'activité.....	525
A. Les principales fréquences.....	525
B. Zones géographiques.....	527
C. Types de toponymes.....	528
II. Odonymes : de la fonction de « coordonnées » à l'effet « international ».....	529
A. La fonction « coordonnées ».....	529
B. Formes partagées et emprunt de signalétique.....	534
C. Odonymes au service d'un « effet international ».....	536
III. Noms de villes : d'une géographie de l'actualité secondaire aux places financières internationales.....	538
A. Le rôle de repère et les dispositifs de « coordonnées » dans les CDN.....	538
B. À villes moyennes, actualité de second ordre.....	541
C. Places financières.....	542
IV. Pays : la « dimension internationale » et la structuration géographique de l'activité.....	545
A. La dimension internationale du groupe bancaire.....	545
B. La structuration géographique du développement international.....	549
1. Le classement des entités du groupe bancaire.....	551
2. Les formes du développement international.....	554
C. La structuration géographique de l'activité.....	558
1. Réutilisation d'une signalétique.....	562
2. France VS monde : une nouvelle organisation de la restitution de l'activité bancaire.....	565
3. Les pays vus comme des marchés.....	573
4. Une « géographie » pour le marché : développement durable et mutualisme.....	576
Chapitre 10 Contenu des dénominations propres et vocabulaire du corpus.....	587
I. Le fonctionnement discursif des dénominations propres.....	591
A. Une candidate possible au thème de discours ?.....	591

B. Un moyen d'accéder aux objets de discours	594
C. Le contenu de la dénomination propre	598
II. Formes du contenu d'une dénomination propre.....	601
A. Contenu ouvert et cadre de la collecte d'informations	601
B. Un exemple de contenu formaté.....	606
C. Une structure topique ?	609
III. Identification et caractéristiques des structures de contenu.....	613
A. Structure de contenu et rôle de la dénomination propre	613
1. Sélection des principaux représentants des classes de dénominations identifiées.....	613
2. Repérage des dénominations propres exactes.....	617
3. Le « rôle » des dénomination propres : entre structures et activités.....	621
a. La démarche suivie	621
b. Les dénominations propres de personnes morales.....	623
c. Noms de lieux	625
d. Noms de personnes physiques	627
e. Dénominations propres d'objets	629
B. Structure type du contenu d'une dénomination propre	631
1. Identification des structures types.....	631
a. Champ textuel pour l'identification d'une structure type de contenu.	633
b. Méthode utilisée pour l'identification d'une structure type de contenu	634
2. Structures types des classes de dénominations propres étudiées.....	636
a. Structure type du contenu des dénominations propres de personnes morales.....	636
b. Structure-type du contenu des noms de lieux	640
c. Structure type du contenu des dénominations de personnes physiques	642
d. Structure-type du contenu des dénominations propres d'objets	644
3. Structure-type et hiérarchisation des classes de dénominations propres	647
C. Structure type et collecte d'informations	650
1. Choix parmi les relations composant les structures-types de contenu....	650

2. Choix des classes de dénominations propres	652
Chapitre 11. Un référentiel en mouvement.....	659
I. Organisation de la collecte	662
A. Unités-pilotes	662
1. Présentation de la démarche.....	662
2. Dénominations propres d'objets utilisées pour la collecte	663
3. Dénominations propres de personnes morales utilisées pour la collecte	664
B. Une approche textométrique de la collecte	666
1. Traitement textométrique appliqué aux unités de faible et de moyenne fréquence.....	666
2. Traitement textométrique appliqué aux unités de fréquence moyenne à élevée	668
a. Méthode adoptée	668
b. Constitution d'un corpus de deuxième niveau.....	670
c. Contrôle de l'homogénéité des stocks de co-occurents du corpus C ²	672
d. Sélection des formes les plus caractéristiques de chaque stock de co-occurents.....	677
II. Réalisation de la collecte.....	682
A. Matérialisation linguistique des relations-types retenues pour la collecte.	682
1. Constructions discursives de la relation de catégorisation	682
a. Prise en charge énonciative et opérations discursives.....	682
b. Procédés de catégorisation : entre ellipse et périphrase.....	685
c. Catégorisations et classes d'entités nommées : entre exubérance et laconisme	691
2. Constructions discursives du rattachement à un nom de secteur d'activité ou de clientèle	694
a. Domaines, marchés, secteurs : saisir le découpage de l'activité.....	694
b. Domaines, marchés, secteurs : identifier la norme du corpus.....	702
3. Critères de sélection des unités	705
B. Intégration des informations collectées dans un référentiel terminologique	706
1. La catégorisation de l'entité nommée et la relation générique	706

2. Le rattachement d'une entité nommée et les relations associatives.....	710
III. Évaluation de la collecte	715
A. Caractéristiques générales de la collecte.....	715
1. Évaluation de la productivité de la collecte	715
a. La collecte varie selon le type d'entité désignée par une dénomination propre	715
b. La productivité varie en fonction des relations étudiées.....	716
2. Données collectées et organisation de l'information.....	720
a. Catégorisations multiples	720
b. Diversité des rattachements d'une même entité.....	721
B. L'accès au vocabulaire du corpus	723
1. Rappel des limites apportées par les procédures utilisées pour la collecte	723
2. Variétés de langue et catégorisation	724
3. Variétés de langue et rattachement des entités nommées	725
C. Mise en évidence de la complexité du découpage de l'activité	727
1. Construction d'un modèle théorique.....	727
2. Confrontation du modèle avec les unités collectées	732
a. Occultation de l'entreprise bancaire.....	732
b. Spécialisation et diversification de l'activité bancaire.....	733
c. Microsecteurs et macrosecteurs d'activité	735
D. L'accès aux flux d'information.....	741
1. Cartographie de la restitution des secteurs d'activité	741
a. Présentation générale	741
b. Les métasecteurs d'activité	743
c. Photographie d'une dynamique centrifuge	747
2. Suivi des manières de découper l'activité, ou l'acculturation des terminologies exogènes.....	751
a. Évolution du contenu d'une dénomination propre.....	751
b. Phénomène de substitution	755
c. Phénomène de complémentation.....	756
d. Phénomène de subsomption.....	758

Conclusion générale.....	765
Corpus, lexiques et dictionnaires utilisés.....	791
1. Corpus.....	792
2. Lexiques et dictionnaires.....	792
Bibliographie.....	795
Contextes.....	796
Discours et textes.....	799
Linguistique générale.....	805
Textométrie et traitement automatique du langage naturel.....	815
Sciences de l'information.....	823
Terminologie et terminographie.....	831
LISTE DES ANNEXES.....	839
Annexes de la première partie.....	840
Annexes de la deuxième partie.....	840
Annexes de la troisième partie.....	841
INDEX DES NOMS PROPRES.....	843
TABLE DES TABLEAUX.....	849
TABLE DES FIGURES.....	855
TABLE DES MATIÈRES.....	861

Résumé :

L'organisation de l'information sur un intranet (réseau informatique interne d'une organisation fonctionnant avec les technologies d'Internet) nécessite de nouvelles approches pour traiter la question de l'adéquation entre l'arborescence des sites et les usages linguistiques de leurs publics. Une façon de prendre en compte ces usages consiste à explorer les données textuelles représentatives d'une situation de communication spécifique. Une telle exploration est effectuée à l'aide de techniques textométriques, comme l'index hiérarchique des formes, les concordances, les segments répétés, la carte des sections d'un texte, le calcul des co-occurrences et l'analyse factorielle des correspondances. On extrait alors d'un corpus de textes de communication d'entreprise (rapports d'activité) les unités lexicales destinées à la construction d'un référentiel terminologique d'un type particulier.

Afin de prendre en compte le contexte de communication on propose d'utiliser trois sortes de repères :

- le référentiel d'objets propre à une organisation,
- les propriétés pragmatiques des noms propres,
- la collecte d'une partie du vocabulaire caractéristique du corpus utilisé comme source du référentiel terminologique, réalisée à partir d'une sélection de noms propres.

Ainsi, cette collecte ne se limite pas aux seules unités terminologiques : elle comprend également des mots relevant de la langue commune et des noms propres. Les unités appartenant au vocabulaire du corpus sont choisies en fonction du type de relations sémantiques établies avec les noms propres dans les discours. Enfin, les résultats obtenus sont évalués en termes de productivité, de fiabilité et de représentativité.

Corporate discourses and information organization – Contribution of the textual statistics to the construction of terminological thesaurus adaptable to the context

Abstract :

Information organization on an intranet (internal network of an organization, using technologies of Internet) needs new approaches handling the question of the adequacy between the structures of intranet sites and the language use of their visitors. A way to take into account these usages is to explore textual data which are representative of a specific situation of communication. Such an exploration is carried out with textual statistics tools, like hierarchical index, concordance, repeated segment, textual map, co-occurrence and cluster analysis. This corpus-based approach allows us to extract linguistic units belonging, for example, to texts of corporate communication (annual reports). Recognition and storage of such lexical data aim at the construction of a terminological thesaurus of a peculiar type.

We suggest taking into account the context of communication by using three sorts of marks :

- the particular ontology of an organization such as it is evoked in discourses,
- the pragmatic properties of the proper names,
- a selection of proper names allows gathering a part of the characteristic vocabulary of the corpus used as source for the terminological thesaurus.

This collection does not thus limit itself to the only terminological units, but also contains words of the common language and proper names. Elements belonging to the vocabulary of the corpus are selected according to the type of semantic relations established with the proper names in the texts. Finally, the results are assessed in terms of productivity, reliability and representativeness.

Mots-clés :

Intranet – organisation de l'information – linguistique de corpus – textométrie – terminologie – nom propre

Keywords :

Intranet – information organization – corpus linguistics – textual statistics – terminology – proper name

*EA 2290 SYLED – CLA2T
Université Sorbonne nouvelle – Paris 3
19, rue des Bernardins – 75005 – France*