Université de Toulouse

Mémoire présenté pour l'obtention d'une
habilitation à diriger des recherches
spécialité : linguistique

Contributions à la description de la structure morphologique du lexique et à l'approche extensive en morphologie

Nabil Hathout

CLLE-ERSS, CNRS & Université de Toulouse II-Le Mirail

Remerciements

Je remercie Jacques Durand d'avoir accepté de m'accompagner dans la présentation de cette habilitation et pour ses conseils et sa disponibilité.

Je remercie également :

– Dany Amiot, Marie-Paule Péry-Woodley, Vito Pirrelli, Michel Roché et Pierre Zweigenbaum qui ont bien voulu faire partie de mon jury d'habilitation ;
– Marc Plénat auquel mes recherches doivent beaucoup ;
– Dany Amiot, Michel Aurnague, Gilles Boyé, Myriam Bras, Jacques Durand, Cécile Fabre, Marc Plénat, Josette Rebeyrolles, Michel Roché, Ludovic Tanguy et Pierre Zweigenbaum pour leur aide dans la rédaction de ce mémoire ;
– Michel Aurnague, Andrée Borillo, Gilles Boyé, Myriam Bras, Georgette Dal, Jacques Durand, Cécile Fabre, Bruno Gaume, Hélène Giraudo, Olivier Hautecœur, Françoise Kerleroux, Stéphanie Lignon, Fabio Montermini, Philippe Muller, Fiammetta Namer, Marc Plénat, Franck Sajous, Jeannine Richard-Zappella, Michel Roché, Ludovic Tanguy et Jesse Tseng pour leur amitié, leurs conseils et les nombreuses collaborations.

Mes remerciements vont enfin à Aïda, à nos enfants Emna et Hamza et à mes parents Mohamed-Salah et Zina pour leur affection et leur soutien.

Table des matières

Introduction **9**

1 Parcours scientifique **13**
1.1 De la théorie du gouvernement et du liage à la morphologie lexicale paradigmatique . . . 13
1.2 Découvrir la structure morphologique du lexique . . . 15
1.2.1 Verbaction, TLFindex et Lexeur . . . 15
1.2.2 DéCor . . . 16
1.2.3 Analogies morphosynonymiques . . . 17
1.2.4 Proximité et paradigmes morphologiques . . . 19
1.3 La morphologie extensive . . . 20
1.3.1 Webaffix . . . 20
1.3.2 Contribution méthodologique et technique . . . 22
1.3.3 Description morphologique . . . 23
1.3.4 Productivité morphologique . . . 25
1.3.5 Diffusion et réutilisation des données . . . 26
1.4 Conclusion . . . 26

2 Un modèle théorique de la morphologie lexicale **27**
2.1 Introduction . . . 27
2.2 Principes généraux . . . 28
2.3 Un modèle à quatre niveaux . . . 29
2.4 Proximité morphologique . . . 32
2.5 Proximité lexicale . . . 34
2.6 Organisation morphologique du lexique . . . 35
2.6.1 Paradigmes morphologiques et paradigmes lexicaux . . . 35
2.6.2 Analogie . . . 40
2.7 Sémantique d'instance . . . 41
2.8 Système de contraintes . . . 42

2.9 Contraintes sur la position lexicale 43
2.10 Contraintes morphophonologiques 44
2.10.1 Dissimilation 45
2.10.2 Taille 46
2.11 Contraintes lexicales sur la forme 46
2.11.1 Recyclage 46
2.11.2 Transparence sémantique 48
2.11.3 Transparence catégorielle 52
2.12 Le jeux des contraintes : le cas des dérivés en *anti-* 53
2.13 Conclusion 55

3 Nouveau paradigme d'analyse morphologique automatique 57
3.1 Introduction 57
3.1.1 Analyse morphologique lexématique 57
3.1.2 Combiner la proximité morphologique et l'analogie formelle 60
3.1.3 Réseau morphologique 61
3.1.4 Travaux connexes 63
3.2 Proximité morphologique 64
3.2.1 Traits formels et traits sémantiques 64
3.2.2 Connecter les lexèmes *via* leurs traits 66
3.2.3 Estimer la similarité morphologique des mots 67
3.2.4 Voisinages morphologiques 69
3.3 Analogie 70
3.3.1 Connecter les familles et les séries 70
3.3.2 Analogies formelles 72
3.3.3 Implémentation 73
3.4 Premiers résultats 76
3.5 Conclusion et développements futurs 77

4 Analyse unifiée de la préfixation en *anti-* 81
4.1 Introduction 81
4.2 *Anti-* est un préfixe 83
4.2.1 Radicaux nominaux, adjectivaux et verbaux 83
4.2.2 Radicaux pluriels 85
4.2.3 *Anti*-VN 87
4.2.4 *Anti-* devant des syntagmes nominaux sans déterminant 88
4.2.5 *Anti-* devant des noms propres, des propositions, des interjections et des onomatopées 88
4.2.6 *Anti-* devant des pronoms 89
4.2.7 *Anti-* devant des syntagmes nominaux comportant un déterminant . 90
4.3 La sémantique des dérivés en *anti-* 92
4.3.1 Notions générales 94
4.3.2 Interprétations spatiales endocentriques 97
4.3.3 Interprétations spatiales exocentriques 98

4.3.4 Interprétations logiques endocentriques 99
4.3.5 Interprétations adversatives exocentriques 101
4.3.6 Interprétations adversatives endocentriques 103
4.3.7 Unité de famille 104
4.4 Caractérisation catégorielle 106
4.5 Radicaux d'emprunt 107
4.5.1 Parasynthèse 108
4.5.2 Radicaux suffixés 109
4.5.3 Allomorphie 110
4.5.4 Supplétion 111
4.5.5 Substitution de suffixes 111
4.5.6 Surcharge 111
4.5.7 Identité de forme 112
4.5.8 Élision 113
4.5.9 Suffixation d'expressions nominales 114
4.5.10 Bases désadjectivales 114
4.5.11 Périmètre du voisinage morphologique 115
4.6 Conclusion 116

Conclusion **119**

Bibliographie **123**

Introduction

Contrairement à la thèse dont la durée est souvent liée à celle de la bourse ou du financement obtenu, il n'y a pas de réelle limite au nombre d'années de carrière pour la présentation d'une habilitation à diriger les recherches. Dans mon cas, ce temps a été relativement long parce que j'ai changé de thématique de recherche après mon recrutement au CNRS et que mes travaux se répartissent dans deux domaines distincts de la linguistique — la morphologie et le traitement automatique des langues —, mais aussi parce que la direction de travaux de recherche comporte une part de responsabilité qu'il faut être prêt à assumer. L'un des préalables à cet engagement est, je crois, d'avoir soi-même un programme scientifique en état de marche, bien délimité, avec des objectifs et des orientations clairement identifiés. Je pense que mon programme de recherche a aujourd'hui atteint ce niveau de maturité.

Mon mandat au comité national de la recherche scientifique [1] et ma fonction de secrétaire scientifique de la section 34 (langues, langage, discours) ont eu un rôle déterminant dans cette maturation. Cette expérience m'a permis d'avoir une idée claire de ce que l'on attend d'un chercheur en linguistique, de ce qu'est un bon dossier scientifique, d'observer la diversité des parcours et des stratégies professionnels, de m'imprégner des nombreux exemples de réussites scientifiques et de tirer des leçons des quelques cas problématiques ou difficiles que nous avons eu à examiner. Mon passage au comité national m'a aussi permis de comparer mes recherches à celles du reste de la communauté et surtout de n'avoir plus aucun complexe sur la position de mes travaux dans la discipline. Il m'a aussi permis d'avoir une idée assez précise des recherches menées en linguistique en France. Cette connaissance du champ s'ajoute à celle que j'ai dans les domaines de la morphologie et du traitement automatique des langues. Mes collaborations avec les morphologues français ont été nombreuses et régulières, dans le cadre du GDR 2220 (description et modélisation en morphologie) dirigé par Bernard Fradin et de la rédaction collective de l'ouvrage *Aperçus de morphologie du français* (Fradin et al. 2009). Au niveau européen et international, l'organisation, en collaboration, du 3e Forum de morphologie (Fradin et al. 2003a) et des

1. J'ai en fait été élu deux fois. La première fois en 2004, sur un poste vacant, pour une durée de 6 mois seulement.

Décembrettes (Hathout et Roché 2003, Hathout et Montermini 2007, Montermini et al. 2007), devenue aujourd'hui l'une des principales conférences européennes de morphologie, m'a donné l'occasion de rencontrer une bonne partie des figures importantes de la discipline et de connaître les travaux qui sont réalisés dans les principaux centres de recherche en Europe. J'ai aussi une bonne connaissance du domaine du traitement automatique des langues. Je fais partie du comité de rédaction de la revue TAL (traitement automatique des langues) et je participe régulièrement aux comités de lecture et de programme de la conférence TALN (conférence sur le traitement automatique des langues naturelles), dont j'ai organisé en collaboration avec Philippe Muller l'édition de 2007 (Benamara et al. 2007)[2].

Je ne présente dans ce mémoire que les aspects scientifiques de mon travail. Mes autres expériences et activités sont listées dans mon CV. J'ai par ailleurs choisi de mettre l'accent sur mes chantiers en cours et de présenter plus succinctement mon parcours scientifique. Il m'a en effet semblé important de présenter en détail mes thématiques de recherches principales, la façon dont je les aborde et quelques-uns des résultats récemment obtenus. Ces chantiers définissent le cadre de mes recherches à venir et celles dans lesquelles j'envisage d'encadrer des travaux d'étudiants de master et de thèse. Les recherches en cours dans ces différents chantiers n'ont pas toutes le même état d'avancement et de finalisation. Les plus récentes comme le modèle théorique du chapitre 2 sont encore en développement. À l'inverse, le modèle informatique présenté dans le chapitre 3 est plus abouti. Ces différences se retrouvent aussi au sein d'une même étude comme l'analyse de la préfixation en *anti-* (chapitre 4) où celle des dérivés comme *anti-obèse* pourrait être affinée davantage tandis que celle de la sous-série des adjectifs comme *antiparlementaire* ou *antilimace* est plus assurée.

Ce mémoire se compose de quatre chapitres. Dans le premier, je présente mes activités scientifiques depuis mon DEA, obtenu en 1989, jusqu'à aujourd'hui. J'y montre la cohérence de mon programme de recherche et la façon dont mes chantiers de recherche actuels s'inscrivent dans la continuité de mes travaux antérieurs. J'explique notamment comment mes travaux de morphologie informatique et descriptive s'organisent autour de mes deux thèmes de recherche principaux : la découverte de la structure morphologique du lexique et la morphologie extensive.

Le deuxième chapitre décrit un modèle théorique de la morphologie dérivationnelle. Il m'est en effet apparu nécessaire de replacer l'ensemble de mes recherches tant informatiques que descriptives dans un contexte théorique général. Mon objectif est avant tout de caractériser de façon précise les paradigmes morphologiques et les objets dont ils sont constitués et d'expliciter les contraintes qui portent sur l'expression morphologique du sens et sur la forme des mots.

Je décris dans le chapitre 3 un modèle informatique qui permet de renouveler la façon dont est conçue l'analyse morphologique automatique. J'y propose un changement de paradigme qui consiste à passer d'une analyse de mots à une analyse globale du lexique. La mise en œuvre de ce type d'analyse fait intervenir les notions de proximité, de familles et de séries présentées dans le chapitre 2. Je montre comment on peut implémenter de façon simple cette proximité et comment on peut utiliser l'analogie formelle pour découvrir une partie des relations qui s'établissent au sein des familles et des séries morphologiques.

2. Farah Benamara et Sylwia Ozdowska sont les organisatrices de la conférence associée RECITAL 2007.

Le dernier chapitre présente une étude approfondie de la préfixation en *anti-* réalisée dans le cadre de la morphologie extensive. Trois problèmes y sont abordés. Le premier concerne la diversité des séquences introduites par *anti-* et les conséquences de cette pluralité sur la caractérisation de la nature de *anti-* : est-ce un préfixe ou une préposition ? Le deuxième problème est celui de l'unité de la préfixation en *anti-*. Je montre qu'il existe cinq interprétations possibles pour les dérivés en *anti-* et qu'elles peuvent être rassemblées au sein d'une même famille de sens. Enfin, j'aborde la question de la parasynthèse. J'en propose une analyse qui découle directement des contraintes morpholexicales qui sont décrites dans le chapitre 2.

Avant d'entrer dans le vif du sujet, je voudrais signaler que je travaille principalement sur le français car c'est la seule langue pour laquelle mes intuitions sont assez fiables. Le français n'est cependant pas ma langue maternelle. Mon autre langue est le tunisien, ce qui pourrait expliquer mon penchant pour une conception paradigmatique de la morphologie des langues, et notamment de celle du français. Par ailleurs, je ne travaille que sur l'écrit parce que je m'intéresse essentiellement à la sémantique lexicale et à l'interface entre la sémantique et la morphologie. Je voudrais aussi indiquer qu'une grande partie du travail présenté dans ce mémoire est issue de collaborations à la fois nombreuses et régulières comme en témoigne la liste de mes publications, et souligner que les propositions et les analyses détaillées dans les chapitres 2 et 4 sont essentiellement une interprétation personnelle des idées, des principes et des réflexions qui constituent le substrat scientifique de l'axe DUMAL, le groupe des morphologues de CLLE-ERSS [3].

3. Les morphologues et les phonologues de CLLE-ERSS sont aujourd'hui rassemblés au sein de l'axe MorPho-Lex.

CHAPITRE 1

Parcours scientifique

Je retrace dans ce chapitre mon parcours scientifique en m'attachant à montrer sa cohérence et à présenter la façon dont s'articulent les recherches que j'ai menées durant cette première phase de ma carrière. Je montre notamment comment ces recherches débouchent sur mes chantiers actuels. Ils sont présentés de façon détaillée dans les trois chapitres suivants de ce mémoire. Ce premier chapitre est plus synthétique et décrit seulement les contributions les plus significatives de mes travaux antérieurs. Je profite aussi de ce tour d'horizon pour esquisser quelques-uns de mes projets futurs.

1.1 De la théorie du gouvernement et du liage à la morphologie lexicale paradigmatique

Mon premier travail de recherche a été un DEA dirigé par Mario Borillo sur la sémantique des prépositions directionnelles *devant*, *derrière*, *à droite de*, *à gauche de*. Cette étude a été menée dans un cadre pluridisciplinaire. J'ai proposé une formalisation des descriptions de Vandeloise (1986) et Herskovits (1982) dans une approche inspirée de la « physique qualitative » (Hayes 1985, Davis 1988). J'ai ensuite effectué une thèse sous la direction de Patrick Saint-Dizier sur l'analyse syntaxique automatique du français basée sur la théorie du gouvernement et du liage (GB ; *Government and Binding Theory*) de Chomsky (1981, 1986). Le cadre informatique de cette thèse était la programmation logique avec contraintes (Jaffar et Lassez 1987, Colmerauer 1990, Jaffar et Maher 1994). La principale contribution de ce travail est une modélisation des principes et des théories qui composent GB dans une logique du premier ordre augmentée de contraintes.

J'ai ensuite été recruté comme chargé de recherche CNRS sur un programme de recherche en traitement automatique des langues dans un cadre génératif. J'ai été affecté à l'INaLF à Nancy. L'adéquation entre ce programme et les activités menées au sein de l'équipe de recherche et de services informatiques à laquelle j'ai été rattaché était quasi nulle. Mon programme de recherche a donc entamé une évolution longue et douloureuse vers des thématiques mieux adaptées à mon environnement de travail comme l'acquisition

de connaissances à partir du *Trésor de la Langue Française* (TLF) dont l'informatisation venait à peine de démarrer. Après une première étude linguistique des marqueurs superficiels dans les définitions de botanique et de chimie (Hathout 1996a, b), je me suis lancé dans l'analyse syntaxique des définitions du TLFi (*Trésor de la Langue Française informatisé*) et la réalisation d'un analyseur syntaxique adapté à ces données. Les difficultés que j'ai rencontrées dans le traitement des verbes m'ont donné l'occasion d'entamer ma première collaboration avec Fiammetta Namer sur la construction d'un lexique syntaxique pour le TAL à partir du Lexique-Grammaire du LADL (Hathout et Namer 1997a, b, 1998). Malheureusement, les améliorations attendues ne se sont pas concrétisées. Le développement de cet analyseur a ensuite été progressivement abandonné au profit d'une nouvelle thématique de recherche : la morphologie informatique.

Durant cette période, j'ai aussi participé à un nombre important de projets et collaborations comme l'action d'évaluation des étiqueteurs morphosyntaxiques Grace, le projet européen LE-Parole sur la constitution de corpus et de lexiques au format TEI ou le projet Iliad sur la création d'une chaîne d'analyse linguistique et statistique de l'information en corpus (Toussaint et al. 1998). J'ai également fait, en collaboration avec Pascal Amsili, quelques incursions en sémantique formelle dans le cadre de la DRT (Kamp 1981, Asher 1993). Nous nous sommes intéressés à la représentation de la négation (Amsili et Hathout 1996) et au typage des λ-DRS (Amsili et Hathout 1998).

Si la morphologie informatique est devenue, à l'INaLF, ma principale thématique de recherche, c'est grâce à la disponibilité de gros lexiques flexionnels comme TLFnome [1] et parce que j'ai tout de suite eu la conviction qu'une bonne partie des connaissances morphologiques pouvaient être acquises à partir de ces ressources. Cette conviction et les principes qui guident mes recherches actuelles se sont mis en place dès mon premier travail de morphologie informatique, à savoir la construction du lexique Verbaction, sur une proposition de Didier Bourigault. Le fait de travailler sur des lexiques complets m'a naturellement amené à m'intéresser à la structure de ces objets et m'a convaincu que cette dernière détermine largement les propriétés individuelles des mots. J'ai également été influencé par l'idée d'apprentissage endogène que les travaux de Didier Bourigault sur Lexter ont contribué à diffuser largement (Bourigault 1994). Après mon arrivée à Toulouse, les collaborations et les échanges que j'ai eus avec Marc Plénat ont été déterminants dans l'orientation de mes recherches. J'ai ainsi pris conscience de l'importance des données et surtout de celles qui ne se trouvent pas dans les dictionnaires, de la diversité des conditionnements, de la nécessité de travailler sur des grandes quantité d'exemples et de la difficulté de les collecter. J'ai aussi appris, en travaillant avec Marc Plénat, que les données ne sont pas tout et que l'analyse linguistique est un art difficile.

La suite du chapitre détaille les deux axes principaux qui structurent mes recherches en morphologie. La section 1.2 présente mes travaux de linguistique informatique relatifs à la découverte de la structure morphologique du lexique et la section 1.3 mes recherches autour de la morphologie extensive.

1. TLFnome est un lexique flexionnel construit à l'INaLF en 1996 par Jacques Maucourt et Marc Papin à partir de la nomenclature du TLF. Il a été mis au format XML en 2005 et rebaptisé Morphalou 1.0. Il est disponible sous cette forme sur le site du CNRTL (`www.cnrtl.fr/lexiques/morphalou/`).

1.2 Découvrir la structure morphologique du lexique

La première constante des recherches que je mène depuis une dizaine d'années est leur finalité : la découverte de la structure morphologique dérivationnelle du lexique. Cet objectif initialement un peu flou s'est nettement précisé après mon arrivée à l'ERSS.

1.2.1 Verbaction, TLFindex et Lexeur

J'ai réalisé mes premiers travaux en morphologie en construisant, avec l'aide de Anne Berche et Fabienne Mougin, deux ressources lexicales destinées au traitement automatique du français. La première est dérivationnelle, Verbaction, et la seconde flexionnelle, TLFindex. Verbaction est un lexique de couples verbe : nom [2] dont le nom est morphologiquement apparenté au verbe et peut être utilisé pour exprimer l'action dénotée par ce dernier [3]. L'exemple (1) illustre son format actuel dans lequel les mots sont munis d'étiquettes morphosyntaxiques au format Grace (Rajman et al. 1997). Ce lexique a été réalisé en utilisant *findaffix*, un utilitaire fourni avec le correcteur orthographique *ispell*, qui permet d'apprendre des schémas de suffixation à partir d'une liste de mots. Un schéma de suffixation est un couple de suffixes $s_1 : s_2$ qui permet de mettre en relation deux formes w_1 et w_2 telles que w_2 puisse être obtenue en enlevant à w_1 le suffixe s_1 puis en lui ajoutant le suffixe s_2. En d'autres termes, w_1 et w_2 peuvent être connectées s'il existe un radical r tel que $w_1 = r \odot s_1$ et $w_2 = r \odot s_2$ où $\odot$ dénote la concaténation. Les relations identifiées de cette manière s'établissent entre des mots qui appartiennent à la même famille morphologique. Ces schémas sont communément utilisés pour le traitement automatique de la morphologie, entre autres par Jacquemin (1997), Pirrelli et Yvon (1999), Zweigenbaum et Grabar (1999), Grabar et Zweigenbaum (1999), Gaussier (1999) ou Schone et Jurafsky (2000).

(1)
```
censurer/Vmn----      censure/Ncfs
centraliser/Vmn----   centralisation/Ncfs
centrer/Vmn----       centrage/Ncms
centrer/Vmn----       centration/Ncfs
```

J'ai ainsi adopté, dès le départ, une conception relationnelle du lexique et une approche inductive dans laquelle les relations morphologiques émergent des données, de façon globale. Je signale que ce travail a été réalisé sans connaissance théorique approfondie en matière de morphologie dérivationnelle. Ceci explique en partie le fait qu'il n'ait donné lieu à aucune publication spécifique. Les techniques que j'y ai mises en œuvre sont présentées dans le chapitre 8 de Tanguy et Hathout (2007).

La première version de Verbaction, construite à partir des entrées de la nomenclature du TLF, comprenait 6 742 dérivés. Une deuxième version a été réalisée, avec l'aide de Sylwia Ozdowska, en complétant la première au moyen des entrées secondaires du TLF et par une collecte sur le Web (Tanguy et Hathout 2002a, 2003, Hathout et Tanguy 2002a, 2006). Elle contient 9 393 couples. Verbaction illustre le type de connaissances que doit, selon moi, décrire une base de données morphologiques.

2. Dans la suite du mémoire, je noterai par $x : y$ les instances des relations lexicales.

3. Verbaction est diffusé sous licence *Creative Common* (`w3.erss.univ-tlse2.fr/ressources/verbaction/`).

La seconde ressource que j'ai réalisée est TLFindex, un lexique flexionnel construit à partir des entrées secondaires du TLF et destiné à compléter TLFnome. Les schémas d'affixation utilisés pour le créer ont été appris à partir de TLFnome. L'approche adoptée est exemplariste. Elle consiste à préserver les régularités formelles de ce lexique initial.

J'ai également construit en 2001, en collaboration avec Cécile Fabre et avec l'aide de Franck Floricic et Chrystèle Lesselingue, Lexeur, un lexique dérivationnel de noms en *-eur*. Ce lexique de 3 766 entrées décrit les relations entre ces noms et leurs bases éventuelles (Hathout et Fabre 2002). Il a été réalisé dans le but de caractériser en contexte les similarités des structures argumentales des noms déverbaux en *-eur* avec celles de leurs bases verbales. L'étude a fait apparaître que les verbes et les noms déverbaux partagent une partie très faible de leurs compléments et que la prise en compte de ce paramètre est cruciale si l'on veut décrire l'héritage des structures argumentales (Fabre et al. 2004).

1.2.2 DéCor

La morphologie informatique est devenue, à partir de 1997, mon thème de recherche principal. J'ai participé à la première école d'été de morphologie organisée à Nantes par Paul Boucher et au projet Franlex (base de données lexicales sur la morphologie dérivationnelle en français contemporain) proposé par Georgette Dal et Christian Jacquemin. Franlex a débouché sur le projet MorTAL (analyseur morphologique pour le traitement automatique de la langue) coordonné par Georgette Dal et dont les autres membres étaient Christian Jacquemin, Fiammetta Namer et moi-même (Dal et al. 1999, Hathout et al. 2002, Dal et al. 2004). J'ai également participé activement aux travaux du GDR 2220 (description et modélisation en morphologie) dirigé par Bernard Fradin.

Dans le cadre de MorTAL, j'ai développé l'analyseur DéCor (dérivation pour les corpus) composé d'un ensemble de programmes permettant de découvrir les bases des mots construits par un affixe donné. L'identification de la base d'un dérivé se fait en deux temps : DéCor rassemble dans un premier temps les éléments de la famille morphologique du dérivé, c'est-à-dire l'ensemble des mots pouvant lui être connectés par des schémas de suffixation. Il sélectionne ensuite une base parmi ces parents en fonction (*i*) de la taille de leurs séries analogiques, i.e. du nombre de couples qui participent à l'analogie formelle définie par le schéma qui les connectent au dérivé, et (*ii*) de la taille du radical qu'ils partagent avec ce dernier. Cet analyseur a clairement fait apparaître la distance qui sépare l'approche que j'ai adoptée du modèle SILEX développé par Danielle Corbin (1987, 1991) et son équipe. La différence essentielle est que dans le modèle SILEX les dérivés sont construits par des règles de construction de mots (RCM) indépendantes les unes des autres. Par la suite, la construction de chaque mot particulier est également indépendante si l'on omet le principe de copie. Les spécificités de chaque dérivé peuvent ainsi être réglées au moyen d'ensembles d'exceptions comme l'a implémenté Fiammetta Namer (2005) dans son analyseur DériF (dérivation en français ; sur cette question, voir aussi Dal et Namer (2005)). En revanche, dans une conception relationnelle, chaque dérivé entretient des relations avec l'ensemble des membres de sa famille et celle qui le lie à sa base n'est pas intrinsèquement différente des autres. Une autre différence vient du fait que DéCor n'implémente aucune connaissance linguistique particulière à l'exception de l'observation très générale que deux mots

qui partagent leur radical sont susceptibles d'appartenir à la même famille morphologique. Les bases sont identifiées uniquement à partir des formes des mots du lexique. Or, leurs spécificités et celles des leurs dérivés ne peuvent être prises en compte qu'en utilisant des données plus riches qui contiennent des informations sémantiques, comme je l'ai fait par la suite dans Hathout (2001) et Hathout (2009a).

J'ai au préalable examiné deux autres méthodes susceptibles de rapprocher les résultats de DéCor de ceux de DériF. La première consiste à utiliser des indices statistiques globaux comme le rapport entre la fréquence du dérivé et celle du candidat ou la mesure de l'information mutuelle des suffixes au sein des couples de mots (Hathout 2000). Ces mesures n'ont malheureusement pas permis d'améliorer significativement les performances de DéCor pour les suffixations comme *-iser* dont l'analyse diverge le plus de celle de DériF. Notons que le rapport entre la fréquence d'un dérivé et celle de sa base est aussi utilisé en psycholinguistique par Hay (2000) comme indice de décompositionnalité et d'analysabilité (*parsability*). Les expériences que présente Hay font apparaître que les dérivés qui sont moins fréquents que leur bases tendent à être décomposés lors de l'accès lexical tandis que ceux qui sont plus fréquents tendent à être reconnus sans décomposition, comme des formes entières, par une stratégie d'appréhension globale (*whole word route*). Cet indice contraint également l'ordre des affixes dans les mots. Hay propose la généralisation suivante : les affixes facilement analysables ne doivent pas apparaître dans la portée d'affixes non analysables.

La deuxième voie que j'ai explorée a été de redéfinir la tâche et d'utiliser un référentiel mieux adapté à DéCor que les analyses de DériF. Cette nouvelle tâche consiste à classer les mots du lexique en familles morphologiques au lieu d'associer les dérivés à leurs bases (Hathout 2005). Cette classification est une étape supplémentaire vers la découverte de la structure morphologique du lexique. Je signale que Gaussier (1999) propose une technique similaire de classification en familles morphologiques. Gaussier utilise ensuite ces familles pour identifier des morphèmes et des relations morphologiques.

1.2.3 Analogies morphosynonymiques

DéCor exploite les analogies graphémiques qui impliquent des dérivés suffixés. Il s'appuie sur l'hypothèse que le partage d'un radical commun suffisamment long par deux formes est fortement corrélé à un partage de propriétés sémantiques. Les expériences menées montrent que cette corrélation est insuffisante et qu'il est nécessaire de disposer d'une caractérisation sémantique des mots. Pour répondre à cette difficulté, j'ai mis au point une méthode d'acquisition de relations morphologiques à partir de dictionnaires de synonymes (Hathout 2001, 2002, 2003, 2005). Il s'agit d'utiliser l'analogie pour croiser les familles morphologiques avec des ensembles de synonymes. La méthode consiste à utiliser dans un premier temps DéCor pour calculer l'ensemble des relations morphographiques qui s'établissent à l'intérieur des familles morphologiques. Seules les suffixations sont considérées. On croise ensuite ces relations familiales avec des relations de synonymie pour former des

quadruplets analogiques comme en (2)[4]. En (2a) par exemple, la décoration est à l'action de décorer ce que l'embellissement est à l'action d'embellir. De même, en (2b), une facétie est un acte ou une parole facétieuse de la même façon qu'une drôlerie est un acte ou une parole drôle.

(2) a. *décoration* : *décorer* :: *embellissement* : *embellir*

b. *facétie* : *facétieux* :: *drôlerie* : *drôle*

Ce type de croisement s'avère très efficace pour former des couples de mots morphologiquement apparentés avec une précision de l'ordre de 95% pour le dictionnaire de synonymes de l'ATILF[5] et de 92% en anglais pour les synsets de WordNet (Miller et al. 1990, Fellbaum 1999). Il est également intéressant parce qu'il permet de relier des couples qui relèvent de séries dérivationnelles différentes comme en (2a) ou dans lesquels les relations morphologiques sont inversées comme en (2b). Dans ce quadruplet, *facétie* est la base de *facétieux* mais *drôlerie* est dérivé *drôle*.

J'ai également effectué une comparaison de la sélectivité de différents types de quadruplets. On peut en effet typer les quadruplets morphosynonymiques de la manière suivante :

- un quadruplet est dit « exact » (de type E) si c'est une analogie formelle comme (3) ;
- un quadruplet qui n'est pas exact est dit « hétérogène » (de type H) comme (4).

(3) *déchirer* : *déchirure* :: *érafler* : *éraflure*

(4) *déchirer* : *déchirure* :: *fendre* : *fente*

On peut ensuite projeter ce typage sur les couples qui composent les quadruplets :

- un couple est dit « exact » (de type E) s'il fait partie d'un quadruplet exact ;
- un couple qui n'est pas exact est dit « hétérogène » (de type H).

Le typage des couples permet en retour de caractériser plus finement les quadruplets de type H. Trois sous-types peuvent être distingués :

- un quadruplet est de sous-type H0 si ses deux couples sont de type E ;
- un quadruplet est de sous-type H1 si l'un de ses couples est de type E et l'autre de type H ;
- un quadruplet est de sous-type H2 si ses deux couples sont de type H .

J'ai révisé manuellement des échantillons 200 couples qui apparaissent dans ces différents types de quadruplets et j'ai mis en évidence que la meilleure précision est obtenue pour les couples E issus de quadruplets H0 ou H1. Les couples qui apparaissent dans des quadruplets E ont une précision inférieure. Cette observation, qui peut paraître contre-intuitive, s'explique par le fait que les couples E sont plus fortement contraints lorsque la caractérisation synonymique est distincte de celle qui est issue des séries morphologiques.

4. Je rappelle qu'une analogie est une relation quaternaire qui s'établit entre les éléments d'un quadruplet (a, b, c, d) tel que a est à b ce que c est à d. Un quadruplet analogique (a, b, c, d) est noté $a : b :: c : d$. Par abus de langage, on utilise aussi le terme d'« analogie » pour désigner les quadruplets qui appartiennent à une relation analogique.

5. Voir `www.atilf.fr/synonymes/` pour une présentation de cette ressource. Une version nettoyée de ce dictionnaire est consultable sur le portail lexical du CNRTL (`www.cnrtl.fr/synonymie/`).

1.2.4 Proximité et paradigmes morphologiques

Dans tous les travaux qui viennent d'être présentés, les relations morphologiques sont calculées au moyen de schémas d'affixation. Cette technique présente deux limites. La première est qu'elle nécessite un découpage « morceaulogique » des mots en un radical et un affixe, ce qui n'est pas totalement satisfaisant dans la perspective lexématique dans laquelle je me place. La seconde est qu'un bon nombre de relations lui échappent, notamment celles qui sont peu fréquentes dans les dictionnaires comme *charivari* : *charivaresque* et celles qui font intervenir des affixes non contigus comme dans *fructueux* : *infructueusement*. Il s'agit là d'un verrou pour l'ensemble des recherches en morphologie informatique. J'ai proposé dans Hathout (2008, 2009a, b) une mesure de la proximité morphologique qui permet d'identifier des voisinages au sein desquels on peut retrouver des relations morphologiques sans décomposer les mots en un radical et un ou plusieurs affixes. Cette mesure, présentée en détail dans le chapitre 3, rend compte notamment de la gradabilité des relations morphologiques (Hay et Baayen 2005). Elle constitue un élément nouveau dans la description de la structure morphologique du lexique. Cette mesure est directement inspirée des travaux que j'ai réalisés en collaboration avec Bruno Gaume et Philippe Muller sur la désambiguïsation lexicale (Gaume et al. 2004a, b) et l'acquisition de synonymes à partir des définitions du TLFi (Muller et al. 2006). Nous avons utilisé pour ces deux tâches des graphes lexicaux explorés au moyen d'algorithmes de marches aléatoires. La ressource que j'ai utilisée dans l'expérience détaillée dans le chapitre 3 a aussi été préparée conjointement lors de cette collaboration.

Cette mesure de proximité peut être complétée par l'explicitation d'autres éléments de la structure morphologique, à savoir (*i*) les familles et les séries flexionnelles, dérivationnelles, morphologiques et lexicales et (*ii*) les analogies qui permettent de les interconnecter. Ces paradigmes forment un maillage qui permet de redéfinir la tâche d'analyse morphologique. L'ensemble est décrit dans les chapitres 2 et 3. Dans le chapitre 2, proximité et paradigmes sont intégrés dans un embryon de théorie de la morphologie dérivationnelle afin de montrer comment ces éléments s'organisent et s'utilisent. Il s'agit d'un cadre basé sur l'usage, dans lequel performance et compétence ne sont pas séparées. Ce cadre me permet aussi d'articuler mes travaux en morphologie informatique et mes travaux de description morphologique.

Les recherches que je décris dans le chapitre 3 illustrent la possibilité de pratiquer la linguistique en construisant des modèles informatiques et donnent une idée de la façon dont ces derniers permettent de renouveler les points de vue sur les objets de la discipline. Pour reprendre une expression de Didier Bourigault, je dirais que je suis moi aussi « le linguiste de ce schéma » [6], c'est-à-dire un linguiste à part entière qui s'appuie sur des objectifs et des méthodes de traitement automatique des langues pour comprendre et décrire des phénomènes linguistiques.

6. Didier Bourigault nous a présenté lors des premières grammaires aérées de l'ERSS, un schéma sur la place des outils informatiques dans la recherche en linguistique.

1.3 La morphologie extensive

Si, comme on vient de le voir, la linguistique informatique offre une voie de recherche fructueuse en morphologie, son efficacité dépend de façon cruciale de la bonne connaissance des objets et des données qu'elle traite. Cette connaissance ne peut être atteinte que par la description linguistique qui, comme l'a bien montré Marc Plénat (2000) (voir aussi Plénat et al. (2002)), doit être extensive. La **morphologie extensive**[7] est une pratique qui consiste à baser les descriptions sur la plus grande quantité possible d'exemples authentiques. Nous avons par exemple constitué un corpus de plus de 5 000 adjectifs en *-able* pour décrire les relations qui s'établissent entre ces derniers et leurs noms recteurs, soit plus du triple du nombre d'exemples considérés par les études antérieures de ce suffixe (Hathout et al. 2003). L'approche extensive relève d'une conception de la linguistique comme science d'observation[8]. Elle découle de la constatation que la quantité de données disponibles détermine la résolution des observations qui peuvent être réalisées. Plus ces données sont nombreuses, plus le niveau de détail des observations est élevé. Or, les exemples déterminants pour l'analyse d'une construction particulière sont généralement les plus rares. Pour espérer découvrir un nombre suffisant de tels spécimens, il faut disposer d'un très gros corpus d'exemples. Le linguiste doit par conséquent consacrer une part importante de son activité de recherche à la collecte systématique de tous les spécimens qu'il peut trouver. Leur compilation dans une base de données prend un temps considérable. Si la collecte proprement dite est rapide[9], la validation des données collectées exige un travail philologique long et fastidieux (Hathout et al. 2009a).

1.3.1 Webaffix

La rentabilité de la morphologie extensive est directement liée à la possibilité de collecter facilement de grandes quantités d'exemples. À l'heure actuelle, la seule ressource qui permette d'en rassembler en nombre suffisant est le Web. On peut naturellement aussi utiliser des bases textuelles, des archives de journaux ou des dictionnaires informatisés, mais ces corpus sont trop petits et leur contenu est trop normalisé pour permettre d'observer les constructions sporadiques qui nous intéressent. Le principal objectif de la collecte est en effet de trouver des exemples souvent rares voire exceptionnels capables de nous éclairer sur la manière dont s'organisent les dérivés d'une construction morphologique particulière.

La façon la plus simple de trouver des exemples sur le Web consiste à soumettre manuellement des requêtes à un moteur de recherche comme Google, Yahoo ou plus récemment Bing. Le linguiste s'appuie alors sur son intuition pour prédire des instances de la construction à laquelle il s'intéresse. Cet exercice présente cependant deux limites. La première est le temps qu'il faut pour soumettre les milliers voire les dizaines de milliers de requêtes correspondant aux différentes formes sous lesquelles la construction peut se réaliser. Quel-

7. Cette approche a été initiée et définie par Marc Plénat à qui l'on doit également son nom.

8. Je remercie Marc Plénat de m'avoir fait découvrir cette formule bien connue de Gustave Guillaume.

9. Cette affirmation doit cependant être nuancée pour un domaine connexe comme la phonologie. Dans un projet comme PFC par exemple (Durand et al. 2002, 2005), la collecte des données comporte un important travail d'enregistrement et de transcription.

qu'un qui s'intéresserait par exemple aux déverbaux en *-eur* doit soumettre pour chacun des 10 000 verbes du français [10] au moins une requête pour chacune des finales *-eur*, *-eurs*, *-euse*, *-euses*, *-trice*, *-trices*, soit environ 60 000 requêtes. La seconde limite est plus gênante que la première car elle est liée à la compétence linguistique et à l'imagination du chercheur. Nul n'est en effet capable d'imaginer spontanément la diversité et l'étendue des utilisations que les locuteurs font d'une construction donnée, même de celles qu'on considère *a priori* comme les plus simples comme les suffixations en *-able* ou en *-ion*.

Pour répondre à ces difficultés, j'ai développé, en collaboration avec Ludovic Tanguy, Webaffix, une boîte à outils d'acquisition lexicale à partir du Web (Hathout et Tanguy 2002a, b, 2006, Tanguy et Hathout 2002a, b, 2003). Webaffix permet de créer et de compléter des lexiques de façon semi-automatique par de nouvelles formes construites morphologiquement. Il offre deux fonctionnalités principales : la recherche de ces formes sur le Web et leur filtrage. Webaffix se compose de trois modules :

1. un module de recherche par motif [11] qui permet de découvrir des formes qui correspondent à un patron donné comme par exemple la possession d'une finale particulière ;
2. un composant de prédiction morphologique permettant de calculer les formes des bases ou des dérivés ;
3. un méta-moteur disposant de fonctionnalités dédiées à l'exploration lexicale du Web.

Une partie des techniques implémentées dans le deuxième et le troisième module sont détaillées dans les chapitres 8 et 10 de Tanguy et Hathout (2007). D'autres systèmes comme WaliM développé par Namer (2003b, 2005), proposent des fonctionnalités similaires à celles de Webaffix. WaliM est destiné à la validation par le Web de formes construites morphologiquement.

Webaffix propose deux modes de recherche de mots nouveaux. Le premier exploite un lexique existant ou une base de données morphologiques à partir desquels on génère une liste de formes candidates. On utilise ensuite le méta-moteur pour rechercher sur le Web des attestations de ces formes. Le second mode recourt au module de recherche par motif pour repérer sur le Web des formes qui correspondent à une configuration donnée, par exemple, toutes celles qui finissent par *-able*, sans aucune contrainte sur le radical. C'est de cette façon que nous avons découvert des formes comme *wapisable* [12] ou *copolymérisable* dont le radical ne fait partie ni de la nomenclature du TLF et ni de celle du GRLF (Hathout et al. 2003). Ce mode de recherche présente un intérêt exceptionnel car il permet de découvrir des constructions imprévisibles comme la première ou qui sont difficilement prédictibles à partir des ressources disponibles comme la seconde. Le fait de ne plus pouvoir l'utiliser est pour nous un réel handicap.

10. Ce chiffre correspond approximativement au nombre de verbes que l'on trouve dans le TLF et le GRLF (*Grand Robert de la Langue Française*).

11. Ce module faisait appel au moteur AltaVista. Suite au rachat de AltaVista par Overture en 2003, elle-même rachetée par Yahoo en 2004, ce moteur ne permet plus l'utilisation des caractères jokers nécessaires à la recherche par motif.

12. Cet adjectif est construit sur *wapiser* 'modifier un site Web pour le rendre accessible via le protocole Wap'. Wap est un protocole d'accès à Internet à partir d'un téléphone portable ou d'un assistant personnel.

Même si le premier mode de recherche est toujours opérationnel, nous avons cherché d'autres moyens de collecter des dérivés « inattendus ». La première piste explorée a été Trifouillette [13], un outil d'indexation lexicale du Web développé par Franck Sajous (Sajous et Tanguy 2006). Ce projet n'a malheureusement pas permis d'atteindre les résultats espérés du fait des difficultés liées à l'indexation du Web, et notamment aux pièges à robots (*spider traps*). Plus récemment, nous avons pu utiliser des pages Web collectées par un partenaire, la société Exalead [14]. Le corpus qu'elle a généreusement mis à notre disposition dans le cadre du projet Quaero contient 2,5 millions de pages et 3,3 milliards de mots. Une expérience a été réalisée en collaboration avec Franck Sajous et Ludovic Tanguy sur ces données dans le but d'acquérir de nouveaux noms déverbaux en *-ion*, *-ment* et *-age*. Nous avons ainsi pu estimer qu'en moyenne, moins d'une page sur 1 000 contient un dérivé valide inconnu (i.e. qui ne se trouve pas dans la version actuelle de Verbaction). Les différentes étapes de cette quête de mots nouveaux destinés notamment aux morphologues extensifs sont présentées dans Hathout et al. (2009b).

1.3.2 Contribution méthodologique et technique

Ma contribution à l'approche extensive en morphologie a également été de nature méthodologique et technique. J'ai publié en collaboration deux articles de synthèse destinés à faire connaître cette approche (Hathout et al. 2008, 2009a). Nous y présentons quelques-unes des avancées que l'on peut directement attribuer à l'utilisation de très grandes quantités d'exemples comme la mise en évidence de certains conditionnements complexes qui contrôlent la troncation des radicaux dans les dérivés en *-esque* (Plénat et Roché 2003) ou l'absence de toute contrainte catégorielle sur les bases d'adjectifs en *-able* (Hathout et al. 2003). L'approche extensive est aussi présentée en détail par Durand (2009), dans une perspective épistémologique.

L'un des enseignements des premières études de morphologie extensive qui ont été réalisées est que la qualité des exemples collectés ne constitue pas un problème car toutes les données utilisées font *in fine* l'objet d'une validation par le linguiste. Par ailleurs, la possibilité de réaliser facilement des travaux de morphologie dérivationnelle à partir de données provenant de sources faiblement contrôlées comme des pages Web distingue nettement cette discipline de la syntaxe et de la phonologie. Parmi les raisons qui sont à l'origine de cet avantage, on peut citer :

(*i*) pour la phonologie, le fait que les corpus oraux sont rares et de taille très réduite en comparaison de corpus écrits et des textes disponibles sur le Web. Le développement des techniques de reconnaissance de la parole pourrait, dans le futur, réduire cette avance ;

(*ii*) pour la syntaxe, le fait que l'unité d'indexation privilégiée des textes est le mot graphique.

L'adoption généralisée de l'approche extensive en morphologie est freinée par le fait que la collecte des exemples ne peut pas être réalisée sans un minimum d'outils et de compé-

13. `w3.erss.univ-tlse2.fr/trifouillette/`
14. `www.exalead.com`

tences en linguistique informatique. Pour aider à lever ce verrou, j'ai publié, en collaboration avec Ludovic Tanguy, *Perl pour les linguistes*, un guide pratique à l'usage des linguistes qui veulent exploiter des données langagières (Tanguy et Hathout 2007). Ce livre fait suite à l'école thématique sur « la constitution et l'analyse de corpus et de lexiques pour la recherche en morphologie » que j'ai organisée en juillet 2003 en collaboration avec Fiammetta Namer dans le cadre du GDR de morphologie. L'ouvrage est à la fois un manuel permettant d'acquérir des connaissances élémentaires d'informatique et de programmation en Perl et un livre de « recettes » qui propose des programmes prêts à l'emploi que l'on peut utiliser tels quels ou adapter pour des tâches particulières. Ce livre dispose d'un site Web d'accompagnement qui contient un environnement de travail, l'ensemble des programmes et des exemples de données (corpus et lexiques) qu'ils permettent de manipuler [15].

1.3.3 Description morphologique

J'ai également réalisé la description extensive de deux affixations. La première, menée en collaboration avec Marc Plénat et Ludovic Tanguy, a porté sur la sémantique de la suffixation en *-able* et plus précisément sur la relation qui s'établit entre l'adjectif construit et son nom recteur (Hathout et al. 2003). Nous y avons montré que cette relation ne peut pas être caractérisée en termes de transformation syntaxique à partir de l'équivalent propositionnel du syntagme nominal, par exemple de (5b) pour (5a). On trouve en effet un grand nombre de noms recteurs correspondant à des sujets (6) et à des compléments indirects introduits par les prépositions *à* (7a), *de* (7b), *sur* (7c), *pour* (7d), etc.

(5) a. *un poisson pêchable*

b. *Max pêche un poisson.*

(6) *Le terme de proche s'entend par les personnes non* ***adhérables*** *à l'AGPM et à l'AGPM Familles et qui bénéficient du parrainage préalable d'un adhérent de l'AGPM ou de l'AGPM Familles nominativement identifié* `www.agpm.fr/pdf/groupe/statuts_agpm_familles.pdf`

(7) a. *Le premier PC «* ***parlable*** *». On pourra maintenant causer à son ordinateur.* `www2.canoe.com/techno/nouvelles/archives/1996/09/1996 0926-162324.html`

b. *Soit tout risque de devenir sacré et impraticable. Soit tout devient profane,* ***usable*** *et* ***abusable****.* `rad2000.free.fr/glosds01.htm`

c. *À quelques sandwiches près, l'image qu'il offre n'est pas si éloignée de la réalité, à tout le moins de la réalité «* ***enquêtable*** *».* `www.homme-moderne.org/plpl/l0207/41.html`

d. *C'est pourquoi la démarche islamiste "est théoriquement* ***plaidable***". `www.archipress.org/ts66/forum.htm`

Il n'est pas davantage possible de décrire la suffixation en *-able* en termes de contraintes sur les rôles thématiques. On observe en effet que les noms recteurs de ces adjectifs peuvent recevoir l'ensemble des rôles présents dans la structure argumentale du verbe de base, à

15. `perl.linguistes.free.fr`.

l'exception d'agent. Nous avons aussi mis au jour un grand nombre d'attestations dans lesquelles le noms recteur correspond à un circonstant de lieu (8a), de temps (8b), de cause (8c), etc. et non à un argument du verbe base. Plus intéressant encore est le fait que certains noms recteurs ne correspondent ni à des arguments, ni à des circonstants mais à des compléments de nom d'un argument ou d'un circonstant. En (9) par exemple, *consistance* est complément de *béton*, l'objet du verbe *verser*.

(8) a. *On met 10mn à monter la tente, 50mn à enfoncer les sardines... Jamais vu un terrain de camping aussi peu '**sardinable**'.* perso.club-internet.fr/belyne/corse%20III.htm

b. *Ici, la période **navigable**, est de trois mois plus longue que dans la rive nord de la méditerranée.* www.cap-tunisie.com/HTML/port.htm

c. *pis finalement, c'est plus une situation **pleurable** quand j'y pense que quand ça arrive.* 66.46.177.46/zerotom/archives/005804.php

(9) *La versatilité de Sikadur 42 Grout Pak Multi-Flo permet un ratio agrégat/résine de 6:1 pour une consistance **versable**, 5:1 pour une consistance autonivelante.* 207.61.100.79/fr/constr/cat/Division3/03600_Coulis/CoulisABaseDeResineReactive/Sikadur42GroutPakMulti-Flo_F_03600.pdf

Nous avons établi que la sémantique du suffixe *-able* est relativement abstraite, ce qui induit une sous-spécification sémantique des adjectifs construits et une grande multiplicité dans leurs interprétations. Un adjectif comme *pêchable* peut ainsi se trouver dans une dizaine de relations sémantiques différentes avec ses noms recteurs (Hathout et al. 2003:76). Enfin, nous avons observé qu'il n'existe pas de contrainte catégorielle sur les bases et avons identifié cinq nouvelles séries d'adjectifs dénominaux en *-able* en plus des quatre habituellement reconnues (Gawelko 1975, Plénat 1988).

J'envisage de reprendre cette étude de la suffixation en *-able* afin d'apporter une réponse à la question, toujours en suspens, de l'impossibilité pour le nom recteur d'un adjectif en *-able* d'être un agent. Cette réponse pourrait être exprimée en termes de la dynamique des forces de Talmy (2000) présentée en section 4.3.1. Comme pour les adjectifs préfixés en *anti-* qui ont une interprétation adversative (voir section 4.3.5), l'analyse des dérivés en *-able* ferait intervenir un agent intentionnel qui n'est pas nommé et qui souhaite accomplir l'action évoquée par la base de l'adjectif. Par exemple, dans le cas de (5a), cette action est celle de pêcher ce poisson. Cet agent exerce ainsi une force antagoniste qui tend à la réalisation de cette action. L'utilisation d'un adjectif en *-able* présente l'entité désignée par le nom recteur, par exemple le poisson en (5a), comme exerçant une force agoniste visant à faire échouer cette action. Dans l'exemple (5a), le poisson exerce une force qui tend à empêcher l'agent de le pêcher. En utilisant un adjectif en *-able*, on indique en plus que la force exercée par l'agent est supérieure à celle exercée par l'entité désignée par le nom recteur, par exemple en (5a) que le poisson ne parviendra pas à empêcher le pêcheur de l'attraper. L'impossibilité d'avoir un agent comme nom recteur découle alors simplement du fait que ce dernier ne peut pas en même temps vouloir la réussite de l'action évoquée par la base et exercer une force qui vise à l'empêcher [16]. Notons que cette situation est différente de celle

16. Je remercie Michel Aurnague de m'avoir suggéré cette analyse.

que l'on trouve dans un exemple comme (10) analysé par Talmy (2000:431-433) où l'on a une opposition psychologique entre deux parties de soi, le désir de répondre qui exerce une force agoniste et l'intériorisation des normes sociales qui exerce une force antagoniste.

(10) *Il s'est abstenu de répondre.*

La deuxième description morphologique extensive que j'ai réalisée porte sur la préfixation en *anti-*. Elle fait l'objet du chapitre 4. Elle reprend une étude menée en collaboration avec Marc Plénat et Olena Levkovych que nous avons présentée aux 4es Décembrettes (Plénat et al. 2004a).

1.3.4 Productivité morphologique

La morphologie extensive se distingue nettement de ce que l'on pourrait considérer comme une morphologie de corpus à laquelle on pourrait rattacher par exemple les travaux sur la productivité morphologique réalisés au sein de l'action « bases de données constructionnelles et productivité des procédés morphologiques en corpus » du GDR de morphologie. J'ai réalisé dans ce cadre, en collaboration avec Bernard Fradin et Fanny Meunier, une étude de la suffixation en *-et* et en *-ette* à partir d'un corpus composé de 5 années d'archives du journal *Libération* (Fradin et al. 2003b). Une base de données de 1 145 noms a été constituée. Notons que ce chiffre est très faible en comparaison du nombre d'exemples qu'une étude extensive pourrait rassembler. Plénat (2005) a par exemple rassemblé un corpus de 550 noms déverbaux en *-ette* lorsque le corpus de Fradin et al. (2003b) n'en contient qu'une trentaine, soit un rapport de l'ordre de 20 pour 1. Les 1 145 noms ont été analysés et répartis en 18 sous-séries sémantico-référentielles [17] en fonction de deux critères : la dénotation du dérivé et la relation qu'il entretient avec sa base. Une comparaison de la productivité de la suffixation en *-et* et *-ette* dans ces différentes sous-séries a été réalisée. Nous avons aussi comparé la productivité de la suffixation en *-et* à celle de la suffixation en *-ette*. Les conclusions de cette étude sont qu'il existe une grande variabilité de la productivité de *-et* et *-ette* en fonction des sous-séries et que *-ette* est deux fois plus productif que *-et*. Ce travail a été poursuivi par une étude psycholinguistique qui a confirmé l'existence de plusieurs patrons dérivationnels (Meunier et al. 2007).

J'ai également participé aux travaux de l'action « aspects de la productivité morphologique » du GDR de morphologie qui visent à créer une cartographie de la productivité morphologique d'un ensemble d'affixes du français, notamment *-able*, *-oir*, *-ion*, *-iser*, *-ifier*, *-ité*, *-iser*, *-if*, *-ique* (Grabar et al. 2006a, b). Dans le cadre de cette collaboration, nous avons mis au point une grille d'analyse commune qui distingue différentes catégories de lexèmes suffixés. La productivité morphologique de ces affixes a été calculée dans différents corpus d'articles du journal *Le Monde*.

Les mesures statistiques définies par Baayen (1992) pour estimer la productivité d'un procédé morphologique doivent nécessairement être calculées dans des corpus de taille réduite car il est nécessaire d'y relever tous les mots construits par le procédé. En revanche, ce type de morphologie de corpus n'est pas adaptée à la description des procédés morphologiques du fait de leur forte plasticité sémantique et de la dispersion des créations lexicales.

17. J'emprunte ce terme à Roché (2009, en préparation).

1.3.5 Diffusion et réutilisation des données

La morphologie extensive augmente également l'urgence d'un problème pour lequel il n'existe pas encore de réponse satisfaisante, ni pour le français ni pour aucune autre langue : la conservation, la réutilisation et la complétion des corpus d'exemples. La collecte et le traitement des données absorbe en effet une partie importante de l'activité de recherche des morphologues extensifs, mais l'étendue de ce travail est difficile à faire reconnaître pleinement et à rentabiliser. Elle n'apparaît en effet qu'en filigrane dans les publications que les données extensives permettent de rédiger, à travers les analyses d'une portion infime des exemples rassemblés. Or, relativement à l'effort réalisé, le travail supplémentaire qui serait nécessaire pour préserver ces collections et permettre leur diffusion éventuelle est faible.

La difficulté est de deux ordres : la mise en forme des données et leur mode de diffusion. Elle pourrait être résolue s'il existait un outil simple et facile à utiliser destiné à la description et l'exploitation des corpus d'exemples en morphologie. Cet outil pourrait prendre la forme d'un service Web similaire aux applications Google Docs [18] permettant de décrire les données de façon souple et cumulative. Il serait utilisé par les morphologues dès le début de leur travail de collecte et de vérification. Pour faciliter son adoption, un tel système pourrait fournir à l'utilisateur une partie des exemples dont il a besoin, notamment ceux qui figurent dans les dictionnaires.

J'envisage de travailler, en collaboration avec Gilles Boyé, sur un outil de ce genre. De type Wiki, il disposera d'un canevas formé par les familles et les séries présentées en section 2.6. Ce canevas pourra être construit en utilisant les techniques décrites dans le chapitre 3. Cette proposition se distingue de la plupart des projets antérieurs de création de bases d'exemples morphologiques souvent conçus dans le but de fusionner des bases existantes. Il faut au contraire, selon nous, créer un système destiné aux études nouvelles. Les collections d'exemples existantes pourront y être ultérieurement intégrées, si cet outil s'impose comme une base fiable et pratique.

1.4 Conclusion

Les chantiers présentés dans la suite du mémoire sont les prolongements naturels de mes travaux antérieurs. Le premier, qui est l'objet du chapitre 3, est la poursuite de la caractérisation de la structure morphologique du lexique dans le cadre de la création d'une base de données morphologiques à large couverture. Le second chantier, illustré par le chapitre 4, est la description morphologique dans un cadre extensif et DUMALien (Roché et al. 2010). Il m'est également apparu nécessaire de replacer mes recherches dans un contexte théorique général dont quelques-uns des éléments sont présentés dans le chapitre 2.

18. `docs.google.com`.

CHAPITRE 2

Un modèle théorique de la morphologie lexicale

2.1 Introduction

Le développement d'un modèle théorique de la morphologie lexicale n'est pas un objectif premier dans mes recherches, mais une réponse à un double besoin. Le premier est de disposer d'un cadre théorique et d'un ensemble d'outils adaptés aux analyses morphologiques dérivationnelles (chapitre 4). Le second est d'inscrire le paradigme d'analyse morphologique automatique que je propose (chapitre 3) dans un cadre théorique permettant d'identifier les objectifs à atteindre. Il ne s'agit donc pas ici de décrire un modèle complet permettant de rendre compte de tous les phénomènes de la morphologie dérivationnelle, mais plutôt de proposer un cadre général dont certains aspects font l'objet d'une élaboration plus détaillée. Leur mise au point générale ayant été réalisée en vue de l'analyse de la préfixation en *anti-*, j'utiliserai principalement cette dérivation pour les illustrer.

Dans un modèle par règles comme celui de Corbin (1987, 1991), le dispositif est simple : on dispose d'un ensemble d'*inputs* qui permettent de produire des *outputs* au moyen de règles de construction de mots (RCM). Mais quel est la nature d'un tel dispositif dans un modèle lexématique sans RCM où il n'y a que des *outputs* et où l'*input* est la totalité du lexique ? Un début de réponse consiste à mettre en évidence l'existence de paradigmes morphologiques comme le font par exemple Blevins (2003, 2006) ou Booij (2007, 2008). Ces auteurs s'attachent à montrer les avantages qu'apportent ces paradigmes dans l'analyse des phénomènes morphologiques mais ne disent rien du cadre dans lequel ils s'insèrent. Une exception notable à cette tendance est Roché (en préparation). Reconnaître le rôle des paradigmes en morphologie est une première étape importante, mais quel est le statut de ces paradigmes ? Comment sont-ils utilisés ? Dans quel système s'insèrent-ils ? Comment passer d'un système capable de découvrir des paradigmes à un système capable de prédire des formes nouvelles et de tenir compte des variations comme les allomorphies et les suppléions ? C'est à ces questions que ce chapitre tente d'apporter quelques éléments de réponse.

2.2 Principes généraux

1. Le cadre théorique que je propose peut être vu comme un passage d'une morphologie analytique à une morphologie topologique centrée sur les relations qui existent entre les mots et sur les distances qu'elles induisent, la caractérisation précise de chaque instance de ces relations devenant secondaire. Ce cadre est basé sur l'usage. Il peut être rapproché des modèles proposés par Jackendoff (1975), Bybee (1985, 1988, 1995), Neuvel et Singh (2001), Burzio (2002), Gaume (2004) ou Pirrelli (2007).

2. Je considère qu'il existe une distinction claire entre morphologie et lexique :
   – la morphologie est un ensemble de moyens à la disposition des locuteurs qui permettent d'exprimer des sens à l'aide de mots [1] (en production [2]) et de retrouver des sens à partir de mots (en compréhension) ;
   – le lexique est un espace dédié à la mémorisation des mots, des lexèmes et des relations qui s'établissent entre eux.

Morphologie et lexique sont néanmoins intimement liés : les mots construits par la morphologie [3] sont normalement destinés à entrer dans le lexique ; le lexique fournit à la morphologie l'essentiel des informations dont elle a besoin pour caractériser le sens et la forme des mots.

3. Je considère qu'un mot est caractérisé par son sens, sa forme et sa catégorie :
   – Un mot est dérivé si son sens est construit, c'est-à-dire s'il est défini relationnellement (Corbin et Temple 1994).
   – La forme d'un mot est une collection d'indices qui signalent un ensemble de relations entre ce dernier et d'autres mots du lexique.
   – La catégorie d'un mot est un ensemble de traits morphosyntaxiques qui dépendent du contexte dans lequel le mot est utilisé.

4. Je considère que la détermination de la forme des mots (en production) ou de leur sens (en compréhension) est réalisée individuellement pour chaque occurrence. La morphologie opère ainsi au même niveau que la syntaxe, à savoir celui de la production langagière et de son interprétation. Elle n'opère pas au niveau du lexique mais a accès à l'ensemble des objets lexicaux permettant de faire référence, et notamment aux formes fléchies, aux locutions, aux expressions usuelles, etc.

5. Je considère qu'il n'existe qu'une morphologie qui a en charge à la fois la flexion et

1. J'utilise le terme de « mot » pour désigner les mots-occurrences qui apparaissent dans les productions langagières.

2. Le terme de « production » (resp. « compréhension ») a ici le sens de production de mots dans le cadre de la production d'énoncés (resp. compréhension des mots dans le cadre de la compréhension des énoncés). La production des mots concerne à la fois les mots déjà mémorisés dans le lexique et ceux qui ne le sont pas encore. J'utiliserai le terme de construction pour parler de la formation de mots morphologiquement complexes. Je signale que cette formation ne présuppose pas que le mot construit est nouveau, i.e. absent du lexique mémorisé.

3. Il s'agit ici d'un abus de langage. Ce sont les locuteurs et non la morphologie qui construisent les mots. L'expression « les mots construits par la morphologie » doit donc être comprise comme un raccourci de « les mots construits par les locuteurs en utilisant des moyens morphologiques ».

la dérivation. Je considère aussi que la morphologie est sollicitée chaque fois qu'un mot est utilisé, y compris lorsque ce dernier est déjà mémorisé dans le lexique.

6. Je considère que la forme d'un mot construit est principalement déterminée par le lexique existant et ses structures et par un ensemble de contraintes concurrentes, en suivant (Plénat 2000). Ces contraintes n'entrent cependant pas dans le cadre de la théorie de l'optimalité (Prince et Smolensky 1993, McCarthy et Prince 1993) car (*i*) elles ne sont pas hiérarchisées, et (*ii*) elles sont gradables. Par ailleurs, je ne m'intéresse pas à leur universalité.

7. Le modèle que je propose est non compositionnel : les dimensions sémantique et formelle des mots y donnent lieu à des traitements indépendants. Les contraintes qui déterminent la forme ne portent pas sur le sens, et, réciproquement, les mécanismes qui déterminent le sens ne prennent pas en compte la forme. L'articulation entre forme et sens se situe uniquement au niveau du lexique.

2.3 Un modèle à quatre niveaux

La morphologie opère sur quatre niveaux de représentation : un niveau sémantique, un niveau formel, un niveau catégoriel et un niveau lexical qui supporte l'organisation morphologique. Le niveau lexical est conçu comme un espace discret composé de **positions**. Ces positions remplissent dans le modèle une fonction similaire à celle des lexèmes ou des unités lexicales. Une position est une cellule dans laquelle sont mémorisés des mots décrits entre autres par un sens, une forme, une catégorie et une fréquence. Les mots qui partagent une même position sont ordonnés en fonction de leur fréquence. Les plus fréquents sont les plus accessibles. Celui qui est premier définit les **valeurs par défaut** de la position : son sens (resp. sa forme et sa catégorie) par défaut. Par ailleurs, les mots contenus dans les positions sont connectés les uns aux autres par des relations sémantiques, formelles et catégorielles (Bybee 1985, 1988, 1995). Chaque occurrence de chaque mot est donc mémorisée avec l'ensemble des relations qui la caractérisent. Le niveau lexical mémorise aussi ces relations et leurs fréquences. Ces relations varient selon deux paramètres : (*i*) leurs forces que l'on pourrait estimer par le nombre moyen de traits partagés par les mots qu'elles connectent ; (*ii*) leurs fréquences que l'on pourrait estimer par le nombre de couples de mots entre lesquels elles s'établissent. Les relations entre les mots induisent des relations entre les positions qui les contiennent. Ces relations varient selon les mêmes paramètres de force et de fréquence. Pour prendre une image, le niveau lexical est une sorte de filet multidimensionnel dont les nœuds sont plus ou moins gros selon qu'ils sont plus ou moins fréquents et dont les fils seraient élastiques. Ces fils seraient constitués de brins plus ou moins nombreux en fonction de la fréquence de la relation et plus ou moins courts en fonction du nombre de traits partagés. Certains segments seraient ainsi plus rigides (i.e. plus gros et plus courts) que d'autres. Par ailleurs, ces relations s'organisent en paradigmes, c'est-à-dire que les relations de même type s'orientent toutes de la même manière. Elles renforcent ainsi localement la rigidité du lexique dans cette direction, et par suite la leur (i.e. leur propre rigidité).

L'une des caractéristiques du modèle proposé est que les niveaux sémantique, formel et

catégoriel ne sont pas directement en correspondance. Ils le sont seulement par l'intermédiaire du niveau lexical. Ces correspondances peuvent être représentées schématiquement comme en figure 2.1. L'objectif de la morphologie est de trouver les meilleures correspondances possibles entre ces quatre niveaux. Pour ce faire, les correspondances sont soumises à un système de contraintes permettant de sélectionner celles qui offrent la coïncidence optimale entre sens, positions lexicales, formes et catégories. Le niveau lexical est donc l'équivalent du niveau morphomique proposé par (Aronoff 1994:25).

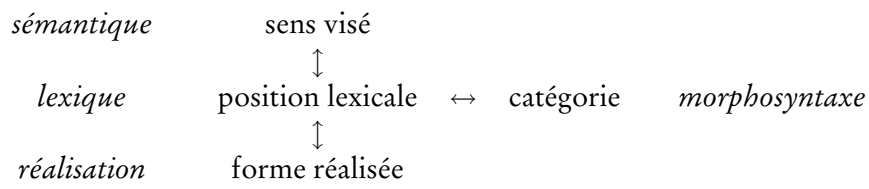


FIGURE 2.1 – Un modèle à quatre niveaux. Le niveau sémantique est composé de sens, le niveau formel de formes, le niveau catégoriel de traits morphosyntaxiques et le niveau lexical de positions lexicales. Chaque niveau possède sa propre structure. Les structures des quatre niveaux sont relationnelles. Les relations à chaque niveau définissent une mesure de proximité. La morphologie met en correspondance les objets qui composent les quatre niveaux en préservant au mieux les proximités qui existent au sein de chaque niveau. Les positions lexicales servent d'interface entre les formes des mots (formes réalisées), leurs propriétés morphosyntaxiques (catégories) et les sens qui leur sont attribués (sens visés).

La morphologie est utilisée en production et en reconnaissance. Dans le premier cas, elle est conçue comme un prestataire de service dont le rôle est d'exprimer un sens à l'aide d'un mot, lorsqu'il existe un moyen d'y parvenir. Pour être exprimé, ce sens, que j'appellerai **sens visé**, doit être apparié à une position p dans le lexique qui peut être soit occupée, soit encore vide, mais cette distinction n'est pas essentielle. La position p ne peut pas être située n'importe où dans l'espace lexical. Seuls les emplacements qui peuvent entrer dans les structures paradigmatiques existantes sont autorisés pour les sens construits. En conséquence, le lexique est un **espace discret**, ce qui distingue ce modèle de celui de Victorri (1994), Ploux et Victorri (1998). La morphologie détermine ensuite la forme la plus appropriée à l'identification de p, tout en tenant compte du contexte particulier dans lequel le mot est utilisé et notamment des propriétés morphosyntaxiques qu'il détermine. Lorsque p contient des mots mémorisés, elle exerce une attraction qui favorise la réutilisation de l'une des formes de ces mots (Burzio 2002). Si p est vide, la morphologie a le choix entre l'utilisation d'une forme existante, i.e. de la forme d'un mot mémorisé qui occupe une position proche de p, ou la création d'une forme nouvelle dont les propriétés définissent les mêmes proximités avec les formes des mots des positions proches de p que celles de ces positions avec p. En reconnaissance, les correspondances sont établies en sens inverse. La première étape consiste à apparier la forme réalisée et les propriétés morphosyntaxiques de l'occurrence avec une position p dans le lexique. Cette position peut être vide ou déjà

occupée. Dans un second temps, p est mise en correspondance avec un sens visé en tenant compte des autres éléments du contexte d'utilisation du mot. Ce sens peut être celui d'un mot mémorisé dans p si p en contient, celui d'un mot qui se trouve dans une position voisine, ou un sens nouveau dont les propriétés définissent les mêmes proximités avec les sens des mots des positions proches de p que celles de ces positions avec p. La morphologie est donc aussi un prestataire de service spécialisé dans l'identification du sens exprimé par les mots.

L'architecture illustrée par la figure 2.1 s'inspire du modèle à deux niveaux de Koskiennemi (1983) notamment par l'existence de correspondances entre les niveaux et par le fait que ces dernières sont contrôlées par des contraintes. Le modèle à deux niveaux est un cadre formel et informatique dans lequel la morphologie opère sur deux types de représentations, formelles et lexicales. Les représentations formelles décrivent les réalisations qui peuvent être décrites par des graphies (par exemple, `dérivation`) ou par des transcriptions phonétiques. Les représentations lexicales sont des concaténations de morphèmes qui satisfont à des contraintes morphotactiques élémentaires (par exemple `dériv+ion`). Les deux types de représentations sont composés de symboles, formels dans un cas et lexicaux dans l'autre. Les symboles qui composent les représentations formelles et lexicales de chaque mot sont en correspondance un à un. Si nécessaire, des symboles vides `0` sont introduits (par exemple, `dérivat0ion` : `dériv00+ion` où `a`, `t` et `+` sont en correspondance avec des symboles vides). Les correspondances entre les deux représentations sont contrôlées par des règles à deux niveaux destinées principalement à rendre compte des allomorphies. Ces règles remplissent un rôle similaire aux contraintes présentées dans les sections 2.8 à 2.11. Le modèle que je propose s'éloigne néanmoins de celui de Koskiennemi sur plusieurs points : (*i*) le cadre général est purement lexématique ; (*ii*) le niveau sémantique est distinct du niveau lexical ; (*iii*) le modèle comporte une mesure de proximité entre les mots ; (*iv*) les correspondances préservent au maximum la structure topologique qui existe dans chacun des quatre niveaux.

Le modèle à quatre niveaux peut également être rapproché du modèle connexionniste de Seidenberg et Gonnerman (2000) et de Plaut et Gonnerman (2000) qui comporte lui aussi un niveau intermédiaire « caché » destiné à représenter les régularités statistiques qui existent entre trois niveaux d'entrée-sortie : un niveau sémantique, un niveau phonologique et un niveau orthographique. Ces quatre niveaux sont implémentés sous forme de réseaux de neurones. Les unités du niveau caché sont connectées à celles des trois autres. En revanche, il n'y a pas de connexions directes entre les unités des trois niveaux d'entrée-sortie.

L'idée d'établir des correspondances entre un espace de formes caractérisées par leurs propriétés phonologiques et un autre niveau de représentation, qui en plus préservent les distances entre les mots est à la base du modèle développé par Vito Pirrelli et ses collaborateurs (Pirrelli et al. 2004, Pirrelli et Herreros 2007, Pirrelli 2007, Calderone et al. 2007). Dans ce modèle, la morphologie flexionnelle de langues comme l'italien, l'anglais ou l'arabe est représentée au moyen de cartes auto-organisatrices (*Self Organizing Maps*) de Kohonen (1995). Ces cartes sont des algorithmes neuronaux qui permettent de réaliser une projection non linéaire d'un espace de grande dimension sur un espace bidimensionnel afin de faire apparaître les proximités topologiques entre les données et les zones denses (i.e. les agrégats d'éléments).

L'existence d'un niveau lexical intermédiaire entre catégorie, forme et sens apporte au modèle un degré de liberté qui lui permet de rendre compte de la plasticité catégorielle et sémantique et de la polyvalence des formes. La plasticité catégorielle repose sur la continuité qui existe entre de nombreuses catégories comme les noms et les adjectifs (comme on peut le voir par exemple dans *une ferme modèle*) ou les participes verbaux et les adjectifs (par exemple dans *un animal blessé* et *un animal qui a été blessé*). La plasticité sémantique des mots peut par exemple être illustrée par les cinq sens de *farfouillette* (1) listés par Plénat (2005) ou par ceux de *russe* (2) présentés dans Roché (2008). Réciproquement, ce degré de liberté permet de rendre compte de l'utilisation d'une même forme pour exprimer des sens différents comme *anti-Sarkozy* dans les exemples (3) et (4). *Anti-Sarkozy* est habituellement utilisé pour désigner une opposition à la politique ou aux idées de Nicolas Sarkozy (3). Mais dans un contexte comme (4), ce mot signifie « qui est tout le contraire d'un Sarkozy ».

(1) a. 'action de farfouiller' ;
b. 'personne qui farfouille' ;
c. 'instrument pour farfouiller' ;
d. 'endroit où l'on farfouille' ;
e. 'ce que l'on farfouille'

(2) a. 'nom du peuple' ;
b. 'gentilé de Russie' ;
c. 'relatif au peuple russe' ;
d. 'relatif à la Russie' ;
e. 'relatif aux citoyens de la Fédération de Russie' ; etc.

(3) *La mobilisation **anti-Sarkozy** continue sur la Toile.* `www.lemonde.fr/cgi-bin/ACHATS/acheter.cgi?offre=ARCHIVES&type_item=ART_ARCH_30J&objet_id=987148&clef=ARC-TRK-NC_01`

(4) *Juppé, c'est l'**anti-Sarkozy**.* `pascaldecaillet.blog.tdg.ch/archive/2008/03/10/juppe-l-anti-sarkozy.html`

2.4 Proximité morphologique

La force des relations qui s'établissent entre les mots a été décrite de manière imagée ci-dessus comme la longueur des segments d'un filet lexical. Les segments courts connectent des positions proches. Cette notion peut être définie de façon formelle comme une mesure de proximité morphologique. Elle peut être implémentée comme je le présente dans le chapitre 3 et dans Hathout (2008, 2009a, b)[4]. Cette mesure peut être décrite de la manière

4. Dans cette implémentation, la mesure n'est pas symétrique. C'est en effet davantage une mesure d'accessibilité qu'une mesure de proximité. Pour obtenir une mesure symétrique, il suffit de remplacer les valeurs de proximité $m(x, y)$ par la moyenne $\frac{m(x,y) + m(y,x)}{2}$.

suivante (D2.1)[5] :

Définition 2.1 PROXIMITÉ MORPHOLOGIQUE

a. Deux formes qui partagent à la fois des propriétés sémantiques et formelles sont morphologiquement proches.

b. La proximité morphologique de deux mots est d'autant plus grande que le nombre des propriétés sémantiques et formelles qu'ils partagent est grand.

c. La proximité morphologique de deux mots est d'autant plus grande que les propriétés qu'ils partagent leur sont spécifiques, c'est-à-dire que peu d'autres mots en sont munis.

Cette mesure de la proximité morphologique est globale : elle prend en compte la totalité du lexique. Inspirée des travaux de Gaume (2004) et Gaume et al. (2002, 2006), elle peut être comparée à la mesure proposée par Jackendoff (1975) pour estimer le coût de l'information ajoutée par une nouvelle entrée lexicale. Elle peut par ailleurs être restreinte aux seules propriétés formelles (resp. sémantiques) afin de définir une mesure de proximité formelle (resp. sémantique). Ces trois mesures induisent chacune une topologie sur le niveau de représentation dans lequel elles sont définies. Notons qu'une mesure similaire peut aussi être définie pour le niveau catégoriel.

Les propriétés sémantiques et formelles qui servent à déterminer D2.1 doivent être maximalement redondantes afin de capter le plus grand nombre possible de similarités qui existent entre les mots. Pour une langue comme le français, on peut par exemple utiliser comme propriétés formelles d'un mot donné l'ensemble des séquences sonores qui apparaissent dans sa forme. Un mot comme *atomique* sera ainsi décrit par les traits listés en (5) où # marque le début et la fin du mot[6].

(5) /#atɔmik#/,
/#atɔmik/, /atɔmik#/,
/#atɔmi/, /atɔmik/, /tɔmik#/,
/#atɔm/, /atɔmi/, /tɔmik/, /ɔmik#/,
/#atɔ/, /atɔm/, /tɔmi/, /ɔmik/, /mik#/,
/#at/, /atɔ/, /tɔm/, /ɔmi/, /mik/, /ik#/,
/#a/, /at/, /tɔ/, /ɔm/, /mi/, /ik/, /k#/,
/a/, /t/, /ɔ/, /m/, /i/, /k/

Pour une langue gabaritique comme l'arabe, on peut utiliser des séquences similaires en remplaçant les représentations phonétiques par des schèmes consonantiques et vocaliques (Pirrelli et Federici 1994, Pirrelli et Herreros 2007). Par exemple, les propriétés associées à la forme /maktabun/ 'bureau' seraient alors les séquences qui apparaissent dans les patrons /#mktb#/ et /#CaCCaCun#/ où C représente une consonne. La finale /un/ étant une marque de cas, elle est traitée comme un élément vocalique. Ces deux patrons peuvent être obtenus

5. Dans la suite du mémoire, les numéros des définitions seront précédés de la lettre « D » et ceux des contraintes par la lettre « C ».

6. Cette liste peut être étendue davantage encore en ajoutant des traits dans lesquels certains des symboles seraient non spécifiés. Cette extension pourrait être réalisée pour tous les traits de (5). Par exemple, à partir de /tɔm/, on ajouterait /_ɔm/, /t_m/, /tɔ_/, /__m/, /_ɔ_/, /t__/ où « _ » représente un symbole joker.

simplement à partir de la graphie vocalisée du mot (6) ; le /un/ final est noté par le dernier diacritique vocalique.

(6) مَكتَبٌ

Le sens étant abstrait, il est plus difficile de connaître précisément les objets qui le composent. Si l'on considère qu'il est de nature relationnelle, les traits sémantiques que l'on pourrait associer à un mot seraient des ensembles redondants de sens connexes (par exemple, des sous-ensembles de la liste (7a) dans le cas de *atomique*), de relations sémantiques (par exemple, des sous-ensembles de (7b)) et de couples (relation sémantique, sens connexe) (par exemple, des sous-ensembles de (7c)).

(7) a. ATOME NUCLÉAIRE BOMBE RADIATION ÉLECTRON etc.
b. 'relatif à' synonyme 'sert à fabriquer' 'émet' 'dans son orbite' etc.
c. ('relatif à', ATOME), (synonyme, NUCLÉAIRE), ('sert à fabriquer', BOMBE), ('émet', RADIATION), ('dans son orbite', ÉLECTRON), etc.

REMARQUE 1.– La proximité morphologique est maximale entre les mots morphologiquement apparentés qui sont sémantiquement les plus proches (Roché 2009, Dal 2007), notamment entre :

- les verbes et les noms processifs correspondants (*diffuser* : *diffusion*) ;
- les noms et les adjectifs de relation correspondants (*parlement* : *parlementaire*) ;
- les adjectifs et les noms de propriété correspondants (*obèse* : *obésité*) ;
- les adjectifs et les adverbes de manière correspondants (*magnifique* : *magnifiquement*).

REMARQUE 2.– Cette mesure permet par ailleurs de distinguer pour un dérivé donné, par exemple *nationaliser*, la relation qu'il a avec sa base, *national*, de celles qu'il a avec ses propres dérivés comme *nationalisation* ou *renationaliser*. Un mot est moins proche de sa base qu'il ne l'est de ses dérivés simplement parce qu'il partage davantage de propriétés sémantiques et formelles avec les seconds. Par exemple, les propriétés sémantiques et formelles induites par la suffixation en *-iser* sont partagées par *nationaliser* avec *nationalisation* mais sont absentes de *national*. L'intersection de l'ensemble des traits de *nationaliser* et *nationalisation* contient donc strictement l'intersection des traits de *nationaliser* et *national*. Par suite, la proximité de *nationaliser* est plus grande avec *nationalisation* et qu'avec *national*.

REMARQUE 3.– Cette mesure ne tient pas compte de la saillance des positions. Par exemple, dans un triplet (*pluie*, *pluvieux*, *pluviosité*), la relation *pluie* : *pluvieux* contribue plus à la caractérisation du sens de *pluvieux* que *pluviosité* : *pluvieux*.

2.5 Proximité lexicale

La proximité morphologique peut être étendue en une proximité lexicale en y intégrant les propriétés catégorielles :

Définition 2.2 PROXIMITÉ LEXICALE

a. Deux formes qui partagent des propriétés sémantiques, formelles ou catégorielles sont lexicalement proches.

b. La proximité lexicale de deux mots est d'autant plus grande que le nombre des propriétés sémantiques, formelles et catégorielles qu'ils partagent est grand.

c. La proximité lexicale de deux mots est d'autant plus grande que les propriétés qu'ils partagent leur sont spécifiques, c'est-à-dire que peu d'autres mots en sont munis.

Les propriétés catégorielles sont des traits morphosyntaxiques qui décrivent la partie du discours et, en fonction de cette dernière, le genre, le nombre, le cas, la personne, le temps, le mode, etc. Par exemple, les traits élémentaires associés aux mots *chat* et *dort* pourraient être ceux présentés en (8). Ces traits élémentaires permettent d'associer à chaque mot des traits complexes, en l'occurrence des ensembles redondants de traits élémentaires comme par exemple (9) pour le nom *chat*.

(8) a. *chat* : nom ; masculin ; singulier

b. *dort* : verbe ; 3^{e} personne ; singulier ; présent ; indicatif

(9) {nom, masculin, singulier} ; {nom, masculin} ; {nom, singulier} ; {masculin, singulier} ; {nom} ; {masculin} ; {singulier}

La contribution des propriétés catégorielles à la proximité lexicale est faible car ce sont les traits les moins spécifiques.

2.6 Organisation morphologique du lexique

Les relations formelles, sémantiques et catégorielles qui s'établissent entre les positions du lexique définissent un ensemble de structures paradigmatiques qui sont utilisées par la morphologie pour analyser les mots et pour en construire de nouveaux (Blevins 2003). Ces paradigmes sont interconnectés les uns avec les autres par des relations analogiques. Ils définissent un maillage qui est spécifique au niveau de représentation lexical et qui renforce l'accessibilité morphologique (D2.1).

2.6.1 Paradigmes morphologiques et paradigmes lexicaux

Les paradigmes qui composent l'organisation morphologique du lexique peuvent être distingués selon trois dimensions : (*i*) famille *vs* série en fonction de leur taille et de leur cohésion ; (*ii*) flexion *vs* dérivation en fonction de la nature des relations sémantiques qui s'établissent entre leurs éléments ; (*iii*) morphologie *vs* lexique. Ces trois dimensions identifient un groupe de paradigmes privilégiés notamment les lexèmes, les séries flexionnelles lexicales, les familles dérivationnelles morphologiques, les familles dérivationnelles lexicales, les séries dérivationnelles morphologiques et les séries dérivationnelles lexicales.

On peut définir les familles morphologiques comme des ensembles de mots qui sont très proches les uns des autres et qui partagent le plus grand nombre possible de propriétés sémantiques et formelles les plus spécifiques possible. De même une série morphologique

est un ensemble aussi grand que possible de mots qui partagent des propriétés morphologiques et catégorielles très générales et qui participent à un nombre aussi grand que possible d'analogies intégrant d'autres mots de la série.

La distinction entre morphologie et lexique se retrouve au niveau de la structure morphologique du lexique :

- Les paradigmes morphologiques rassemblent des ensembles de mots qui entretiennent des relations morphologiques. Par exemple, *modifier* : *modifiable* appartient à un paradigme morphologique car les deux mots partagent en même temps des propriétés de forme et de sens.
- Les paradigmes lexicaux réunissent des éléments qui se trouvent dans des relations qui sont généralement exprimées morphologiquement. Les paradigmes lexicaux étendent les paradigmes morphologiques en y incluant les supplétions. Par exemple, *prison* : *carcéral* appartient à un paradigme lexical qui contient également *commerce* : *commercial*, *université* : *universitaire*, *périphérie* : *périphérique*, etc. *Prison* et *carcéral* ne sont pas morphologiquement apparentés parce qu'ils ne partagent pas de propriétés formelles. Néanmoins, la relation sémantique qui les unit, relation entre un nom et son adjectif de relation, est normalement exprimée en utilisant un procédé morphologique, par exemple les suffixations en *-al*, *-el*, *-aire*, *-ien*, *-éen*, *-ique*, etc. comme on peut le voir dans les exemples ci-dessus. Le paradigme lexical est ainsi caractérisé par un rapport sémantique constant entre les membres des relations qu'il contient.

Il n'existe pas de critère simple permettant de distinguer les relations flexionnelles des relations dérivationnelles. Les relations flexionnelles s'établissent entre des mots de même catégorie (i.e. partie du discours) dont les sens sont strictement interprédictibles (Boyé 2010). Par exemple, la relation entre les formes adjectivales *vert* et *verte* est flexionnelle car le sens de chacune d'elles peut être complètement récupéré à partir de celui de l'autre. La seule différence qui existe entre ces deux mots est la valeur du genre, masculin dans un cas et féminin dans l'autre. Les relations flexionnelles sont d'autre part limitées à un ensemble réduit de propriétés morphosyntaxiques. Par ailleurs, tous les couples du paradigme qui contient *vert* : *verte* présentent exactement la même différence. Les relations flexionnelles se distinguent aussi par la systématicité quasi totale des paradigmes dans lesquels elles s'insèrent. Cette rigidité absolue des relations flexionnelles les distingue des relations dérivationnelles qui sont elles caractérisées par une grande souplesse. Notons cependant que certaines relations dérivationnelles comme celles qui s'établissent entre les noms et les adjectifs de relation correspondants, entre les verbes et les noms processifs correspondants et entre les adjectifs et les noms de propriétés correspondants ou entre les adjectifs et les adverbes de manière correspondants sont relativement proches des relations flexionnelles.

Les **familles flexionnelles morphologiques** regroupent des mots qui partagent deux à deux des propriétés formelles et qui entretiennent des relations flexionnelles. Par exemple, la famille flexionnelle morphologique de *lavons* contient *lavera* et l'ensemble des formes fléchies du verbe *laver*, ce dernier n'ayant pas de radicaux supplétifs [7]. À l'inverse, *allons* a

7. Par radicaux supplétifs, j'entends thèmes au sens de (Bonami et Boyé 2003, Bonami et al. 2009). Par exemple, *écrivons* et *écrira* appartiennent à des familles flexionnelles morphologiques différentes.

dans sa famille flexionnelle morphologique *allaient* mais pas *ira* ni *vont* qui appartiennent à des familles flexionnelles morphologiques différentes. Les familles flexionnelles morphologiques sont les paradigmes les plus denses qui soient car leurs éléments sont maximalement proches les uns des autres du point de vue morphologique. Ces derniers ne diffèrent en effet que sur le plan flexionnel, or les propriétés et les marques flexionnelles sont les caractéristiques les moins spécifiques dans la mesure où elles sont les plus fréquentes.

Les **lexèmes** sont des **familles flexionnelles lexicales** qui regroupent des mots entre lesquels il existe des relations flexionnelles, même s'ils ne partagent pas de propriétés formelles. Ces mots peuvent appartenir à une ou à plusieurs familles flexionnelles morphologiques. Par exemple, le lexème de *lavons* est identique à sa famille flexionnelle morphologique. Celle de *allons*, en revanche, contient les mots des familles flexionnelles morphologiques de *allons*, de *ira* et de *vont*. Les lexèmes sont des paradigmes pratiquement aussi denses que les familles flexionnelles morphologiques.

À l'opposé, les **séries flexionnelles lexicales** et les **séries flexionnelles morphologiques** sont les paradigmes les moins denses. Ils rassemblent des formes qui sont disséminées sur l'ensemble du lexique et qui sont connectées par les relations morphologiques les plus lâches. Ces relations s'établissent entre des formes de lexèmes différents qui portent la même flexion et partagent l'ensemble de leurs propriétés morphosyntaxiques. Par exemple, la série flexionnelle morphologique de *dérivons* contient l'ensemble des formes verbales conjuguées à la 1re personne du pluriel du présent de l'indicatif comme (10). Ces formes partagent la même marque flexionnelle et les mêmes propriétés morphosyntaxiques de temps, de mode et de personne.

(10) *lavons* ; *évitons* ; *verbalisons* ; *mangeons* ; *visons* ; *prenons*

La série flexionnelle lexicale de *dérivons* contient aussi *sommes*, en plus des mots de la série flexionnelle morphologique illustrée en (10).

Les **familles dérivationnelles morphologiques**, ou plus simplement **familles morphologiques**, correspondent à des relations morphologiques intermédiaires qui s'établissent entre des lexèmes dont les formes partagent leurs propriétés les plus spécifiques. Les membres de ces familles partagent donc en même temps des propriétés de forme et de sens. Ces familles correspondent aux paradigmes dérivationnels présentés dans Roché (2008, 2009, 2010b). Par exemple, la famille morphologique de *dérivation* contient des lexèmes comme (11). Les familles sont des ensembles denses dont les éléments sont très proches tant du point vue formel que sémantique. Elles possèdent peu d'éléments. En contrepartie, le lexique contient un grand nombre de familles.

(11) *dérivation* ; *dériver* ; *dérivable* ; *dérivationnel* ; *dérivationnellement* ; *dérivatif* ; *dérive* ; *dériveur* ; *dérivée*

Les **familles dérivationnelles lexicales**, ou plus simplement **familles lexicales**, rassemblent au sein d'un même ensemble les éléments de plusieurs familles morphologiques lexicalement complémentaires. Par exemple, la famille lexicale de *champignon* contient à la fois la famille morphologique de *champignon*, *champignonner*, *champignonneux*, *champignonnière*, etc., celle de *fongus*, *fongique*, *fongicide*, etc. et celle de *mycète*, *mycétique*, *mycotique*, *mycose*, *mycosique*, *mycologie*, etc. Ces trois familles morphologiques sont sé-

mantiquement très proches puisque *champignon*, *fongus* et *mycète* sont par exemple synonymes. Elles sont réunies au sein de la même famille lexicale parce que *fongique* est l'adjectif relationnel de *champignon* et que cette relation sémantique est habituellement exprimée morphologiquement comme dans *robot* : *robotique* ou *relation* : *relationnel*. Dans le cas de *champignon* : *fongique*, la relation n'est pas morphologique. *Champignon* et *fongique* ne peuvent donc appartenir à la même famille morphologique. Cette relation est en revanche lexicale et sémantique, *fongique* occupant dans le lexique la position d'adjectif de relation de *champignon*. C'est en ce sens que ces familles sont lexicalement complémentaires. La situation est similaire pour *champignon* et *mycotique*. La complémentarité des trois familles morphologiques apparaît pour plusieurs autres relations, par exemple avec *fongicide* (resp. *mycocide*) ou *antifongique* (resp. *antimycotique*) qui occupent les positions de dérivés en *-cide* et en *anti-* de *champignon*. Plus généralement, deux familles morphologiques sont réunies au sein de la même famille lexicale si les deux conditions suivantes sont vérifiées :

1. Elles contiennent des éléments maximalement proches sur le plan sémantique. Ces éléments peuvent être par exemple des synonymes (*marseillais* : *phocéen*) ou bien être dans une relation *verbe* : *nom d'action* (*tomber* : *chute*), *nom* : *adjectif de relation* (*oreillons* : *ourlien*) ou *adjectif* : *nom de qualité* (*allemand* : *germanicité*) (Roché 2009).
2. Certains éléments de l'une des familles morphologiques occupent dans le lexique des positions non pourvues dans l'autre. En d'autres termes, l'une des familles fournit à l'autre un radical supplétif.

Ces conditions peuvent naturellement être généralisées à des ensembles de plus de deux familles morphologiques.

Les **séries dérivationnelles morphologiques** correspondent à des relations plus lâches qui s'établissent entre des lexèmes dont les formes partagent des propriétés qui sont non flexionnelles et peu spécifiques. Ces formes partagent notamment leurs propriétés morphosyntaxiques (i.e. elles appartiennent à la même partie du discours). Par exemple, *dérivation* appartient à une série de noms en *-ion* qui inclut *compilation*, *production*, *succion*, *vision*, etc. À l'inverse des familles morphologiques, les séries dérivationnelles morphologiques sont des ensembles de grande taille mais peu denses : leurs éléments sont relativement distants les uns des autres. Le nombre de séries dérivationnelles est très inférieur à celui des familles morphologiques.

Les séries dérivationnelles morphologiques ne contiennent que des mots construits par une même dérivation. Par exemple, *confection* ou *locomotion* n'appartiennent pas à la série dérivationnelle morphologique de *compilation* bien que ce soient des noms d'action finissant en *-ion*. Le lexique contemporain ne contient pas en effet de lexèmes leur permettant de participer à la série des analogies qui rassemble les éléments de la série dérivationnelle morphologique de *compilation*. *Confection* et *locomotion* appartiennent en revanche à la série dérivationnelle lexicale de *compilation*. Cette série contient naturellement aussi les éléments de la série dérivationnelle morphologique de *compilation*. Les **séries dérivationnelles lexicales**, ou **séries dérivationnelles**, généralisent ainsi les séries dérivationnelles morphologiques. Les mots y sont rattachés uniquement sur la base de leur sens et de leur forme. Ces séries sont les paradigmes lexicomorphologiques définis dans Roché (2008, 2009). La

figure 2.2 récapitule les différents types de paradigmes qui ont été présentés. Chacun de ces types est illustré par un couple de l'une de ses instances.

	flexionnelle		dérivationnelle	
famille	*lavons* : *lavera*	*allons* : *ira*	*dériver* : *dérivable*	*prison* : *carcéral*
série	*lavons* : *coupons*	*lavons* : *sommes*	*dérivable* : *variable*	*furieux* : *curieux*
	morphologique	lexicale	morphologique	lexicale

FIGURE 2.2 – Exemples de couples caractéristiques des huit paradigmes proposés. Les paradigmes morphologiques rassemblent des mots qui partagent en même temps des propriétés sémantiques et formelles. Les paradigmes lexicaux étendent les paradigmes morphologiques en relâchant l'obligation de partage de propriétés formelles. Les familles rassemblent des mots qui partagent leurs propriétés les plus spécifiques. Les séries rassemblent des mots qui partagent leurs propriétés les moins spécifiques. Les paradigmes flexionnels rassemblent des mots qui ne diffèrent que par leurs propriétés morphosyntaxiques. Les paradigmes dérivationnels rassemblent des mots dont les différences ne sont pas flexionnelles.

Les familles lexicales que je propose se distinguent nettement des familles synchroniques dérivationnelles de Gruaz (1997a, b) et des familles morphologiques de Corbin (1997). Ces auteurs envisagent en effet les familles essentiellement comme un mode d'organisation lexicographique et non comme une composante fondamentale de la structure morphologique du lexique, leur objectif étant la création de dictionnaires morphologiques, le DISFA pour Gruaz et le DDF pour Corbin. De ce fait, aucun d'eux ne propose de croiser ces familles avec des séries dérivationnelles. Par ailleurs, les familles qu'ils proposent sont constituées à partir de racines, de têtes de famille, de primitifs, comme par exemple *lait* pour la famille de *laitier*. Ma proposition est différente car je considère que les familles et les séries émergent du lexique dans sa globalité et s'agrègent à partir des relations de proximité morphologique qui existent entre les lexèmes et les formes qu'ils rassemblent.

Notons que plusieurs études psycholinguistiques portent sur les effets liés aux paradigmes morphologiques. Ainsi, New et al. (2004) montrent que le lexème (i.e. la famille flexionnelle) intervient dans le traitement des mots qui le composent, notamment via sa fréquence. Ces auteurs ont en effet montré que le traitement des noms singuliers en français dépend de la fréquence cumulée du lexème. Des effets liés aux séries ont également été mis en évidence en néerlandais et en finnois par Bertram et al. (1999, 2000). Les expériences réalisées indiquent que le mode de stockage et la stratégie de reconnaissance des mots dépendent de deux propriétés des séries : (*i*) l'existence de séries homonymes et (*ii*) la productivité. Les conclusions de leurs études sont que les mots qui portent un suffixe productif qui n'a pas un homonyme plus productif sont traités par une stratégie à deux voies (*dual route*) tandis que ceux dont le suffixe n'est pas productif ou qui ont un homonyme plus productif donnent lieu à une stratégie d'appréhension globale (*whole word route*).

Notons aussi que le fait de se placer dans un cadre paradigmatique n'est pas incompatible avec l'utilisation de termes comme « radical », « initiale » ou « finale ». Ces termes

désignent des segments de la forme des mots, considérés comme de simples indices. Ce ne sont ni des objets lexicaux, ni des unités qui auraient une représentation dans la grammaire ou le lexique :

- Par « radical », j'entends la partie de la forme partagée par un mot avec d'autres membres de sa famille morphologique.
- Par « initiale » et « finale », j'entends la partie de la forme qui est partagée par les membres de la série morphologique du mot et qui permet de reconnaître cette dernière.

Les familles et les séries dérivationnelles induisent ainsi une segmentation approximative de la forme des mots qui dégage des signaux destinés à guider l'analyse.

2.6.2 Analogie

Parmi les nombreuses relations lexicales qui s'établissent entre les éléments des paradigmes présentés ci-dessus, les analogies [8] jouent un rôle particulier dans l'organisation morphologique du lexique. Elles sont présentes dans tous les paradigmes, flexionnels, dérivationnels, morphologiques ou lexicaux. Par exemple, *dérivons* et *dérivait* qui appartiennent au même lexème font partie d'analogies comme (12a). De même, au sein de la série flexionnelle de *dérivons*, le couple *dérivons* : *évitons* participe à un ensemble d'analogies qui contient (12b). De façon similaire, *dérivation* et *dériver* qui appartiennent à la même famille morphologique sont impliqués dans une série d'analogies telles que (13a). De la même manière, *dérivation* et *variation* qui se trouvent dans la même série dérivationnelle font partie d'analogies comme (13b).

(12) a. *dérivons* : *dérivait* :: *parlons* : *parlait*
dérivons : *dérivait* :: *finissons* : *finissait*
dérivons : *dérivait* :: *prenons* : *prenait*
dérivons : *dérivait* :: *évitons* : *évitait*

b. *dérivons* : *évitons* :: *dériver* : *éviter*
dérivons : *évitons* :: *dériveras* : *éviteras*

(13) a. *dérivation* : *dériver* :: *navigation* : *naviguer*
dérivation : *dériver* :: *immersion* : *immerger*

b. *dérivation* : *variation* :: *dériver* : *varier*
dérivation : *variation* :: *dérivable* : *variable*

Il existe plusieurs types d'analogies. Les plus contraintes sont les analogies formelles qui doivent être vérifiées par les formes elles-mêmes, par exemple par les graphies comme (12b) (Lepage 2003, Stroppa 2005, Yvon 2006, Stroppa et Yvon 2006). Il existe d'autre part des analogies sémantiques comme (14) où la relation concerne le sens des lexèmes (Gentner 1983). Entre ces deux extrémités, on trouve des analogies dérivationnelles morphologiques qui s'établissent entre des termes qui appartiennent aux mêmes familles morphologiques

8. Pour une présentation générale de l'utilisation de l'analogie comme principe explicatif de faits morphologiques, voir Dal (2003a, 2008).

et aux mêmes séries dérivationnelles comme (13a) qui correspondent aux règles de redondance lexicale de Jackendoff (1975). On trouve aussi des analogies dérivationnelles lexicales comme (15) qui sont des analogies sémantiques dans lesquelles les membres de chaque couple appartiennent à la même famille morphologique. Ce type d'analogie est utilisé par Hathout (2001).

(14) *texte* : *écrire* :: *maison* : *bâtir*

(15) *batterie* : *batteur* :: *violon* : *violoniste*

Les analogies jouent un rôle important dans l'organisation morphologique du lexique car elles connectent les lexèmes aux séries flexionnelles d'une part et les familles dérivationnelles aux séries dérivationnelles de l'autre (voir section 3.3.1). Ces connections permettent de caractériser sémantiquement les séries, les unes par rapport aux autres (Hathout 2003). Les analogies sont également importantes parce qu'elles définissent les modèles constructionnels et les schèmes de nomination présentés dans Skousen (1989, 1992), Becker (1994), Roché (2007, 2009). Ces schèmes correspondent aux relations morphologiques les plus couramment utilisées pour exprimer les sens construits. Chaque schème dispose d'un pouvoir d'attraction qui dépend de sa taille (Jackendoff 1975, Burzio 2002) : une analogie est d'autant plus apte à être utilisée pour former des mots nouveaux qu'elle comporte un grand nombre de quadruplets.

2.7 Sémantique d'instance

Les lexèmes construits ont, au niveau du lexique, une sémantique abstraite déterminée en grande partie par les familles et les séries auxquelles ils appartiennent. La sémantique lexicale d'un lexème est généralement sous-déterminée afin de pouvoir s'adapter à un nombre suffisant de contextes. On peut en effet supposer que la mémorisation d'un lexème dont le sens serait trop contraint et dont la fréquence ne serait pas élevée n'est pas rentable.

En discours, lorsqu'un locuteur utilise un mot, il lui assigne une **sémantique d'instance**. Cette sémantique d'instance doit idéalement être compatible avec la sémantique lexicale du lexème. Elle n'est cependant pas toujours totalement spécifiée (Corbin et Temple 1994). Les sens des instances ne sont ni prévus ni listés lors de la première utilisation du mot. Ils sont créés et assignés par le locuteur chaque fois qu'il se sert d'un mot. De son côté, l'interlocuteur doit découvrir la sémantique de chaque instance à partir :

1. de la sémantique lexicale du lexème que l'on peut voir comme un invariant sémantique défini par les membres de la série morphologique du lexème et par ceux de sa famille lexicale. Cet invariant est un sens très abstrait. Par exemple, l'invariant correspondant à la série des noms en *-ette* est que le dérivé désigne une entité qui est « petite », d'une façon ou d'une autre, dans sa catégorie (Plénat et al. 2004b, Plénat 2005). Celui des dérivés préfixés en *anti-* de la sous-série qui contient *antiparlementaire*, *anti-émeute* et *antilimace* est de désigner la propriété de s'opposer aux nuisances causées aux intérêts d'un agent non explicité (voir section 4.3.6) [9].

9. Dans la suite du mémoire, j'utiliserai le dérivé *antilimace* comme identifiant de cette sous-série.

2. des valeurs sémantiques des instances du lexème déjà rencontrées et des instances des membres de sa famille et de sa série. Ces valeurs définissent un **sens par défaut**, celui que le mot a habituellement, le plus fréquemment. Par exemple, *antilimace* désigne par défaut un produit qui empêche les limaces de manger les plantations et *bleuet* une centaurée bleue. Le sens par défaut doit être compatible avec l'invariant sémantique précédent.
3. du contexte dans lequel elle est utilisée.

Dans tous les cas, le sens de l'instance est recalculé à l'intérieur des limites définies par l'invariant sémantique de la série morphologique et de la famille lexicale d'une part et du contexte de l'autre. Ce calcul peut se limiter à une simple copie du sens mémorisé dans le lexique (i.e. copie du sens par défaut), mais il peut aussi aboutir à un sens différent induit par un contexte particulier comme *anti-Sarkozy* dans l'exemple (4).

2.8 Système de contraintes

Les correspondances entre sens visés, positions lexicales, formes réalisées et catégories sont réglées par des **contraintes** concurrentes, redondantes et non obligatoires. Certaines de ces contraintes sont également contradictoires, c'est-à-dire qu'elles ne peuvent pas être satisfaites en même temps. Les correspondances entre sens visé et position lexicale, entre forme réalisée et position lexicale et entre catégorie et position lexicale sont conçues comme des compromis optimaux entre les exigences de ces différentes contraintes. Je rappelle que les positions lexicales représentent des lexèmes et que ces derniers sont des classes de mots.

La satisfaction de ces contraintes sur la forme, sur le sens et sur la catégorie présente les caractéristiques suivantes :

- L'optimum est recherché séparément pour chaque occurrence de chaque mot. L'importance relative des contraintes est en effet susceptible de varier en fonction du contexte dans lequel un mot est utilisé et des objectifs du locuteur.
- Les contraintes sont toutes gradables [10]. Leur satisfaction n'est donc pas binaire, en oui ou non.
- Certaines contraintes sur la forme et sur la catégorie sont relativement faibles et ne peuvent obtenir seules la sélection des réalisations qui leur conviennent. Elles doivent s'allier à d'autres contraintes pour leur imposer des ajustements satisfaisants.

Dans le reste du chapitre, je m'intéresse essentiellement aux contraintes qui portent sur la production des mots, c'est-à-dire (*i*) sur la sélection d'une position lexicale pour un sens visé donné (section 2.9) et (*ii*) sur la sélection d'une forme réalisée pour une position lexicale donnée (section 2.11). Les contraintes qui gèrent l'interprétation des mots seront traitées dans une étude ultérieure. Je signale qu'une grande part des contraintes présentées dans ce qui suit ne sont pas nouvelles. Elles réinterprètent dans le modèle à quatre niveaux des contraintes qui ont été déjà proposées ou mises en évidence comme :

10. Pour une présentation de synthèse de la gradabilité des phénomènes morphologiques, voir Hay et Baayen (2005).

- les contraintes dissimilatives (section 2.10.1), qui pénalisent les formes qui contiennent des répétitions de sons identiques ou proches (Plénat 2000, Lignon 2000, Roché 2003a, 2007, Lignon et Plénat 2009, Roché 2009) ;
- les contraintes de taille (section 2.10.2) qui pénalisent les radicaux trop courts ou trop longs (Plénat et Roché 2003, Roché 2003a, Plénat 2009, Roché 2008) ;
- les contraintes de transparence (section 2.11.2) qui favorisent les formes qui permettent l'identification la plus exacte possible du sens construit des lexèmes (Corbin 2001a, b) ;
- les contraintes sur le marquage catégoriel (section 2.11.3) qui favorisent les formes qui permettent la meilleure identification possible de leur catégorie grammaticale (Corbin 2001b).

Ces contraintes sont toutes illustrées par leur contribution à l'analyse des dérivés en *anti-* (voir aussi section 2.12).

2.9 Contraintes sur la position lexicale

Deux contraintes règlent la sélection de la position lexicale associée à un sens visé. Cette sélection dépend de la qualité de la position et de celle de la correspondance entre le sens visé et les sens mémorisés dans la position. La qualité d'une position lexicale est déterminée par :

(*i*) la fréquence cumulée du lexème et celle des membres de sa famille dérivationnelle lexicale (Bybee 1985) ;

(*ii*) la densité de la famille de la position ;

(*iii*) la taille du réseau de relations dans lequel la position est impliquée.

Ces deux dernières caractéristiques sont liées à la richesse et la force de la définition de la position dans le lexique. Les contraintes permettant de sélectionner la position lexicale optimale (C2.1 ; C2.2) sont formulées en utilisant les notions générales suivantes (D2.3 ; D2.4 ; D2.5 ; D2.6) :

Définition 2.3 Relation morphologique
Une relation morphologique est une relation lexicale qui peut être exprimée morphologiquement. Cette relation fait partie des paradigmes morphologiques qui structurent le lexique.

Définition 2.4 Force d'une relation morphologique
Une relation morphologique est d'autant plus forte :

a. que les couples de positions qu'elle connecte sont proches morphologiquement.

b. qu'elle connecte un grand nombre de couples de positions (i.e. qu'elle induit une analogie morphologique qui contient un grand nombre de couples).

c. que les couples de positions qu'elle connecte participent chacun à un grand nombre d'analogies morphologiques.

Définition 2.5 Force de la définition lexicale
Une position lexicale est d'autant plus fortement définie :

a. que les positions avec lesquelles elle est en relation représentent des mots fréquents et que ces positions participent elles-mêmes à un grand nombre de relations et d'analogies morphologiques.

b. que la zone du lexique dans laquelle elle se trouve est dense (i.e. que la position a un grand nombre de voisins proches).

c. que les relations dans lesquelles elle se trouve sont nombreuses et qu'elles participent à un grand nombre d'analogies morphologiques.

Lorsque le sens visé doit être exprimé morphologiquement, la zone dense pertinente pour D2.5b est la famille dérivationnelle lexicale et les relations à considérer pour D2.5c sont les relations morphologiques. Un corollaire de D2.5c est qu'une position lexicale optimale doit appartenir à une série dérivationnelle morphologique aussi grande que possible.

Définition 2.6 Saillance d'une position lexicale
Une position lexicale est d'autant plus saillante :

a. qu'elle représente un mot fréquent.

b. qu'elle est fortement définie.

Contrainte 2.1 Contrainte de préservation de la proximité sémantique
La position lexicale p *qui convient le mieux à un sens visé* s *est celle qui est morphologiquement la plus proche possible des positions* $q_1, \ldots, q_n$ *associées aux mots* $m_1, \ldots, m_n$ *qui sont sémantiquement les plus proches de* s.

En d'autres termes, s'il existe des mots mémorisés dont le sens est proche du sens visé, alors la position lexicale à sélectionner doit être proche de celles de ces mots.

Contrainte 2.2 Contrainte sur la saillance de la position lexicale
La position lexicale qui convient le mieux à un sens visé est la plus saillante possible.

Je signale que les définitions et les contraintes proposées dans cette section ne font intervenir que cinq notions : (*i*) la proximité morphologique ; (*ii*) la proximité sémantique ; (*iii*) la fréquence des mots représentés par une position ; (*iv*) le nombre de relations auxquelles participent une position ; (*v*) le nombre d'analogies auxquelles participent une relation.

2.10 Contraintes morphophonologiques

Je ne considère ici que les contraintes morphophonologiques déjà mises en évidence, à savoir les contraintes dissimilatives et les contraintes de taille. Les dérivés en *anti-* illustrent l'optionalité de ces contraintes. Ces contraintes ne les affectent en effet pas du tout.

2.10.1 Dissimilation

Les contraintes dissimilatives ont été décrites et étudiées en détail par Plénat (2000), Roché (1997, 2003a), Lignon (2000), Lignon et Plénat (2009). Elles pénalisent la création de formes qui contiennent des sons identiques ou proches situés à proximité. Par exemple, elles expliquent pourquoi les locuteurs utilisent la forme *camionneur* et non *camionnier* et pourquoi ils préfèrent *moliéresque* à *moliérien*. Ces contraintes ne sont naturellement pas obligatoires et l'on trouve sur le Web plusieurs attestations de l'adjectif *moliérien*.

Les dérivés en *anti-* témoignent également de l'optionalité de ces contraintes. Elles n'ont en effet aucun impact sur la formation de ces derniers comme le montre la possibilité d'itération du préfixe dans *anti anti bayesien* (16), *anti-anti-CPE* (17) ou *anti-anti-avortement* (18). L'insensibilité à ces contraintes s'observe aussi sur les dérivés dont le radical commence par /ti/ comme *antititan*, *antititanesque* ou *antititiste* (19), par une syllabe phonologiquement proche comme *anti-Tintin* (20) et même par la voyelle /i/ comme *anti-idiot* (21). Cette particularité est signalée par Dell (1995). Dans cet article, Dell s'appuie sur le maintien de la voyelle /i/ à l'initiale du radical dans des configurations comme (21), prononcé /ɑ̃tiidjo/ et non /ɑ̃tidjo/, pour conclure que *anti-*, comme d'autres préfixes, est un mot phonologique à part entière.

(16) *J'ai installé spamassassin depuis qq temps et ça marche pas mal sauf pour les spams* ***anti anti bayesien****.* `linuxfr.org/~lordcow/,2.html`

(17) *Les «* ***anti-anti-CPE*** *» saisissent la justice, comme des salariés le feraient dans une usine bloquée par un piquet de grève.* `www.marianne2.fr/Qui-sont-les-anti-anti-CPE-_a65697.html`

(18) *Une contre-manifestation d'****anti-anti-avortement*** *avait lieu en même temps. Ça a failli dégénérer...* `actualite.alvinet.com/article-1161362-anti-ivg-manifestation-devant-urbain-v-le-flop-des-anti-avortement.html`

(19) *Il connaissait bien Rajk et pensait organiser un gouvernement antititiste en exil.* `www.persee.fr/.../xxs_0294-1759_1990_num_25_1_2225`

(20) *En cela c'est un* ***anti-Tintin****.* `lechoucas.dupuis.com/lechoucas/serie/auteur2.html`

(21) *Imaginons la tête du monsieur qui a installé cela, si demain il y avait un boitier anti-vieux ! ! ! ou pire un boitier* ***anti-idiot*** *(pour rester correct)* `www.rue89.com/2008/04/30/le-boitier-anti-jeunes-est-bien-illegal`

L'insensibilité des dérivés en *anti-* aux contraintes dissimilatives se manifeste également par le fait qu'il n'y a pas de semi-vocalisation de /i/ en /j/ ni d'insertion de yod épenthétique devant une voyelle autre que /i/ (Lombard 1964, Morin 1971, Durand 1982)[11] : *anti-atomique* est prononcé /ɑ̃tiatɔmik/ et non /ɑ̃tjatɔmik/ ni /ɑ̃tijatɔmik/ ; *anti-érotique* est prononcé /ɑ̃tiɛʁɔtik/ et non /ɑ̃tjɛʁɔtik/ ni /ɑ̃tijɛʁɔtik/.

11. Lombard (1964) et Morin (1971) sont cités par Dell (1995).

2.10.2 Taille

La formation des mots est aussi soumise à des contraintes de taille qui pénalisent la création des formes trop longues ou trop courtes. Ces contraintes sont responsables du raccourcissement ou de l'allongement des radicaux afin que la taille de ces derniers soit proche de l'optimum qui est de deux syllabes (Plénat et Roché 2003, Plénat 2009). Par exemple, le nom propre *Récamier* est accourci dans *récamiesque* pour éviter de produire la forme *récamiéresque* dont le radical comporte trois syllabes et deux /e/ consécutifs. La forme *récamiesque* satisfait ainsi à la fois les contraintes dissimilatives et les contraintes de taille. À l'inverse, la forme *goutte* est allongée dans *gouttelette* pour éviter la forme *gouttette* qui présente deux /t/ consécutifs ; là encore, c'est la coalition des contraintes de taille et des contraintes dissimilatives qui induit la forme interfixée.

Les contraintes de taille sont optionnelles et n'ont elles aussi pas d'influence sur la préfixation en *anti-*. Par exemple, le radical de *anticonstitutionnel*, sur lequel est formé le fameux *anticonstitutionnellement*, comporte cinq syllabes. Plus généralement, la possibilité de préfixer *anti-* à des groupes nominaux sans déterminant comme *anti-fil à pendre le linge* (22) légitime la présence parmi ces dérivés de mots longs voire très longs.

(22) *J'ai bien compris que tu étais* ***anti-fil à pendre le linge*** *alors hop, enlevé! Quant à savoir si la levée a été faite, tu n'as pas dû bien regarder.* `www.vacanceo.com/albums_photos/voir-photo_396772.php`

2.11 Contraintes lexicales sur la forme

Les contraintes sur la position lexicale portent sur la correspondance entre le sens visé et la position lexicale. Les contraintes morphophonologiques portent uniquement sur la forme. Il existe un troisième groupe de contraintes qui contrôlent les correspondances entre la position lexicale et la forme et entre la catégorie et la forme, via la position lexicale. Ces contraintes favorisent les formes qui sont les mieux à même d'exprimer morphologiquement le sens des dérivés et leur catégorie. Certaines sont bien connues comme les contraintes de transparence qui, dans un modèle morphématique, stipulent que la forme idéale d'un dérivé est la concaténation des morphèmes qui composent sa structure. À l'inverse des contraintes morphophonologiques, ces contraintes interviennent de manière prépondérante dans la détermination de la forme des dérivés en *anti-*.

2.11.1 Recyclage

La première des contraintes à laquelle sont soumises les formes des mots est celle de la primauté du lexique mémorisé. Cette contrainte favorise la réutilisation des formes ayant déjà servi à exprimer le sens visé par le locuteur. Elle correspond au principe d'économie proposé par Roché (2008). La rareté relative des créations lexicales dans les discours témoigne de l'importance de cette contrainte. Dans les corpus, le nombre de néologismes est infinitésimal par rapport à celui des occurrences. Cette contrainte est aussi responsable de la préservation des formes allomorphiques et supplétives. Elle peut être formulée de la manière suivante (C2.3) :

Contrainte 2.3 Contrainte de recyclage
La forme optimale d'un mot est aussi proche que possible d'une forme mémorisée dans une position lexicale aussi saillante que possible.

Idéalement, cette forme est donc celle d'un mot existant très fréquent. Cette contrainte permet par ailleurs de légitimer les conversions, i.e. les formes obtenues par dérivation non affixale.

La contrainte de recyclage intervient de façon nette dans la sélection des radicaux adjectivaux des formes des dérivés. Par exemple, la sélection de la forme *antipiratable* (23) pour un dérivé dont la base peut être exprimée par le lexème *piratage* [12] peut être en partie attribuée à cette contrainte qui favorise le radical mémorisé *piratable* à un radical nouveau ou à un radical plus rare comme *piratesque*, *piratier*, *piratien*, etc. La sélection de ce radical est également conditionnée par la série des dérivés de la forme *anti-Xable* présentée en section 4.3.6. La table 2.1 fait apparaître la corrélation qui existe entre la fréquence des lexèmes et la disponibilité de leurs formes comme radicaux de dérivés en *anti-* [13]. La contrainte de recyclage rend compte de cette corrélation.

(23) *aucune console n'est* ***anti-piratable***
les mini dvd pour xbox vierge existe a la vente
`forum.hardware.fr/hfr/JeuxVideo/piratage-xbox-commence-sujet_34065_1.htm`

lexème	pages	dérivés en *anti-*
piratage	2 200 000	220 000
piratable	20 000	20
piratesque	940	0
piratier	580	0
piratien	63	0

Table 2.1 – Fréquences de quelques lexèmes de la famille dérivationnelle de *piratage* et des dérivés en *anti-* dont le radical est une de leurs formes. Les chiffres présentés sont la somme des nombres de pages en français indexées par Google qui contiennent ces formes. Une corrélation nette peut être établie entre la fréquence du lexème et la disponibilité de ses formes comme radicaux de dérivés en *anti-*. Décompte réalisé le 27 août 2009.

Le recyclage concerne aussi les formes entières des dérivés et pas seulement leurs radicaux. Par exemple, *antimauritanien* peut à la fois signifier 'opposé à la Mauritanie' (24) ou 'opposé aux mauritaniens' (25). Deux sens construits différents sont ainsi exprimés par la même forme. Notons que le recyclage de la forme *antimauritanien* est renforcé par celui de *mauritanien* qui est à la fois l'adjectif relationnel correspondant à Mauritanie et son

12. *Antipiratable* peut également être analysé comme un dérivé de la sous-série *antihéros* dont le sens serait 'qui a des propriétés contraires à celles d'un système piratable'. Sa base serait alors l'adjectif *piratable*. Je remercie Marc Plénat de m'avoir signalé cette possibilité.

13. Je renvoie à Hathout et al. (2009a, 2008) pour les précautions d'emploi de ce type de statistiques.

gentilé (Roché 2008). Le recyclage contribue ainsi à l'organisation analogique induite par le principe de copie (Dell 1970). Par ailleurs, Roché (2008) note que les adjectifs sont plus souvent concernés par le recyclage que les autres catégories parce qu'ils sont dénués de capacité référentielle directe et qu'ils ont davantage de plasticité sémantique. Je rappelle que les contraintes qui contrôlent les correspondances entre les quatre niveaux de représentation sont redondantes et qu'elles s'appliquent systématiquement même lorsque l'analyse d'un mot particulier n'impose pas d'y faire appel.

(24) *Selon le journal, les deux Français, qui ont quitté samedi Nouakchott pour la France, ont été accusés "d'activités **anti-mauritaniennes**", notamment "d'espionnage et de menées subversives".* `membres.tripod.fr/JournalRefus/actualite10.html`

(25) *Les émeutes anti-commerçants au Sénégal, par exemple, sont devenues des émeutes **antimauritaniennes** dans la presse d'abord, qui a fait l'amalgame entre commerçants et mauritaniens, mais de ce fait, dans la rue dès le lendemain, puis antisénégalaises en Mauritanie, et d'une révolte contre la marchandise et les marchands on est passé à un conflit interethnique.* `www.teleologie.org/OT/deboard/2889.html`

2.11.2 Transparence sémantique

Les contraintes de transparence correspondent à la contrainte de fidélité bien connue en morphophonologie. Dans un cadre lexématique classique, cette contrainte est maximalement satisfaite par les formes des dérivés affixaux obtenus par concaténation directe de l'affixe dérivationnel à la forme de la base comme dans *acide* > *acidité*[14] ou *roman* > *antiroman*. À l'inverse, une forme comme *antiparlementaire* n'est pas optimale car elle décrit un sens construit que l'on peut paraphraser par 'opposé au parlementaire', différent du sens usuel de ce mot, 'opposé au parlement'. Dans ces modèles, la fidélité est conçue comme la configuration par défaut dans laquelle les formes sont les reflets exacts des sens construits compositionnels qu'elles servent à exprimer. Cette contrainte est notamment proposée dans Corbin (2001a) et (2001b) sous forme respectivement, d'une « contrainte de conformité structuro-sémantique » et d'une « contrainte de conformité sémantique »[15].

Dans le modèle paradigmatique que je propose, la transparence est une contrainte gradable qui favorise les formes qui permettent une reconnaissance optimale de la position lexicale qui correspond à un sens visé par le locuteur. Cette contrainte ne concerne que les mots dérivés, c'est-à-dire ceux dont la position lexicale est associée à un sens construit. Dans ce qui suit, cette position est appelée **position sélectionnée**. Je rappelle qu'un sens construit est un sens défini relationnellement, i.e. par les relations qui existent entre sa position et les autres positions du lexique.

La forme idéale d'un dérivé, celle qui est maximalement transparente, doit permettre d'identifier toutes les relations que ce mot entretient avec les autres mots du lexique. Il

14. J'écrirai A > B pour indiquer que B est un dérivé dont la base peut être exprimée par A.

15. Contrainte de conformité structuro-sémantique : La forme optimale d'une unité construite est la plus proche possible de la forme attendue à partir de la structure et du sens construit (Corbin 2001a).

Contrainte de conformité sémantique : La forme optimale d'une unité construite par affixation est la plus proche de la forme obtenue par la concaténation de toutes et rien que les formes des constituants concourant à la construction de son sens (Corbin 2001b).

n'est évidemment pas possible de décrire en même temps, au moyen d'une seule forme, l'immense collection de ces relations et des positions auxquels la position sélectionnée est connectée. La solution de rechange la plus efficace consiste à identifier cette position par l'une de ces relations, celle qui est la plus forte au sens de D2.4. Une fois cette relation choisie, il suffit de l'exprimer de la manière la plus transparente possible. L'identification de la position sélectionnée se fait ainsi au moyen d'un couple (relation, origine) où la relation est celle qui a été choisie pour identifier la position sélectionnée et où l'origine est la position à laquelle la position sélectionnée est connectée par la relation (voir figure 2.3). L'origine est la position de la base du dérivé. La notion de base devient ainsi secondaire. Elle découle ainsi simplement des contraintes de transparence.

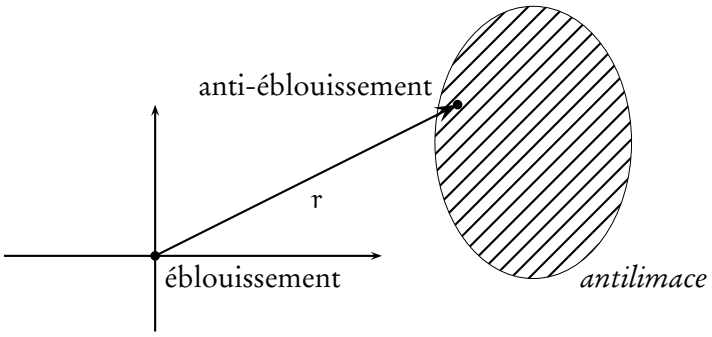


FIGURE 2.3 – Identification de la position de *anti-éblouissement* à partir de celle de sa base *éblouissement* et de la relation morphologique r qui les connecte. La position de *anti-éblouissement* est caractérisée par sa proximité avec la position de *éblouissement* et avec celles des dérivés morphologiques de la sous-série *antilimace*.

La base est donc une position parmi d'autres à partir de laquelle l'identification de la position sélectionnée peut être réalisée. Pour une position sélectionnée donnée, il existe autant de couples qu'il y a de relations qui connectent cette position à d'autres positions lexicales. L'expression du sens construit consiste ainsi à sélectionner le meilleur couple et à l'exprimer de la façon la plus transparente possible. On a donc deux groupes de contraintes : l'un pour choisir le couple (relation morphologique, base) le mieux à même d'identifier le sens construit ; l'autre pour favoriser l'expression la plus transparente possible de ce couple. Les contraintes du premier groupe peuvent être formulées comme suit (C2.4 ; C2.5) :

Contrainte 2.4 CONTRAINTE SUR LA SÉLECTION DE LA RELATION MORPHOLOGIQUE
La relation morphologique qui permet l'identification optimale de la position lexicale associée à un sens construit est la relation la plus forte possible qui connecte cette position à une autre position du lexique.

Cette relation morphologique détermine la série dérivationnelle à laquelle appartient le mot. Je rappelle que la force d'une relation morphologique est définie en D2.4.

Contrainte 2.5 Contrainte sur la sélection de la base
La base qui permet l'identification optimale de la position lexicale associée à un sens construit est :

a. la position la plus proche possible de la position sélectionnée ;

b. la position la plus saillante possible ;

c. la position connectée à la position sélectionnée qui participe avec cette dernière au plus grand nombre possible d'analogies morphologiques.

Comme pour D2.5b, un corollaire de C2.4 et C2.5c est que la série dérivationnelle qui permet la meilleure identification du sens construit est la plus grande possible (i.e. celle qui contient le plus grand nombre possible de lexèmes). Notons que le fait qu'une position sélectionnée p ait une base b n'interdit pas à cette dernière d'avoir p comme base. C'est le cas des mots qui sont en relation de motivation réciproque comme *carrossier* et *carrosserie* ou *pessimiste* et *pessimisme* (Roché 2010a).

Le deuxième groupe de contraintes porte sur l'expression transparente de cette base b et de la série dérivationnelle qui correspond à la relation morphologique r qui connecte b à la position sélectionnée. Cette position peut être vide ou contenir des mots mémorisés. Dans le second cas, la position appartient à une série qui exprime r et qui est sélectionnée. Dans le premier, il faut identifier la série la mieux à même d'exprimer r en utilisant D2.7.

Considérons par exemple un dérivé adjectival dont le sens construit est 'qui protège de l'éblouissement'. Cette position est définie par le couple (*éblouissement*, r) où r(x, y) est une relation glosable par 'x protège de y'. Cet adjectif occupe une position caractérisée par sa proximité avec la position de *éblouissement* et avec celles des dérivés morphologiques de la sous-série *antilimace*, de la série des dérivés en *pare-* (*pare-soleil*) et en *para-* (*parapluie*) et des composés en *protège-* (*protège-cahier*). Supposons que la fréquence et la productivité de la préfixation en *anti-* conduise à sélectionner cette sous-série, i.e. à utiliser ou créer une position dans cette sous-série. L'identification de cette position peut alors être illustrée comme en figure 2.3.

La contrainte de transparence de la base (C2.6) favorise les formes qui permettent d'identifier la base, i.e. la position correspondant au mot *éblouissement*. La meilleure façon d'y parvenir est simplement d'y faire apparaître la séquence *éblouissement*. De même, la transparence de la série dérivationnelle (C2.7) favorise les formes qui permettent d'identifier la série du dérivé, en l'occurrence celle des mots préfixés en *anti-*, et plus précisément ceux de la sous-série *antilimace*. La meilleure façon pour une forme d'atteindre cet objectif est de posséder les marques caractéristiques de cette série, c'est-à-dire l'initiale *anti-*. Plus généralement, la transparence peut être imposée au moyen des deux contraintes suivantes (C2.6 ; C2.7) complétées par la définition D2.7 :

Définition 2.7 Série dérivationnelle d'un dérivé
La série dérivationnelle d'une position est la série dont les positions sont les plus proches possible de cette position.

Contrainte 2.6 Contrainte de transparence de la base

La forme optimale permettant l'identification de la position lexicale d'un dérivé est formellement la plus proche possible de l'une des formes d'un mot qui est morphologiquement aussi proche que possible de la base du dérivé.

Contrainte 2.7 Contrainte de transparence de la série dérivationnelle
La forme optimale d'un dérivé est formellement la plus proche possible du plus grand nombre possible de formes de mots appartenant à la série dérivationnelle du dérivé.

La contrainte C2.6 légitime à la fois les formes *antipère* et *antipaternel* pour exprimer un sens construit comme 'qui s'oppose aux pères' tout en favorisant la première forme par rapport à la seconde. La forme *père* étant plus proche que *paternel* de *père*, *antipère* est plus transparent que *antipaternel*. Le radical *paternel* est néanmoins acceptable selon C2.6 car c'est une forme d'un mot dont la position lexicale est contiguë à celle de la base *père* (voir section 2.4). Le fait que *paternel* appartienne à la famille lexicale de *père* découle simplement de la définition de ce paradigme. La situation est identique pour *antiparlement* et *antiparlementaire* : la première forme est plus transparente que la deuxième. La contrainte C2.7 rend compte d'une forme comme *européen* dont la finale *-éen* est proche mais n'est pas identique à *-ien*, la marque caractéristique des mots de sa série dérivationnelle (Lignon 2000, 2010).

Un corollaire important des contraintes C2.6 et C2.7 peut être formulé sous la forme d'une contrainte C2.8 :

Contrainte 2.8 Contrainte d'adéquation
La forme optimale d'un dérivé ne contient que du matériel phonologique qui contribue directement à l'identification de sa position lexicale.

Cette contrainte rend compte du fait qu'il ne suffit pas de disposer des marques caractéristiques de la série dérivationnelle et de contenir l'une des formes de la base pour être maximalement transparent. Par exemple, toutes les formes qui contiennent la séquence *éblouissement* et qui commencent par *anti-*, comme par exemple *antiséblouissementière*, ne sont pas aussi satisfaisantes que *anti-éblouissement*. La contrainte C2.8 pénalise les formes qui contiennent du matériel sonore qui ne sert pas directement à identifier la position du sens construit comme le *-s-* épenthétique et la finale *-ière* dans *antiséblouissementière*. C'est un corollaire des contraintes C2.6 et C2.7 car les formes qui contiennent du matériel parasite sont plus éloignées de la base et de la série dérivationnelle que celles qui ne contiennent que le minimum indispensable, les propriétés relatives à ces segments additionnels n'étant partagées ni avec la base ni avec les membres de la série dérivationnelle.

Les formulations de C2.6 et C2.7 ne font pas intervenir directement les notions de radical et de marque caractéristique. Ces deux notions peuvent être définies comme des ensembles de propriétés formelles : le radical correspond à l'ensemble maximal de propriétés formelles partagées avec la base ou un mot de sa famille lexicale ; les marques caractéristiques correspondent à l'ensemble des propriétés les plus spécifiques partagées par le maximum de formes de mots de la série dérivationnelle du dérivé. Le cas optimal est

celui de *anti-éblouissement* dont le radical est une des formes de la base [16]. Le dérivé *anti-éblouissement* partage ainsi le maximum possible de propriétés formelles avec *éblouissement*. *Anti-éblouissement* dispose de l'initiale *anti-* comme tous les dérivés de sa série. Il partage donc avec le maximum d'éléments de cette série la propriété la plus spécifique possible, i.e. la présence du préfixe.

2.11.3 Transparence catégorielle

Le niveau de représentation catégoriel décrit l'organisation du lexique en catégories. Cette organisation exerce sur la formation des mots une pression similaire à celle des séries. Il n'existe en effet qu'un nombre très réduit de combinaisons de traits morphosyntaxiques qui induisent un nombre tout aussi réduit de classes. Chaque lexème fait partie d'une catégorie (i.e. partie du discours) qui détermine les différentes manières dont on peut l'utiliser. Les propriétés catégorielles constituent ainsi l'un des éléments fondamentaux de la caractérisation des lexèmes, qu'ils soient construits ou non. Les catégories sont essentiellement des ensembles de même nature que les séries, même si leur fonction est nettement différente. La contrainte qu'elles exercent sur la sélection de la forme des mots est similaire à celle des séries (C2.9). Je rappelle (*i*) qu'en production, le rôle de la morphologie est d'exprimer un sens au moyen d'un mot, (*ii*) que ce dernier doit entrer dans un énoncé et (*iii*) que c'est ce contexte qui détermine sa catégorie. Pour cette tâche, la catégorie du dérivé peut donc être considérée comme une donnée.

Contrainte 2.9 Contrainte de transparence catégorielle
La forme optimale d'un dérivé est formellement aussi proche que possible du plus grand nombre possible de formes de mots appartenant à la catégorie grammaticale du dérivé.

La contrainte C2.9 favorise l'utilisation de formes qui ressemblent à celles des mots de leur catégorie qui sont déjà mémorisés et notamment aux mots construits. C2.9 est en fait simplement un cas particulier d'une règle plus générale qui favorise la création de mots conformes aux régularités formelles déjà établies au sein du lexique. En l'occurrence, il s'agit ici de régularités formelles qui s'établissent à l'intérieur des catégories grammaticales. Cette contrainte reprend l'idée des intégrateurs paradigmatiques proposée par Corbin (1991). Elle est également proche de la « contrainte de conformité catégorielle » [17] introduite par Corbin (2001b). En français, le marquage morphologique dérivationnel se situe généralement à la fin des mots. Cette situation est liée au fait que les mots dérivés constituent une partie importante des unités du lexique et au fait que les mots suffixés représentent la plus grande partie d'entre eux. Par ailleurs, les marques flexionnelles sont aussi toutes situées à la fin des mots (Dell 1970).

16. *Anti-éblouissement* peut également être analysé comme un dérivé formé sur le verbe *éblouir* dont le sens serait 'qui protège du risque d'être ébloui'. Dans ce cas, cette forme ne satisferait pas la contrainte de transparence de la base mais pourrait être sélectionnée parce qu'elle serait optimale relativement à la contrainte sur la sélection de la relation morphologique (C2.4) .

17. Contrainte de conformité catégorielle : En français, la forme optimale d'un mot construit se mesure à la conformité de sa finale aux finales prototypiques, s'il y en a, de la catégorie lexicale à laquelle il appartient. En particulier, la forme optimale d'un adjectif construit est celle d'un adjectif suffixé (Corbin 2001b).

La ressemblance formelle d'un mot avec les mots de sa catégorie ou au moins avec un nombre suffisant de ses éléments est évaluée sur la base du partage de la même finale. Par exemple, pour qu'un dérivé ressemble à un adjectif il doit avoir une finale caractéristique des mots construits de cette catégorie comme *-el*, *-ique*, *-eux*, *-if*, *-ien*, *-ard*, *-able*, *-é*, etc. Ainsi, lors de la construction d'un mot qui doit être utilisé comme un adjectif, la contrainte C2.9 favorisera les formes qui comportent l'une de ces finales. Dans le cas des suffixations, C2.9 n'a pas d'effet puisque les mots suffixés ont trivialement une finale déjà présente parmi les éléments de leur catégorie. Ainsi, les adjectifs dérivés par suffixation ont une finale qui est bien celle d'un adjectif. Dans le cas des préfixations, la satisfaction de la contrainte de transparence catégorielle n'est pas automatique. Diverses stratégies visant à la satisfaire peuvent être observées comme le recours à un radical adjectival morphologiquement apparenté à la base comme dans *parlement* > *antiparlementaire* (voir section 4.5) ou l'élision d'une partie des constructions qui expriment cette base comme dans *attaque aérienne* > *anti-aérien* (voir section 4.5.8). Les préfixations en *post-*, *pré-*, *sub-*, *sous-*, *anté-*, etc. sont soumises aux mêmes contraintes et présentent des dérivés similaires à *anticancéreux* et *antiparlementaire* (Corbin 1980, 2001b). Les radicaux des dérivés en (26 ; 27 ; 28 ; 29 ; 30) sont des formes adjectivales permettant de satisfaire la contrainte de transparence catégorielle.

(26) *je vais me faire opérer et je voulais savoir ce qu'on faisait comme recherche dans le sang lors de ce bilan* ***pré-opératoire*** *mis à part le groupe sanguin.* `forum.doctissimo.fr/sante/chirurgie/operatoire-bilan-sujet_606_1.htm`

(27) *Le symptôme de douleur* ***post-opératoire*** *est subjectif et limité dans le temps.* `www.institut-upsa-douleur.org/Layout/presentation.aspx?ItemID=f2f49c47-07a0-456b-a047-c51720bc2267`

(28) *L'archéologie* ***subaquatique*** *a pour objet d'inventorier, d'identifier et d'étudier les sites à caractère préhistorique ou historique.* `archeologie.ffessm.fr/`

(29) *La fonte du glacier forme un torrent appelé torrent émissaire, composé des eaux des torrents* ***sous-glaciaires****.* `www.ceillac.com/projet-eau-pages/glacier-blanc.htm`

(30) *Pour montrer combien il est facile de s'abuser en tirant des conséquences de certains rapprochemens entre la civilisation* ***antécolombienne*** *de l'Amérique et celle de l'ancien continent, ...* `fr.wikisource.org/wiki/La_civilisation_Mexicaine_avant_Fernand_Cortez`

La caractérisation des contraintes qui portent sur les correspondances entre les niveaux lexical et catégoriel sera développée davantage dans le cadre d'une future description détaillée de la structure flexionnelle des lexèmes dans le modèle à quatre niveaux.

2.12 Le jeux des contraintes : le cas des dérivés en *anti-*

Conjointement, les contraintes présentées ci-dessus permettent de rendre compte de la grande variété des relations qui existent entre les formes des dérivés et leur sens. Elles permettent notamment de traiter des phénomènes habituellement analysés au moyen du

principe de copie (Corbin 1987) et du principe d'économie (Roché 2008). Par ailleurs, ces contraintes affectent uniformément l'ensemble des dérivés en *anti-* indépendamment de leurs spécificités sémantiques. Pour illustrer l'interaction de ces contraintes, considérons l'exemple d'un dérivé de la sous-série *antilimace* dont la base pourrait être exprimée par *cancer*. Ses formes optimales sont déterminées de la manière suivante :

- La contrainte de transparence de la base (C2.6) favorise les formes qui ressemblent le plus à la base du dérivé et notamment celles dont les radicaux sont des formes de la base (*cancer*) ou à défaut, de ses voisins morphologiques immédiats (*cancéreux*, *cancéré*, *cancérien*, *cancérique*, *cancérer*, *cancérable*) ou éventuellement plus distants (*cancérigène*, *cancérologue*, etc.).
- La contrainte de transparence de la série dérivationnelle (C2.7) favorise les formes qui portent l'initiale *anti-* caractéristique de la série.
- La contrainte de transparence catégorielle (C2.9) favorise les formes qui portent une finale adjectivale ou dont la finale ressemble à celle d'un adjectif (*anticancéreux*, *anticancéré*, *anticancérien*, *anticancérique*, *anticancérable*, etc.). Elle pénalise en revanche *anticancer* qui ne dispose pas d'une telle finale.
- La contrainte de recyclage (C2.3) favorise les formes déjà présentes dans le lexique ou qui ressemblent à un mot existant et celles dont le radical est une forme connue ou qui ressemble à un mot connu. Les radicaux *cancer*, *cancéreux* et *cancérigène* sont avantagés par cette contrainte.

Le bilan des exigences placées par les différentes contraintes fait apparaître deux finalistes :

1. *anticancéreux* est la forme optimale pour les contraintes de transparence de la série dérivationnelle et de transparence catégorielle. Cette forme est également tout à fait satisfaisante pour les contraintes de transparence de la base et de recyclage. C'est leur deuxième forme préférée après *anticancer*.

2. *anticancer* est la forme optimale pour les contraintes de transparence de la série dérivationnelle, de transparence de la base (c'est la forme la plus proche de *cancer*) et de recyclage (*cancer* est un mot beaucoup plus saillant dans le lexique existant que les autres radicaux candidats).

Le fait que les formes *anticancer* et *anticancéreux* coexistent en français ne pose aucun problème à la théorie proposée ici. Cette variation découle simplement du fait que pour certaines occurrences, les locuteurs donnent plus d'importance à la contrainte de transparence catégorielle et que, pour d'autres, ils préfèrent satisfaire pleinement les contraintes de transparence de la base et de recyclage. Cet exemple montre qu'il n'y a pas de hiérarchie universelle entre les contraintes et que la contrainte de transparence de la base et la contrainte de recyclage sont beaucoup plus fortes que la contrainte de transparence catégorielle : la forme *anticancéreux* est sélectionnée parce qu'elle est quasi optimale pour les deux premières contraintes alors que *anticancer* est la plus attestée bien qu'elle soit totalement inacceptable pour la troisième.

2.13 Conclusion

Ce chapitre propose un cadre théorique général qui permet d'articuler les travaux présentés dans la suite du mémoire. Ce modèle prend comme point de départ l'organisation relationnelle du lexique et la notion de proximité morphologique. Il reconstruit ensuite les notions de famille morphologique, de série dérivationnelle, de base à partir de l'hypothèse que la morphologie s'occupe de l'expression du sens au moyen de mots. Il comporte un système de contraintes permettant de rendre compte de la grande diversité des stratégies que les locuteurs utilisent pour exprimer un sens au moyen d'un mot dérivé. Ce modèle est remarquable par plusieurs aspects :

- il est radicalement lexématique ;
- il est paradigmatique ;
- il est non compositionnel ;
- il ne comporte ni règle de construction de mots, ni patrons de construction à la Booij (2002, 2007) ;
- c'est un modèle exemplariste. Il ne contient aucune structure ou ni aucun objet qui représente des abstractions. Les paradigmes sont de simples ensembles de mots.

L'utilité de ce modèle reste à confirmer notamment en le confrontant à différentes affixations, en français et dans d'autres langues. Il doit être étendu à la flexion, à la dérivation non affixale et à la composition en mettant au point des modes de représentation adaptés pouvant servir de support aux analogies. Certains de ses aspects doivent être explicités davantage, notamment l'ensemble des informations associées aux positions lexicales, la représentation des paradigmes et en particulier des lexèmes, la nature précise des objets mis en correspondance. Cette élaboration sera réalisée parallèlement au niveau informatique et théorique afin d'identifier les besoins et de vérifier l'opérationnalité des réponses proposées.

Mais la tâche la plus délicate est la conception de protocoles expérimentaux permettant de confirmer la réalité des notions et des contraintes qui ont été proposées. Il est inutile d'espérer atteindre des confirmations similaires à celles de Plénat (2000), Plénat et Roché (2003), Roché (2003a), Plénat (2009) mais il devrait être possible de réaliser des expériences sur corpus et des expériences psycholinguistiques permettant de les tester.

CHAPITRE 3

Nouveau paradigme d'analyse morphologique automatique

Mes travaux en linguistique descriptive et en linguistique informatique sont deux facettes d'un seul et même programme. L'ensemble repose sur les mêmes hypothèses théoriques, synthétisées dans le chapitre 2. Ces hypothèses permettent de définir un nouveau paradigme d'analyse morphologique automatique qui consiste à passer d'une analyse morphologique des mots à une analyse morphologique du lexique. Cette évolution découle directement de l'approche radicalement lexématique qui est adoptée ici.

3.1 Introduction

3.1.1 Analyse morphologique lexématique

Classiquement, la tâche d'analyse morphologique automatique consiste à décomposer les mots en morphèmes (Déjean 1998a, b, Gaussier 1999, Schone et Jurafsky 2000, Goldsmith 2001, Creutz et Lagus 2002, Bernhard 2006a, b). C'est notamment ainsi qu'elle est définie dans les campagnes d'évaluation *Morpho Challenge* (Creutz et Lagus 2005, Kurimo et al. 2007, Kurimo et Varjokallio 2008). Le cadre théorique sous-jacent est celui de la morphologie morphématique qui considère que les mots sont constitués de morphèmes et que ces derniers se combinent relativement à des règles de flexion, de dérivation et de composition. Les mots seraient donc pourvus d'une structure généralement représentée de façon arborescente comme en figure 3.1.

Les limites de la morphologie morphématique sont bien connues (Anderson 1992, Aronoff 1994). Elles peuvent être illustrées par les quelques exemples suivants :

Exemple 1.– Quel est le statut d'une initiale comme *per-* que l'on retrouve dans les mots de la liste (1) ? Pour chacun de ces mots, *per-* peut être remplacée par une autre initiale comme le montrent les mots entre parenthèses.



FIGURE 3.1 – Structures du nom *dérivabilité* et de l'adjectif *dérivationnel* telles qu'elles sont prévues par la morphologie morphématique.

(1) *percevoir* (*recevoir*), *percuter* (*discuter*), *perdition* (*rédition*), *perfection* (*défection*), *performer* (*déformer*), *perfuser* (*refuser*), *permanent* (*rémanent*)

EXEMPLE 2.– Le nom *dérivation* et l'adjectif *dérivatif* sont formés sur un thème du verbe *dériver*. Pourquoi n'existe-t-il pas de verbe similaire pour le nom *locomotion* et l'adjectif *locomotif* ? En d'autres termes, étant donné (2a), quelles sont les solutions des équations en (2b) ?

(2) a. $dérivation = [dériv_{V}\ \textit{-ion}]_{N}$
$dérivatif = [dériv_{V}\ \textit{-if}]_{A}$

b. $locomotion = [x\ \textit{-ion}]_{N}$
$locomotif = [x\ \textit{-if}]_{A}$

EXEMPLE 3.– L'analyse des conversions comporte un morphème zéro dont les propriétés sémantiques sont difficiles à établir. En (3a), il permet par exemple de former un nom de qualité, en (3b) un nom d'action et en (3c) un nom qui désigne une entité d'après sa qualité.

(3) a. $pouvoir_{V} \rightarrow [pouvoir_{V}\ \emptyset]_{N}$

b. $voler_{V} \rightarrow [vol_{V}\ \emptyset]_{N}$

c. $petit_{A} \rightarrow [petit_{A}\ \emptyset]_{N}$

EXEMPLE 4.– Quel est le statut des interfixes comme en (4) ? Quelles sont leurs propriétés sémantiques ? Ces séquences, formellement identiques à des suffixes, ont un rôle uniquement épenthétique (Plénat et Roché 2003, Roché 2003b, Plénat 2009).

(4) *tarte* → *tart**el**ette* ; *goutte* → *gutt**el**ette*

Les réponses théoriques apportées à ces problèmes s'inscrivent quant à elles dans le cadre de la morphologie lexématique (Aronoff 1976, 1994, Anderson 1992, Bybee 1988, 1995, Neuvel et Singh 2001, Burzio 2002, Blevins 2003, 2006). Dans cette approche, les mots sont des unités atomiques qui n'ont donc pas de structure morphologique. La structure morphologique est conçue comme un niveau d'organisation du lexique qui se compose des relations morphologiques s'établissant entre les mots. Elle correspond au niveau de représentation lexical présenté en section 2.3.

La validité de la conception lexématique de la morphologie a fait l'objet de nombreuses études en psycholinguistique. En effet, les modèles psycholinguistiques se distinguent nettement selon que la morphologie soit considérée comme morphématique ou lexématique. Dans un cas, les objets stockés dans le lexique mental sont conçus comme des morphèmes et les mots sont décomposés lors de leur traitement. Dans l'autre, le lexique mental contient des formes complètes et le traitement des mots ne comporte pas de décomposition. Parmi les tenants de la morphologie lexématique, on peut citer Hay et Baayen (2005) dont l'une des conclusions est que la gradabilité des phénomènes morphologiques est compatible avec cette conception.

Ces changements dans la manière dont la morphologie est conçue n'ont en revanche pas eu jusqu'ici de réelles conséquences sur son traitement automatique si l'on excepte les travaux de Neuvel et Singh (2001), Neuvel et Fulop (2002), Hathout (2001, 2003, 2005) sur la découverte de relations morphologiques entre les mots et l'acquisition de schémas d'affixation. Ce chapitre apporte une autre réponse à cette évolution.

Le problème à résoudre peut être formulé comme suit : comment réaliser une analyse morphologique automatique dans le cadre d'une morphologie lexématique, c'est-à-dire sans recourir aux notions de morphème, d'affixe ni d'exposant morphologique ? La solution que je propose passe par un changement de paradigme. Elle consiste à redéfinir la tâche d'analyse morphologique pour que son objectif ne soit plus de découper les mots en morphèmes mais de découvrir l'organisation morphologique du lexique. Cette organisation se compose de l'ensemble des paradigmes identifiés en section 2.6.1 et des analogies qui les connectent, présentées en section 2.6.2. En d'autres termes, il s'agit de construire le maillage défini par les familles, les séries et les analogies morphologiques. L'analyse des mots particuliers est ensuite obtenue comme un sous-produit du maillage global. Par exemple, le nom *dérivation* n'est plus découpé en deux morceaux *dérivat* et *ion*, le premier identifié au morphème verbal *dériv* et le second au suffixe déverbal *-ion* comme dans l'approche morphématique. Son analyse consiste ici à identifier sa position dans le lexique. Cette position est caractérisée d'une part par son appartenance à la famille des mots en (5a) et à la série des mots en (5b) et d'autre part par les analogies qui contiennent les relations qui connectent *dérivation* aux éléments de ces deux ensembles comme (6a) pour *dérivation* : *dérivable* ou (6b) pour *dérivation* : *variation*.

(5) a. *dériver* ; *dérivable* ; *dérivatif* ; *dérivatoire* ; *dérivationnel* ; *dérivabilité* ; *dérive* ; *dériveur* ; *dérivette* ; *dérivement*

b. *acclimatation* ; *bonification* ; *création* ; *évolution finition* ; *éducation* ; *localisation* ; *séduction* ; *variation*

(6) a. *dérivation* : *dérivable* :: *modification* : *modifiable*
dérivation : *dérivable* :: *adaptation* : *adaptable*
dérivation : *dérivable* :: *observation* : *observable*

b. *dérivation* : *variation* :: *dériver* : *varier*
dérivation : *variation* :: *dérivationnel* : *variationnel*
dérivation : *variation* :: *dérivabilité* : *variabilité*
dérivation : *variation* :: *dérivable* : *variable*

3.1.2 Combiner la proximité morphologique et l'analogie formelle

Dans la suite de ce chapitre, je présente un modèle informatique qui implémente certains aspects du modèle théorique décrit dans le chapitre 2. Ce modèle a été mis au point en utilisant un dictionnaire informatisé, en l'occurrence le *Trésor de la Langue Française informatisé* [1] (TLFi). Il permet de vérifier que le modèle théorique est opératoire. Cependant, mon objectif principal est de créer une ressource morphologique fiable et à large couverture. Pour l'heure, il n'en existe pas pour le français. Ce lexique offrira une description globale d'un fragment central et de grande taille du vocabulaire du français.

Les relations morphologiques étant des relations de partage de forme et de sens, proposer une modélisation qui intègre les deux types d'information me semble indispensable d'autant plus si l'on a l'ambition de rendre compte de phénomènes difficiles comme l'allomorphie, les supplétions ou la conversion. Les descriptions sémantiques utilisées dans l'expérience présentée dans la suite du chapitre est en revanche relativement rudimentaire. Elles servent avant tout à montrer comment il serait possible de les intégrer dans ce modèle. De plus, les ressources électroniques qui décrivent des connaissances sémantiques sont relativement rares. Les plus complètes d'entre elles sont certainement les dictionnaires informatisés et les bases de données lexicales de type WordNet (Miller et al. 1990, Fellbaum 1999). Je suis naturellement conscient des limites des descriptions lexicographiques, mais je pense que les bénéfices que l'on peut en tirer dépassent largement les difficultés qu'elles induisent. Ces descriptions sont par ailleurs plus faciles à utiliser que des connaissances extraites de corpus telles que des voisinages distributionnels (Galy et Bourigault 2005) car elles ne contiennent que les définitions des sous-sens les plus représentatifs de l'usage des mots.

La méthode d'analyse morphologique du lexique que je propose s'appuie d'une part sur une mesure de proximité morphologique qui rapproche les membres des familles morphologiques et des séries dérivationnelles. Cette mesure prend en compte à la fois les similarités formelles et sémantiques entre les mots. Elle s'appuie d'autre part sur la découverte d'analogies formelles entre voisins morphologiques. Le recours à l'analogie est relativement répandu en morphologie informatique (Skousen 1989, 1992, Lepage 1998, 2000, 2003, Van den Bosch et Daelemans 1999, Pirrelli et Federici 1994, Pirrelli et Yvon 1999, Hathout 2003, 2005, Yvon et al. 2004, Stroppa et Yvon 2005, 2006, Stroppa 2005, Yvon 2006, Langlais 2009). La principale nouveauté de ma méthode est de combiner l'analogie et la proximité morphologique. J'utilise dans un premier temps la proximité morphologique pour sélectionner des quadruplets de mots susceptibles d'être morphologiquement apparentés puis, dans un deuxième temps, je les filtre pour ne garder que ceux qui forment une analogie formelle. Les deux techniques sont complémentaires : les voisinages morphologiques peuvent être calculés en grand nombre, mais ils sont trop grossiers pour discriminer entre les mots qui sont effectivement morphologiquement apparentés et ceux qui ne le sont pas ; l'analogie formelle permet un filtrage fin mais elle est coûteuse à calculer.

Plus généralement, le paradigme d'analyse morphologique proposé est original à plusieurs égards. Le premier est que le modèle informatique est purement lexématique. La

1. `atilf.atilf.fr/tlf.htm`

découverte des relations morphologiques qui existent entre les mots ne fait pas intervenir de morphèmes, d'affixes, d'exposants morphologiques ni aucune représentation de ces notions. Le deuxième est que les relations morphologiques entre les mots n'ont pas toutes la même force (voir définitions D2.4). De ce fait, l'appartenance aux familles et aux séries est graduelle aussi. Par exemple, dans la famille de *dériveur*, *dérive* est plus proche de ce mot que ne l'est par exemple *dérivationnellement*. De même, dans la série de *formation*, *compilation* est plus proche de ce mot que ne l'est *vision* : *compilation* partage en effet plus de matériel phonologique avec *formation* et participe avec lui à un plus grand nombre d'analogies, en supposant que les traits sémantiques partagés par *formation* avec ces mots sont les mêmes. Mon modèle est ainsi compatible avec les observations de Hay et Baayen (2005) sur la gradabilité des phénomènes morphologiques. Le troisième est sa contribution à l'implémentation des propositions théoriques de Bybee et de Burzio.

La méthode que je propose est aussi très robuste et peut être facilement adaptée à d'autres langues romanes, à l'anglais et probablement aussi à des langues gabaritiques comme l'arabe ou l'hébreu. Cette méthode dispose par ailleurs d'une efficacité suffisante pour être utilisée pour construire de façon semi-automatique une ressource morphologique à large couverture. Elle intègre de manière uniforme les connaissances sémantiques et formelles. Les traits utilisées dans l'expérience présentée ci-dessous pourraient être complétés ou éventuellement remplacés par d'autres comme des traits phonologiques, des éléments de descriptions extraits de dictionnaires de synonymes ou de WordNet ou des propriétés distributionnelles acquises dans des corpus. Tous ces traits peuvent être cumulés sans difficulté malgré les différences de nature et d'origine. Le modèle tire en effet profit de la redondance des traits. Par ailleurs, il est peu sensible aux variations et aux exceptions.

D'un point de vue technique, ma méthode associe :

(*i*) la représentation du lexique sous la forme d'un graphe et son exploitation par un algorithme de marches aléatoires, dans la lignée de Gaume (2003, 2004), Gaume et al. (2002, 2004a, b, 2006), Muller et al. (2006) ;

(*ii*) l'analogie formelle sur les mots (Lepage 1998, 2000, 2003, Yvon et al. 2004, Stroppa et Yvon 2005, 2006, Stroppa 2005, Yvon 2006, Langlais et Patry 2007, 2008, Langlais 2009, Langlais et al. 2009). Cette technique ne fait pas appel à la notion de morphème. Les correspondances entre les mots sont calculées directement au niveau des chaînes de caractères (graphies ou transcriptions phonologiques).

3.1.3 Réseau morphologique

La ressource morphologique qui sera construite à terme est un réseau qui décrit la structure morphologique du lexique présentée dans le chapitre 2. Ce réseau est composé de mots interconnectés par les relations qui s'établissent au sein des paradigmes présentés en section 2.6.1, notamment les familles et les séries dérivationnelles morphologiques et lexicales. Cette structure est complétée par les analogies qui unissent les couples de lexèmes morphologiquement apparentés. Le lexique aura une grande flexibilité lui permettant de représenter de façon adéquate différents phénomènes morphologiques. Le principal est l'allomorphie qui correspond à une position pour laquelle il y un décalage entre les analo-

gies formelles et l'organisation en familles en séries morphologiques. Par exemple, la forme déverbale *dénivellation* sera identifiée comme une forme allomorphique parce que (*i*) elle appartient à la série des mots en *-ion* et (*ii*) à la famille de *déniveler* de par sa proximité avec les éléments qui composent chacun de ces ensembles, et parce que (*iii*) c'est le nom le plus proche de ce verbe. Les noms qui finissent en *-ion* et notamment en *-ation* se trouvent dans des analogies avec leur verbe le plus proche, comme (7). L'absence d'une analogie similaire pour *dénivellation* apparaît alors comme un trou dans la grille analogique. Un second indice est la proximité formelle qui existe entre la forme de *dénivellation* (à la fois graphique et phonologique /denivɛlasjɔ̃/) et celle qui permettrait de combler ce trou (*dénivelation* ; /denivlasjɔ̃/ [2]).

(7) *dérivation* : *dériver* :: *compilation* : *compiler*
formation : *former* :: *variation* : *varier*

Ce réseau permettra de représenter par exemple de la même façon les similarités et les différences qui existent entre *curieux* et *furieux* (Jackendoff 1975). Le fait que *furieux* puisse être analysé comme un adjectif construit sur le nom *furie* découle de l'appartenance des deux mots à la même famille et de leur proximité morphologique ; en revanche, *curieux* n'étant plus actuellement sémantiquement apparenté au nom *cure* — dans le sens qu'il a dans une expression comme *n'avoir cure de* —, il ne fait pas partie de sa famille morphologique. Le réseau décrira également le fait que *furieux* : *furie* participe à des analogies avec d'autres couples de mots comme (8) alors que *curieux* ne se trouve dans aucune analogie similaire. À l'inverse, l'appartenance de *furieux* et *curieux* à la même série lexicale découle simplement du fait qu'ils possèdent chacun les propriétés sémantiques et formelles des adjectifs en *-eux*. Cet exemple illustre la souplesse du modèle et le pouvoir descriptif apporté par les paradigmes et les analogies. Par comparaison, la similarité entre *curieux* et *furieux* ne peut pas être décrite de façon simple dans un modèle morphématique ou dans tout autre modèle ne disposant pas de séries lexicales ou d'un concept équivalent.

(8) *mélodieux* : *mélodie* ; *harmonieux* : *harmonie* ; *facétieux* : *facétie*

La représentation des mots qui contiennent des interfixes comme en (4) ne présente pas non plus de difficulté. Ces mots sont des membres à part entière de leurs familles et de leurs séries. La relation dérivationnelle qu'ils entretiennent avec leurs bases est tout à fait normale. Elle est représentée par leur appartenance à la même famille morphologique et par leur proximité morphologique. D'autre part, la présence de l'interfixe renforce leur intégration formelle dans leurs séries.

L'organisation d'un tel lexique en réseau permet aussi de mieux répondre aux besoins des outils de traitement automatique des langues et de recherche d'information. Les ressources morphologiques sont notamment utilisées dans ces domaines pour la désambiguïsation des rattachements prépositionnels et pour l'expansion de requête. Par exemple, un analyseur syntaxique comme Syntex (Bourigault et Fabre 2000, Bourigault 2007) utilise Verbaction, un lexique composé des couples de noms et de verbes proches qui appartiennent aux mêmes familles morphologiques. Ces informations seront directement dis-

2. Prononciation en français standard. Dans les variantes du français où l'on dit /denivəle/, on aurait naturellement /denivəlasjɔ̃/.

ponibles dans le réseau proposé et pourront ainsi être sélectionnées avec davantage de précision. Dans le domaine de la recherche d'information, les performances de recherche peuvent être améliorées en ajoutant aux requêtes des mots morphologiquement apparentés à ceux qu'elles contiennent (Jing et Tzoukermann 1999, Jacquemin et Tzoukermann 1999, Claveau et L'Homme 2005, Moreau et al. 2007). Ces mots appartiennent tous aux familles morphologiques des mots présents dans les requêtes initiales. De plus, l'estimation de la proximité morphologique associée à chaque relation permet un réglage fin des expansions. Par ailleurs, la nature relationnelle de ce réseau n'ajoute aucune difficulté particulière comme en témoigne le nombre considérable d'applications qui utilisent la base WordNet.

Ce lexique pourra également être utilisé pour la constitution de matériel expérimental en psycholinguistique. Par exemple, le fait qu'un mot soit construit morphologiquement ou non peut être directement obtenu en comparant ses séries morphologiques et lexicales. Parmi les autres caractéristiques prises en compte lors de la sélection du matériel, on peut citer la ressemblance formelle et l'appartenance à la même famille morphologique. Ces informations seront présentes sous forme explicite dans le réseau morphologique.

3.1.4 Travaux connexes

L'approche globale adoptée ici distingue nettement le modèle informatique que je propose d'autres initiatives visant à constituer des bases de données morphologiques comme le projet MorTAL (Dal et al. 1999, 2004, Hathout et al. 2002). Dans ce projet, la base de données est construite en analysant un à un les dérivés construits par différents affixes comme *-able*, *-ité*, *-iser*, *-ifier*, *-ion*, *-ment*, etc. au moyen de l'analyseur par règles Dérif (Namer 2003a, 2005). Si l'approche théorique sous-jacente est morphématique, Dérif réalise de fait une décomposition des mots en racines et affixes. À l'inverse, dans le travail présenté ici, la totalité du lexique est analysée en une fois.

De nombreux travaux dans le domaine de la morphologie informatique visent à retrouver les relations morphologiques entre les unités lexicales. Tous s'appuient en premier lieu sur les similarités qui existent entre les formes graphiques. Ces relations sont principalement préfixales et suffixales, à deux exceptions près, Yarowsky et Wicentowski (2000) et Baroni et al. (2002) qui utilisent la distance d'édition de Levenshtein pour estimer la similarité formelle entre les mots. À ma connaissance, tous les autres réalisent d'une façon ou d'une autre une segmentation des mots, y compris ceux dont le but n'est pas de trouver des morphèmes comme Hathout (2000, 2001, 2003, 2005) ou Neuvel et Singh (2001), Neuvel et Fulop (2002). Le modèle proposé ici se distingue de ces approches par le fait que les similarités sont déterminées uniquement en fonction du partage des traits formels et sémantiques. C'est l'une des principales contributions de ce travail.

Par ailleurs, le modèle proposé peut aussi être comparé à d'autres approches qui combinent les indices graphémiques et sémantiques pour identifier les morphèmes et les relations morphologiques. Les connaissances sémantiques utilisées sont habituellement acquises automatiquement à partir de corpus au moyen de techniques comme l'analyse sémantique latente (Schone et Jurafsky 2000), l'information mutuelle (Baroni et al. 2002) ou la co-ooccurrence dans une fenêtre de n-mots (Xu et Croft 1998, Zweigenbaum et Grabar 2003a, b, Zweigenbaum et al. 2003). Dans l'expérience présentée ci-dessous, les informa-

tions sémantiques sont extraites d'un dictionnaire informatisé et les similarités formelles sont calculées au moyen de marches aléatoires (*random walks*) dans un graphe lexical. Ce modèle s'inscrit dans la continuité de travaux antérieurs (Hathout 2001, 2003, 2005) dans lesquels j'ai utilisé des ressources lexicographiques pour acquérir des connaissances morphologiques, notamment le dictionnaire de synonymes de l'ATILF et la base de données lexicales WordNet.

3.2 Proximité morphologique

J'adopterai ici une définition minimaliste de la parenté morphologique : deux mots sont morphologiquement apparentés s'ils partagent en même temps des propriétés phonologiques et sémantiques. Dans l'expérience décrite dans ce chapitre, j'ai utilisé les propriétés graphémiques des mots à la place de leurs propriétés phonologiques parce que le TLFi ne fournit pas la prononciation de l'ensemble des entrées, notamment des entrées secondaires qui se trouvent dans les rubriques « Remarques » et « Dérivés » et dans les articles consacrés aux affixes et aux éléments formants. J'envisage de reprendre cette expérience en utilisant Liaphon, le phonétiseur du domaine public de Béchet (2001) qui permet de calculer des représentations phonologiques directement utilisables.

L'idée développée dans cette section est d'utiliser une mesure de similarité morphologique comme approximation de la parenté morphologique : deux mots morphologiquement similaires ont de grandes chances d'être morphologiquement apparentés. Cette similarité morphologique est estimée en utilisant un bigraphe dont la figure 3.2 présente un extrait. Ce bigraphe comporte deux ensembles de sommets : d'un côté ceux qui représentent les lexèmes (dans la partie gauche de la figure) et de l'autre ceux qui représentent leurs traits sémantiques et formels (dans la partie droite). Les sommets qui représentent les lexèmes sont identifiés par leurs lemmes (i.e. leurs formes de citation) et leurs catégories grammaticales. Je signale que ce graphe est conforme au modèle à quatre niveaux (voir section 2.3). Les sommets des traits sémantiques et des traits formels auraient en effet pu être représentés de part et d'autre des sommets qui représentent les lexèmes.

3.2.1 Traits formels et traits sémantiques

Les traits formels associés à un lexème sont les n-grammes de lettres qui apparaissent dans son lemme. On impose aux n-grammes une taille minimale ($n \geqslant 3$). Le début et la fin des lemmes sont marqués par des \$. La figure 3.3 présente une partie des n-grammes associés au mot *orientation* où n varie de 13 à 3. Elle permet de voir que les traits formels associés à un mot donné sont fortement redondants. Une caractéristique importante de cette description est qu'aucun de ces n-grammes n'a le statut de morphème. Tous les n-grammes sont utilisés de la même façon. Ils ont tous la même fonction : rapprocher les mots qui contiennent les mêmes sons.

Il aurait été possible d'utiliser comme traits formels des lexèmes les n-grammes qui apparaissent dans leurs formes fléchies. Une telle description serait plus fidèle à la morphologie de mots (*word-based morphology*). Elle rendrait aussi les allomorphies flexion-

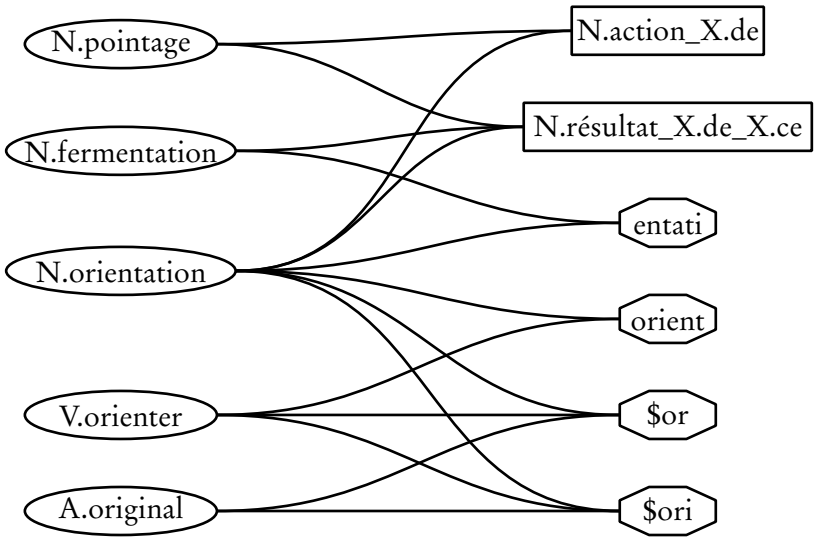


FIGURE 3.2 – Extrait du bigraphe qui représente le lexique. Les lexèmes se trouvent dans des ovales, les traits formels dans des octogones et les traits sémantiques dans des rectangles. Le graphe est symétrique.

nelles disponibles au niveau dérivationnel. Cette option n'a pas été cependant retenue parce qu'elle réduit beaucoup l'homogénéité des représentations formelles. Par exemple, avec un seuil $n \geqslant 3$, un verbe comme *malaxer* se retrouverait fortement rapproché de tous les mots qui contiennent la séquence `xie` comme *anxieux*, *apexien*, *asphyxie*, *orthodoxie*, etc. à cause de la forme *malaxiez*. Pour réduire l'incidence de ce type de traits, il faudrait pondérer la contribution de chaque n-gramme par une estimation de la fréquence des formes du lexème dans lesquelles il apparaît. Cette fréquence pourrait être calculée par exemple à partir d'un gros corpus. Une forme comme *malaxiez* est probablement très rare voire totalement absente dans la plupart des corpus[3]. Les connections superflues amenées par les formes fléchies seraient ainsi affaiblies ou éliminées.

Les traits sémantiques associés à un lexème sont les n-grammes de mots qui apparaissent dans ses définitions. Les n-grammes qui contiennent des ponctuations autres que des apostrophes sont éliminés. En d'autres termes, on utilise uniquement les segments qui apparaissent entre deux ponctuations. Les mots des définitions sont catégorisés et lemmatisés. Les étiquettes utilisées sont : A pour les adjectifs ; R pour les adverbes ; N pour les noms ; V pour les verbes ; X pour toutes les autres catégories. Par exemple, les traits sémantiques induits par la définition (9) du nom *orientation* sont présentés en figure 3.4. On y voit notamment que les traits sémantiques sont tout aussi redondants que les traits formels. Par exemple, deux mots dont les définitions contiennent *résultat de cette action* auront au moins 10 traits sémantiques en commun.

3. Elle est par exemple absente de la base Frantext (`www.frantext.fr`).

```
$orientation$
$orientation orientation$
$orientatio orientation rientation$
$orientati orientatio rientation ientation$
...
$ori orie rien ient enta ntat tati atio tion ion$
$or ori rie ien ent nta tat ati tio ion on$
```

FIGURE 3.3 – Liste partielle des traits formels associés au mot *orientation*. Les traits sont les n-grammes de lettres qui apparaissent dans le mot, avec $13 \geqslant n \geqslant 3$.

(9) *Action d'orienter, de s'orienter ; résultat de cette action.*

```
N.action_X.de_V.orienter
N.action_X.de   X.de_V.orienter
N.action   X.de   V.orienter
X.de_V.s'orienter
V.s'orienter
N.résultat_X.de_X.ce_N.action
N.résultat_X.de_X.ce   X.de_X.ce_N.action
N.résultat_X.de   X.de_X.ce   X.ce_N.action
N.résultat   X.ce
```

FIGURE 3.4 – Traits sémantiques induits par la définition *Action d'orienter, de s'orienter ; résultat de cette action*.

Cette représentation très grossière de la sémantique des mots est inspirée des segments répétés de Lebart et Salem (1994), Lebart et al. (1998). Elle présente plusieurs avantages :

1. elle est fortement redondante et peut ainsi capter différents niveaux de similarité qui existent entre les définitions ;
2. les n-grammes permettent d'intégrer des informations de nature syntagmatique sans nécessiter une véritable analyse syntaxique des définitions ;
3. elle gomme légèrement les variations qui existent dans le traitement lexicographique des entrées, notamment dans les découpages en sous-sens et la formulation des définitions.

3.2.2 Connecter les lexèmes *via* leurs traits

Les traits sémantiques et formels sont utilisés dans le même graphe. Ce graphe est construit en connectant symétriquement chaque entrée à tous ses traits sémantiques et formels (voir figure 3.2). Par exemple, *orientation* est connecté aux traits formels `$or`, `$ori`, `$orie`, `$orien`, etc. qui à leur tour sont connectés aux lexèmes *orienter*, *orientable*, *orientement*, *orienteur*, etc. De même, *orientation* est connecté aux traits sémantiques

`N.action_X.de`, `N.résultat_X.de_X.ce_N.action`, etc. qui sont reliés aux noms *orientement*, *harmonisation*, *pointage*, etc. Les traits sémantiques et formels sont ainsi utilisés de façon similaire. Cette représentation correspond précisément au Modèle en réseau (*Network Model*) de Bybee (1985, 1988, 1995).

Cette représentation redondante peut aussi être rapprochée de celle qui est utilisée par Perruchet et Vinter (1998, 2002) dans Parser, un modèle psycholinguistique de reconnaissance des mots d'un langage artificiel de taille réduite. L'objectif de ces auteurs est de modéliser l'apprentissage de la langue en s'appuyant sur les lois de la mémoire et de l'apprentissage associatif. Ce modèle peut être vu comme une variante dynamique de la méthode proposée par Harris (1955) et affinée par Déjean (1998a, b). Perruchet et Vinter (1998, 2002) défendent l'idée que l'acquisition du langage est conscient. Cela se traduit par l'utilisation de représentations explicites des séquences qui composent le flux analysé par Parser. Ce modèle se distingue sur ce plan des systèmes connectionnistes habituellement utilisés en psycholinguistique pour modéliser l'acquisition du langage comme par exemple par Rumelhart et MacClelland (1986). Il est proche, à l'inverse, de celui qui est présenté ici. Il l'est aussi par l'hypothèse que les représentations linguistiques émergent des données auxquelles apprenants et les locuteurs sont soumis. Le modèle rend compte en outre de l'adaptation de ces représentations au fur et à mesure de l'exposition aux productions langagières. Cette dimension dynamique, pour l'instant absente du modèle que je propose, pourra être l'objet de recherches futures. Notons que Parser est en revanche incapable de traiter des corpus en langues naturelles.

La représentation du lexique par un bigraphe est intéressante parce qu'elle permet de propager une activation de façon synchrone vers les traits sémantiques et formels. Elle l'est aussi parce que le graphe pourrait être utilisé pour la description morphologique. On pourrait par exemple s'en servir pour caractériser la sémantique des adjectifs en *-able* en s'intéressant aux traits sémantiques récurrents dans leurs définitions. De même, il serait possible de découvrir les finales caractéristiques des noms de bateaux en français (10) en identifiant les traits formels qui sont associés de façon privilégiée aux traits sémantiques qui contiennent des mots comme ceux de la liste (11).

(10) a. *voilier* ; *pétrolier* ; *bananier* ; *thonier* ; *sardinier*
b. *patrouilleur* ; *torpilleur* ; *caboteur* ; *dériveur* ; *dragueur*

(11) *bateau* ; *navire* ; *barque* ; *embarcation* ; *bâtiment* ; *chaloupe*

3.2.3 Estimer la similarité morphologique des mots

La similarité morphologique entre un mot et ses voisins est estimée en simulant la propagation d'une activation initiée au niveau du sommet qui représente le mot. Le graphe étant bipartie, l'activation doit être propagée un nombre pair de fois. Dans un graphe fortement redondant comme celui qui est utilisé ici, deux étapes de propagation suffisent à obtenir les estimations visées. Par exemple, si l'on veut déterminer quels sont les voisins les plus proches de *orientation*, on génère une activation au niveau du sommet qui le représente. Cette activation est ensuite propagée de façon uniforme vers les traits sémantiques et formels de *orientation*. Lors de l'étape suivante, l'activation située au niveau des sommets

qui représentent ces traits est propagée vers ceux qui représentent les lexèmes. L'activation reçue par ces derniers est d'autant plus forte que le nombre de traits qu'ils partagent avec *orientation* est élevé et que ces traits sont spécifiques (i.e. qu'ils sont peu fréquents). L'hypothèse sous-jacente est que le niveau d'activation est une estimation du degré de parenté morphologique.

D'un point de vue technique, la propagation est simulée par une marche aléatoire dans le graphe (Gaume 2003, 2004, Gaume et al. 2002, 2004a, b, 2006, Muller et al. 2006). Elle est calculée de façon classique en multipliant la matrice d'adjacence stochastique du graphe. Plus précisément, soit $G = (V, E)$ un graphe composé d'un ensemble de sommets $V = \{v_1, \ldots, v_n\}$ et d'un ensemble d'arcs $E \subset V \times V$. Soit A la matrice d'adjacence de G. A est une matrice $n \times n$ telle que :

$$A_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{sinon} \end{cases}$$

Les lignes de A sont normalisées afin d'obtenir une matrice stochastique M :

$$\forall i \in [1, n], \forall j \in [1, n], M_{ij} = \frac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}$$

Cette valuation des arcs prend en compte le nombre des traits associés à chaque entrée (du fait de la normalisation des valeurs des arcs qui partent des entrées) et la fréquence des traits (du fait de la normalisation des valeurs des arcs qui partent des traits).

Pour simuler une marche de longueur n, on multiplie M n fois. $(M^n)_{ij}$ est alors la probabilité d'atteindre le sommet v_j à partir du sommet v_i après une marche de n étapes. Cette probabilité peut être interprétée comme un niveau d'activation de v_j après une propagation de n étapes initiée au niveau du nœud v_i.

Dans l'expérience présentée dans la suite du chapitre, l'activation est propagée dans le graphe complet pour moitié vers les traits sémantiques et pour moitié vers les traits formels. Cette répartition est obtenue en modifiant les valeurs de la matrice M de la manière suivante. On répartit les arcs du graphe dans trois ensembles $E = F \cup S \cup L$ où F contient les arcs qui connectent une entrée à un trait formel, S les arcs qui connectent une entrée à un trait sémantique et L les arcs qui connectent un trait sémantique ou formel à une entrée. Les valeurs de M sont calculées comme suit :

$$\text{si } (v_i, v_f) \in F, \quad M_{if} = \begin{cases} 0.5 \dfrac{A_{if}}{\sum\limits_{e_{ih} \in F} A_{ih}} & \text{si } v_i \text{ est connecté à un trait sémantique} \\ \dfrac{A_{if}}{\sum\limits_{e_{ih} \in F} A_{ih}} & \text{sinon} \end{cases}$$

$$\text{si } (v_i, v_s) \in S, \quad M_{is} = \begin{cases} 0.5 \dfrac{A_{is}}{\sum\limits_{e_{ih} \in S} A_{ih}} & \text{si } v_i \text{ est connecté à un trait formel} \\ \dfrac{A_{ik}}{\sum\limits_{e_{ih} \in S} A_{ih}} & \text{sinon} \end{cases}$$

$$\text{si } (v_i, v_l) \in L, \quad M_{il} = \frac{A_{il}}{\sum\limits_{e_{ih} \in L} A_{ih}}$$

3.2.4 Voisinages morphologiques

Le graphe utilisé dans l'expérience a été construit à partir des entrées et des définitions du TLFi. Les définitions qui relèvent des usages non standards comme les emplois vieillis, archaïques, argotiques, poétiques, les proverbes, les régionalismes, etc. ont été supprimés. Ce filtrage a aussi éliminé 2 349 entrées. En revanche, les définitions qui concernent des locutions ont été conservées. Le dictionnaire en compte 51 321. L'extraction et le nettoyage des définitions a été réalisée en collaboration avec Bruno Gaume et Philippe Muller, avec l'aide de Jean-Marc Destabeaux. Le graphe bipartie a été créé à partir des 225 529 définitions restantes qui décrivent 75 024 entrées (lexèmes). Elles ont induit environ 9 millions de traits, dont 90% ne sont associés qu'à une entrée. Ces traits hapaxiques ont été supprimés parce qu'ils ne contribuent pas au rapprochement des entrées. Aucune entrée n'a été supprimée du fait de cette réduction : toutes contiennent au moins un trait formel (par exemple un trigramme) ou un trait sémantique (par exemple un mot grammatical) présent ailleurs dans le dictionnaire. La table 3.1 montre que la réduction est plus forte pour les traits sémantiques (93%) que pour les traits formels (69%) : il y a plus de variations dans l'expression du sens que dans la description des formes.

traits	complet	réduit	hapax
formels	1 306 497	400 915	69%
sémantiques	7 650 490	548 641	93%
total	8 956 987	949 556	90%

TABLE 3.1 – Nombre de traits sémantiques et formels du bigraphe. Les traits hapaxiques associés à une seule entrée sont supprimés.

L'utilisation du graphe est illustrée par la figure 3.5 qui présente les 40 voisins les plus proches du verbe *fructifier* pour trois configurations de propagation. Le premier bloc pré-

sente ses voisins dans un graphe qui ne contient que les traits formels. Il fait apparaître que les membres de sa famille morphologique tendent à apparaître en premier et que ceux de sa série dérivationnelle, les verbes en *-ifier* viennent ensuite. Cette répartition découle directement du fait que les familles sont des ensembles plus denses que les séries. La proximité morphologique entre les membres des familles est plus forte que celle qui existe entre les éléments des séries. L'implémentation de la mesure de proximité formelle capture ainsi en partie cette distinction. Elle pourra être exploitée dans des développements futurs pour séparer les familles des séries au sein des voisinages. Les voisins dans le graphe qui ne contient que les traits sémantiques sont présentés dans le deuxième bloc de la figure 3.5. On observe une forte dégradation des résultats due à la plasticité du sens et à la dispersion des acceptions qui en découle, mais aussi à la variabilité dans les descriptions lexicographiques signalée dans le paragraphe précédent. Par exemple, *fructifier* se trouve rapproché de *missionnaire* à cause de la définition (12b) de la locution (12a). La faiblesse de ces résultats montre la nécessité d'améliorer significativement les traits sémantiques extraits des définitions. Le troisième bloc de la figure 3.5 présente les voisins de *fructifier* dans le graphe complet.

(12) a. *Faire fructifier la (les) vigne(s) du Seigneur*
b. *Propager la foi.*

La figure 3.5 montre clairement que les traits formels sont les plus prédictifs et que les traits sémantiques sont les moins fiables. Elle permet également de voir que la proximité morphologique n'est pas suffisamment sélective et que les voisinages ne peuvent pas être utilisés tels quels. Il faut les filtrer davantage. Ce filtrage peut être réalisé au moyen de l'analogie formelle.

3.3 Analogie

3.3.1 Connecter les familles et les séries

Les familles et les séries qui composent la structure morphologique du lexique sont interconnectées par des analogies. Par exemple, la famille du verbe *fructifier* permet de connecter les séries auxquelles appartiennent les mots qui la composent. Ainsi, un couple *fructifier* : *fructification* participe à un nombre important d'analogies avec les autres membres des séries de *fructifier* et de *fructification* comme (13a). De façon duale, la série de *fructifier* permet de connecter les familles des éléments qu'elle contient. Un couple comme *fructifier* : *rectifier* prend part à une petite série d'analogies qui implique les membres des familles de ces deux mots (13b). On voit ainsi que chaque quadruplet analogique est une interconnexion entre deux familles et deux séries (voir figure 3.6).

(13) a. *rectifier* : *rectification* ; *certifier* : *certification* ; *sanctifier* : *sanctification* ; *vitrifier* : *vitrification*
b. *fructificateur* : *rectificateur* ; *fructification* : *rectification* ; *fructifiant* : *rectifiant*

Ces analogies peuvent être utilisées pour sélectionner les parents morphologiques des mots parmi leurs voisins. L'idée générale peut être illustrée en utilisant l'exemple du quadruplet (14). *Fructification* appartenant à la famille de *fructifier* et *rectifier* appartenant à sa

formel	**V.fructifier N.fructification A.fructificateur A.fructifiant A.fructifère V.sanctifier V.rectifier** *A.rectifier V.fructidoriser N.fructidorien N.fructidor* **N.fructuosité R.fructueusement A.fructueux** *N.rectifieur A.obstructif A.instructif A.destructif A.constructif* **N.infructuosité R.infructueusement A.infructueux V.transsubstantifier V.substantifier V.stratifier V.schistifier V.savantifier V.refortifier V.ratifier V.quantifier V.présentifier V.pontifier V.plastifier V.notifier V.nettifier V.mystifier V.mortifier V.justifier V.idiotifier V.identifier**
sémantique	**V.fructifier** *V.trouver N.missionnaire N.mission A.missionnaire N.saisie N.police N.hangar N.dîme N.ban V.affruiter N.melon N.saisonnement N.azédarach A.fruitier A.bifère V.saisonner N.roman N.troubadour V.contaminer N.conductibilité N.alevinage V.profiter* **A.fructifiant** *N.pouvoir V.agir N.opération V.placer N.rentabilité N.jouissance N.avocat N.report* **A.fructueux** *V.tourner V.chiper N.économat N.visa N.société N.réserve N.récréance*
complet	**V.fructifier A.fructifiant N.fructification A.fructificateur** *V.trouver* **A.fructifère V.rectifier V.sanctifier** *A.rectifier V.fructidoriser N.fructidor N.fructidorien N.missionnaire N.mission A.missionnaire* **A.fructueux R.fructueusement N.fructuosité** *N.rectifieur N.saisie N.police N.hangar N.dîme N.ban A.fruitier V.affruiter A.instructif A.obstructif A.destructif A.constructif N.conductibilité V.saisonner N.melon N.saisonnement N.azédarach A.bifère V.contaminer N.roman N.troubadour N.alevinage*

FIGURE 3.5 – Les 40 plus proches voisins du verbe *fructifier* lorsque l'activation est propagée seulement vers les traits formels (premier bloc), uniquement vers les traits sémantiques (deuxième bloc) ou à la fois vers les traits formels et sémantiques (troisième bloc). Les mots qui appartiennent à la famille morphologique ou à la série dérivationnelle *fructifier* sont en gras ; les autres sont en italiques.

série, ces deux mots sont des voisins morphologiques de *fructifier*. De même, *rectification* appartenant à la famille de *rectifier* et à la série de *fructification*, c'est un voisin morphologique de ces deux mots. Ces relations peuvent être représentées graphiquement comme en figure 3.7.

(14) *fructifier* : *fructification* :: *rectifier* : *rectification*

Réciproquement, si on suppose que les voisins d'un mot lui sont morphologiquement apparentés, on peut les utiliser pour former des quadruplets analogiques de la manière suivante :

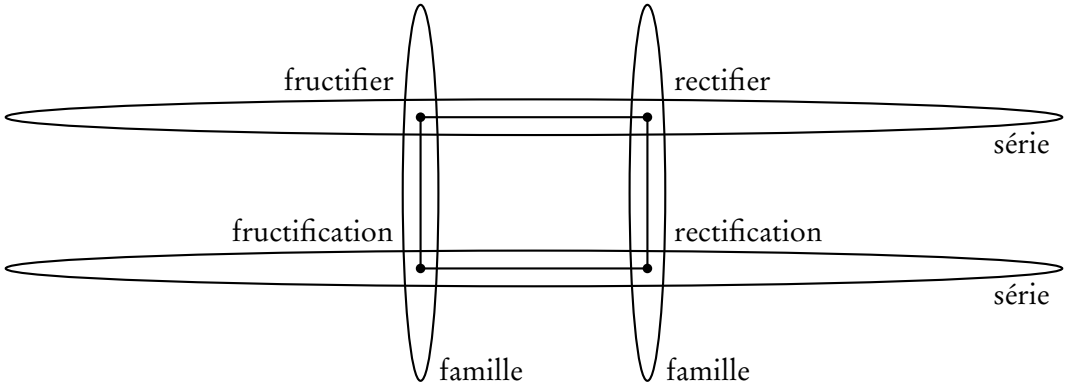


FIGURE 3.6 – L'analogie *fructifier* : *fructification* :: *rectifier* : *rectification* interconnecte les familles de *fructifier* et de *rectifier* et les séries des verbes en *-ifier* et des noms en *-ion*.

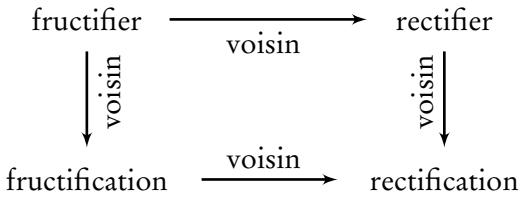


FIGURE 3.7 – Relations de voisinage au sein du quadruplet *fructifier* : *fructification* :: *rectifier* : *rectification*.

pour chaque *mot* a **faire**
  pour chaque *couple* (b, c) *de voisins de* a **faire**
    pour chaque d *voisin de* b *et de* c **faire**
      vérifier que (a, b, c, d) est une analogie
    fin
  fin
fin

3.3.2 Analogies formelles

Une analogie formelle ou analogie graphémique est une relation a : b :: c : d entre quatre chaînes de caractères telles que les différences qui existent entre a et b sont les mêmes que celles qui existent entre c et d. C'est par exemple le cas pour le quadruplet (14). Cette identité des différences est représentée dans la partie gauche de la figure 3.8. Naturellement, les couples de graphies peuvent présenter plusieurs différences comme on peut le voir pour l'analogie entre les quatre mots arabes en (15). Ces mots sont respectivement les

transcriptions du verbe 'écrire', du nom 'document', du verbe 'faire' et du nom 'effet' [4]. Les différences entre les mots du premier couple et entre ceux du second sont représentées dans la partie droite de la figure 3.8. Elles sont identiques. Cet exemple montre que les analogies formelles s'établissent aussi entre les parents morphologiques dans une langue gabaritique comme l'arabe. Je rappelle que la mesure de proximité pourrait facilement être adaptée à la morphologie de ce type de langue (voir section 2.4).

(15) *kataba* : *maktoubon* :: *fa3ala* : *maf3oulon*

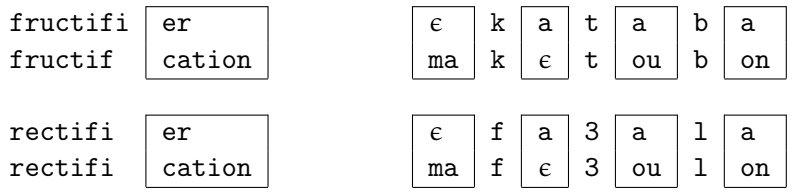


FIGURE 3.8 – Analogies formelles *fructifier* : *fructification* :: *rectifier* : *rectification* et *kataba* : *maktoubon* :: *fa3ala* : *maf3oulon*. Les différences se trouvent dans les parties encadrées. ϵ représente la chaîne vide.

Formellement, ces analogies peuvent être définies en termes de factorisation (Yvon et al. 2004, Stroppa et Yvon 2005) de la manière suivante :

Définition 3.1 FACTORISATION
Soit Σ *un alphabet et* $a \in \Sigma^\star$ *une chaîne de caractères définie sur* Σ. *Une factorisation de longueur* n *de* a *est une séquence de* n *facteurs* $f = (f_a^1, \cdots, f_a^n) \in \Sigma^{\star n}$ *telle que* $a = f_a^1 \odot \cdots \odot f_a^n$ *où* $\odot$ *représente la concaténation.*

Par exemple, (ma, k, ϵ, t, ou, b, on) est une factorisation de longueur 7 de la chaîne *maktoubon*.

Définition 3.2 ANALOGIE FORMELLE
Soient $(a, b, c, d) \in \Sigma^{\star 4}$ *un quadruplet de chaînes de caractères définies sur un alphabet* Σ. (a, b, c, d) *est une analogie formelle si et seulement si il existe* $n \in \mathbb{N}$ *et quatre factorisations de longueur* n (f_a, f_b, f_c, f_d) *de* a, b, c *et* d *telles que* $\forall i \in [1, n], (f_b^i, f_c^i) \in \{(f_a^i, f_d^i), (f_d^i, f_a^i))\}$.

Par exemple, les factorisations présentées en figure 3.9 permettent de vérifier que (15) est bien d'une analogie formelle. Ainsi, on a $(f_b^2, f_c^2) = (f_a^2, f_d^2) = (k, f)$ et $(f_b^3, f_c^3) = (f_d^3, f_a^3) = (\epsilon, a)$.

3.3.3 Implémentation

Les analogies formelles sont vérifiées au niveau graphémique. Les différences entre les deux couples de chaînes de caractères peuvent être calculées à partir des séquences d'opérations d'édition qui permettent de transformer le premier élément de chaque couple en

4. Cet exemple est inspiré de Lepage (1998, 2003).

	f^1	f^2	f^3	f^4	f^5	f^6	f^7
a	ɛ	k	a	t	a	b	a
b	ma	k	ɛ	t	ou	b	on
c	ɛ	f	a	3	a	l	a
d	ma	f	ɛ	3	ou	l	on

FIGURE 3.9 – Factorisations de longueur 7 permettant de vérifier que (15) est une analogie formelle.

l'autre (Lepage 1998). Les opérations d'édition considérées sont l'insertion d'un caractère (notée I), la suppression d'un caractère (notée D) et le remplacement d'un caractère par un autre (notée S). On notera M les caractères qui sont inchangés. La figure 3.10 présente une séquence permettant de transformer *fructueux* en *infructueusement*.

I	I	M	M	M	M	M	M	M	M	S	I	I	I	I	I
ɛ	ɛ	f	r	u	c	t	u	e	u	x	ɛ	ɛ	ɛ	ɛ	ɛ
i	n	f	r	u	c	t	u	e	u	s	e	m	e	n	t

FIGURE 3.10 – Séquence des opérations d'édition qui transforme *fructueux* en *infructueusement*. Le type des opérations est indiqué sur la première ligne : D pour les suppressions, I pour les insertions, S pour les substitutions et M pour les identités.

Chaque séquence a un coût qui est calculé à partir de celui des opérations élémentaires qui la composent. On utilise habituellement les valeurs suivantes : 1 pour l'insertion d'un caractère ; 1 pour la suppression d'un caractère ; 1 pour la substitution d'un caractère ; 0 pour la conservation d'un caractère. La distance de Levenshtein d'un couple de chaînes est le coût minimum des séquences d'édition permettant de transformer la première en la seconde. Elle est calculée par un algorithme de programmation dynamique en construisant une matrice semblable à celle qui est présentée en figure 3.11 (Jurafsky et Martin 2000). Cette matrice représente un treillis d'édition. Chaque séquence d'édition correspond à un chemin dans le treillis.

Le chemin utilisé pour déterminer les différences entre deux chaînes de caractères est calculé à partir de cette matrice de la manière suivante. Le point de départ est la dernière cellule de la matrice (cellule encadrée dans la figure 3.11) et celui d'arrivée est la première. Le chemin est constitué en sélectionnant comme cellule suivante la voisine de coût le plus faible. En cas d'égalité, on préfère la cellule de gauche (qui correspond à une insertion) ou à défaut celle du dessus (suppression) ou enfin celle qui se trouve sur la diagonale supérieure gauche (substitution). L'idée sous-jacente est d'utiliser le plus court chemin afin de préserver les segments communs aux deux chaînes et d'appliquer des critères de sélection uniformes pour permettre la comparaison des différences. L'ordre de préférence relatif des trois types d'opération n'est donc pas important. La figure 3.11 présente le chemin utilisé pour identifier les différences entre les chaînes *fructueux* et *infructueusement*. La séquence d'édition décrite par ce chemin est présentée en figure 3.10.

	◇	i	n	f	r	u	c	t	u	e	u	s	e	m	e	n	t
◇	0	← 1	← 2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
f	1	1	2	↖ 2	3	4	5	6	7	8	9	10	11	12	13	14	15
r	2	2	2	3	↖ 2	3	4	5	6	7	8	9	10	11	12	13	14
u	3	3	3	3	3	↖ 2	3	4	5	6	7	8	9	10	11	12	13
c	4	4	4	4	4	3	↖ 2	3	4	5	6	7	8	9	10	11	12
t	5	5	5	5	5	4	3	↖ 2	3	4	5	6	7	8	9	10	11
u	6	6	6	6	6	5	4	3	↖ 2	3	4	5	6	7	8	9	10
e	7	7	7	7	7	6	5	4	3	↖ 2	3	4	5	6	7	8	9
u	8	8	8	8	8	7	6	5	4	3	↖ 2	3	4	5	6	7	8
x	9	9	9	9	9	8	7	6	5	4	3	↖ 3	← 4	← 5	← 6	← 7	← 8

FIGURE 3.11 – Matrice des distances d'édition de Levenshtein et chemin décrivant la séquence d'opérations d'édition permettant de transformer *fructueux* en *fructueusement*. ◇ représente le début de la chaîne. La cellule (i, j) de la matrice indique la distance d'édition entre la sous-chaînes des i premiers caractères de *fructueux* et celle des j premiers caractères de *infructueusement*.

Les séquences d'opérations d'édition peuvent être simplifiées en fusionnant les suites d'identités (M). La séquence de la figure 3.10 devient alors (16). Cette représentation fait ressortir les différences entre les chaînes. Ces différences sont par exemple les mêmes que celles du couple *soucieux* : *insoucieusement* (17) avec lequel *fructueux* : *infructueusement* forme une analogie. Les représentations (16) et (17) peuvent être rendues totalement identiques en remplaçant leurs sous-chaînes communes par un caractère joker @ comme en (19). On obtient alors une représentation que l'on peut utiliser comme une signature σ du couple. Cette signature correspond précisément aux parties encadrées de la figure 3.8. L'analogie formelle (18) peut être ainsi réduite à une comparaison des signatures $\sigma($`fructueux`, `infructueusement`$)$ et $\sigma($`soucieux`, `insoucieusement`$)$. Plus généralement, quatre chaînes $(a, b, c, d) \in \Sigma^{\star 4}$ constituent une analogie formelle si et seulement si $\sigma(a, b) = \sigma(c, d)$.

(16) ((I,ϵ,i), (I,ϵ,n), (M,fructueu,fructueu), (S,x,s), (I,ϵ,e), (I,ϵ,m), (I,ϵ,e), (I,ϵ,n), (I,ϵ,t))

(17) ((I,ϵ,i), (I,ϵ,n), (M,soucieu,soucieu), (S,x,s), (I,ϵ,e), (I,ϵ,m), (I,ϵ,e), (I,ϵ,n), (I,ϵ,t))

(18) *fructueux* : *infructueusement* :: *soucieux* : *insoucieusement*

(19) ((I,ϵ,i), (I,ϵ,n), (M,@,@), (S,x,s), (I,ϵ,e), (I,ϵ,m), (I,ϵ,e), (I,ϵ,n), (I,ϵ,t))

Notons qu'il possible de réduire le nombre de quadruplets considérés par l'algorithme de la section 3.3.1 en utilisant quelques heuristiques simples comme le fait que :

– les différences de longueur dans chaque couple doivent être identiques (Lepage 2000) : $l(a) - l(b) = l(c) - l(d)$ où $l(x)$ représente la longueur de x ;

– les catégories des entrées sont en relation analogique : $(cat(b), cat(c)) \in \{(cat(a), cat(d)), (cat(d), cat(a))\}$ où $cat(x)$ représente la catégorie de x.

En revanche, l'heuristique de Langlais et Yvon (2008) qui consiste à imposer aux premiers caractères des quatre chaînes et à leurs derniers caractères de former des analogies est trop forte.

Pour finir, j'insiste sur le fait que la vérification des analogies formelles ne fait pas intervenir les notions de morphème, d'affixe ou d'exposant morphologique. Elle consiste uniquement en une mise en correspondance de couples de chaînes de caractères.

3.4 Premiers résultats

Le modèle informatique présenté dans ce chapitre n'est pas un analyseur complet permettant de créer l'ensemble du réseau morphologique décrit en section 3.1.3. Il s'agit d'un travail exploratoire dont l'objectif est de confirmer l'opérationnalité de ce modèle et de mettre au point l'implémentation des composants de base de l'analyseur que sont la proximité morphologique et l'analogie formelle. Une évaluation détaillée de ces composants est donc pour l'heure prématurée. C'est pourquoi je ne présente que quelques résultats partiels et préliminaires.

Ces résultats ont été obtenus en identifiant les 100 plus proches voisins de chaque mot du lexique. J'ai ensuite calculé les analogies formelles obtenues à partir de 22 entrées, appartenant à 4 familles morphologiques de la tranche *fr-* du *TLF* pour laquelle je dispose d'une classification révisée manuellement par Josette Lecomte et moi (20). J'ai révisé ces analogies manuellement en considérant comme corrects les quadruplets (a, b, c, d) qui satisfont (21a) ou (21b). Des exemples de quadruplets corrects sont présentés en (22) et de quadruplets erronés en (23). On peut voir que les analogies collectées réunissent des dérivés et leurs bases comme en (22a), mais aussi des dérivés qui partagent la même base comme en (22b) ou des relations plus distantes entre un dérivé et la base de sa base comme en (22c).

(20) a. S.fructidor ; A.fructidorien ; S.fructidorien ; S.fructidorienne ; V.fructidoriser

b. A.fructifiant ; S.fructification ; V.fructifier ; A.fructificateur ; A.fructifère ; A.fructivore ; S.fructose ; R.fructueusement ; A.fructueux ; S.fructuosité

c. A.frugiforme ; A.frugivore ; S.frugivorisme

d. S.fruit ; S.fruitage ; A.fruiteux ; A.fruitarien ; S.fruitarisme ; S.fruiterie ; A.fruitier ; S.fruitier ; S.fruitière ; A.fruité ; A.fruticole

(21) a. b appartient à la famille de a ; c appartient à la série de a ; d appartient à la série de b et à la famille de c

b. b appartient à la série de a ; c appartient à la famille de a ; d appartient à la famille de b et à la série de c

(22) a. N.fructification:N.identification :: V.fructifier:V.identifier

b. A.fructifiant:A.fructificateur :: A.glorifiant :: A.glorificateur

c. A.fructueux:A.affectueux :: N.infructuosité:N.inaffectuosité

d. A.frugivore:A.végétivore :: R.frugalement:R.végétalement
e. A.fruitarien:A.végétarien :: N.fruitarisme:N.végétarisme
f. A.fruitier:A.laitier :: N.fruiterie:N.laiterie
g. R.fructueusement:R.affectueusement :: N.fructuosité:N.affectuosité

(23) a. A.fruité:N.fruste :: A.truité:N.truste
b. N.fruit:N.frumentaire :: A.instruit:A.instrumentaire
c. N.fruiterie:N.friterie :: V.effruiter:V.effriter

Les trois configurations de propagation présentées dans la section 3.2.4 ont été testées pour ces 22 entrées. Les résultats sont résumés dans la table 3.2. Ils font apparaître une qualité satisfaisante pour les analogies obtenues à partir des traits formels et dans le graphe entier. En revanche, la contribution des traits sémantiques est insuffisante. Ces chiffres confirment la tendance générale qui se dégage des exemples de la figure 3.5. Les traits sémantiques réduisent d'environ un quart le nombre des analogies collectées ; en contrepartie, ils améliorent la précision de la méthode.

configuration	analogies	correctes	erreurs
formel	169	163	3.6%
sémantique	5	5	0.0%
complet	130	128	1.5%

TABLE 3.2 – Nombre d'analogies collectées pour un échantillon de 22 entrées pour les trois configurations présentées dans la section 3.2.4. La dernière colonne indique le taux d'erreur.

Ces résultats montrent également que la méthode s'appuie avant tout sur les proximités formelles. De ce fait, ses performances sont directement liées à la taille des entrées, la proximité formelle étant d'autant plus grande que les mots sont longs. Par exemple, *départementaliser* est plus proche de *départementalisation* que ne l'est *créer* de *création* du fait du rapport entre le nombre de traits qu'ils partagent et le nombre total des traits qu'ils possèdent en propre. La table 3.3 met en évidence cette corrélation. Elle présente le nombre d'analogies collectées et le taux d'erreur pour 13 échantillons de 5 entrées sélectionnés aléatoirement. Les analogies ont été collectées dans le graphe complet. Les entrées dans chaque échantillon ont toutes la même longueur. Cette dernière varie entre 4 et 16 caractères. On observe que les analogies collectées pour les mots de 10 lettres et plus sont toutes correctes.

3.5 Conclusion et développements futurs

J'ai présenté dans ce chapitre un modèle informatique qui implémente quelques-uns des éléments fondamentaux du modèle théorique du chapitre 2. Ce modèle permet de faire émerger une partie de la structure morphologique du lexique à partir des entrées et des

longueur	analogies	correctes	erreurs
4	29	14	51.7%
5	22	14	36.4%
6	8	7	12.5%
7	10	8	20.0%
8	55	54	1.8%
9	29	27	6.9%
10	30	30	0.0%
11	32	32	0.0%
12	19	19	0.0%
13	11	11	0.0%
14	35	35	0.0%
15	63	63	0.0%
16	39	39	0.0%

TABLE 3.3 – Nombres d'analogies et taux d'erreur pour 13 échantillons sélectionnés aléatoirement et composés chacun de 5 entrées. Les entrées dans chaque échantillon sont de même longueur. Les longueurs varient 4 à 16 caractères. Les analogies obtenues à partir des entrées de 10 caractères et plus sont toutes correctes.

définitions d'un dictionnaire informatisé, en l'occurrence le TLFi. Ce modèle est purement lexématique. Il associe deux techniques complémentaires. La première est une mesure de la proximité morphologique calculée en utilisant un algorithme de marche aléatoire dans un bigraphe qui intègre de manière uniforme les traits sémantiques et formels des mots. La marche aléatoire simule la propagation d'une activation dans le graphe lexical. Le niveau d'activation obtenu après cette propagation permet d'identifier les voisins de chaque entrée. L'autre technique est l'analogie formelle utilisée pour identifier les membres des familles morphologiques et des séries dérivationnelles à l'intérieur des voisinages morphologiques.

Ce modèle me semble prometteur parce qu'il est avant tout informatique. Sa mise en œuvre repose sur une définition minimaliste de la parenté morphologique et sur l'hypothèse que les mots apparentés forment des analogies. Il exploite essentiellement la capacité de la mémoire et la puissance de calcul des processeurs. Ce modèle est également intéressant parce que les propriétés sémantiques et formelles sont indépendantes: elles ne sont connectées que via les entrées. Le bigraphe utilisé est ainsi une implémentation relativement fidèle du modèle à quatre niveaux présenté en section 2.3. L'une des conséquences de cette indépendance est la capacité du système à traiter de manière uniforme toutes les relations dérivationnelles, y compris celles qui sont parasynthétiques (voir section 4.5).

Ce travail, encore préliminaire, ouvre un chantier relativement vaste dont l'objectif premier est la constitution d'une base de données morphologiques à large couverture. J'envisage d'en réaliser une première version à court terme en utilisant une méthode de *bootstrap* en créant dans un premier temps un réseau partiel à partir des entrées les plus longues puis en le complétant itérativement avec des mots de plus en plus courts. Pour créer cette base,

il faudra aussi concevoir une méthode permettant de séparer les familles des séries au sein des voisinages morphologiques. On pourra utiliser le fait que la famille et la série d'un mot sont généralement disjoints, qu'ils correspondent approximativement à des composantes connexes dans les voisinages, que les éléments des familles sont plus fortement interconnectés que ceux des séries, que ces dernières sont des ensembles de grande taille, etc. Une fois construit le réseau des paradigmes morphologiques, on pourra le compléter par les paradigmes lexicaux et traiter la question de l'allomorphie. Ces deux sous-tâches mettent en jeu de façon cruciale les relations sémantiques. L'amélioration de leur description est un préalable à leur réussite.

Le modèle devra d'autre part être complété en proposant un traitement de la composition et de la conversion. Les techniques proposées dans ce chapitre peuvent être utilisées telles quelles pour retrouver les relations morphologiques qui impliquent les mots formés par ces procédés. En revanche, les paradigmes décrits en section 2.6.1 ne permettent pas de représenter ces relations. Il faudrait probablement les compléter par des **séries compositionnelles** et des **séries de converts**. Ces dernières ne pourront pas être constituées sans une bonne caractérisation sémantique des mots. L'intégration des conversions au système dépend ainsi principalement de la qualité des traits sémantiques utilisés.

Le deuxième volet de ce programme est l'amélioration des traits sémantiques qui sera réalisée en procédant à l'analyse syntaxique des définitions. Il sera ainsi possible de remplacer les segments répétés extraits des définitions par des fragments redondants de leurs arbres de dépendances syntaxiques. On pourra par exemple utiliser une chaîne de traitement linguistique comme celle du projet Alpage (Boullier et al. 2005). Ces données pourront être complétées par des connaissances distributionnelles similaires aux *Voisins de Le Monde* (Galy et Bourigault 2005).

Le troisième volet est l'évaluation du système. Un protocole d'évaluation adapté aux résultats de l'analyse devra être mis au point. À ma connaissance, il n'existe à l'heure actuelle aucune ressource de référence du type de celle que je propose. Cette évaluation sera d'abord réalisée en anglais pour bénéficier des analyses contenues dans la base CELEX (Baayen et al. 1995). La construction du réseau morphologique anglais pourra être réalisée en utilisant le dictionnaire informatique GCIDE basé sur l'édition de 1913 du *Webster*. Elle se fera d'autre part en français en comparant le réseau morphologique aux résultats d'autres systèmes comme *Linguistica* de Goldsmith (2001) ou l'analyseur de Bernhard (2006a, b) qui a remporté l'édition 2007 de *Morpho Challenge* pour l'anglais et le finnois (Kurimo et al. 2007). La comparaison avec CELEX et les résultats de ces systèmes impliquera un découpage des mots et une identification des allomorphes. Cette tâche n'est pas triviale (Langlais 2009).

J'envisage également, en collaboration avec Gilles Boyé, d'utiliser le réseau morphologique comme support pour la description morphologique du français en l'intégrant dans un système de type Wiki permettant de l'annoter en couches successives (voir section 1.3.5). Le réseau morphologique pourra également être utilisé à des fins lexicographiques pour normaliser les définitions des mots construits morphologiquement en comparant les proximités sémantiques calculées à partir des définitions et les proximités morphologiques dans le réseau. Ce modèle informatique pourra également être utilisé à des fins typologiques

pour comparer les structures des réseaux morphologiques de langues différentes (Prévot et al. 2008).

CHAPITRE 4

Analyse unifiée de la préfixation en *anti-*

4.1 Introduction

L'analyse de la préfixation en *anti-* présente trois types de difficultés bien connues qui seront abordées successivement dans ce chapitre. Ces difficultés ont été identifiées dans les nombreux travaux antérieurs qui traitent de cette préfixation, notamment Rey (1968), Dell (1970), Zribi-Hertz (1973), Corbin (1980), Durand (1982), Voir (1982), Corbin (1987), Fradin (1996, 1997a, b), Montermini (2002), Levkovych (2004), Plénat et al. (2004a), Heyna (2008), Montermini (2008). Ce chapitre présente une analyse unifiée de cette préfixation [1]. Mon analyse n'est naturellement pas complète. Le traitement de certaines questions est seulement esquissé et reste à approfondir. Il va de soi que mes propositions sont en grande partie inspirées des travaux cités ci-dessus.

Le premier point que j'aborde est la variété des séquences susceptibles d'apparaître après *anti-*. On trouve en effet parmi ces dernières des noms communs au singulier (1a) ou au pluriel (1b), des noms propres (1c), des composés VN lexicalisés (1d) ou non (1e), des syntagmes nominaux sans déterminant (1f) ou avec déterminant (1g), des pronoms (1h), des interjections (1i), des groupes propositionnels (1j), des adjectifs (1k) et des verbes (1l). Cette variété soulève plusieurs questions : *anti-* est-il un préfixe ou une préposition ? Existe-t-il une préposition *anti* similaire à *contre* ? Si c'est un préfixe, quel est le statut des radicaux syntaxiques et des radicaux au pluriel ? Je détaille en section 4.2 ce petit inventaire des objets qui apparaissent après *anti-* et je les compare à ceux que l'on trouve devant différents suffixes.

(1) a. *antilimace* ; *antihéros* ;
b. *antirides* ; *anti-animaux* ;
c. *anti-Sarkozy* ; *anti-Sartre* ;

1. Je remercie Marc Plénat et Michel Roché pour les exemples et les arguments qu'ils m'ont proposés.

d. *anti-porte-bagage* ;
e. *vernis spécial anti-mange-ongles* ; *la cuillère anti-brûle-la-langue* ;
f. *anti-gros-lard* ; *anti-homme de lettres* ; *anti-crottes de chiens sur les trottoirs* ;
g. *anti les mecs qui font ça sur route ouverte* ;
h. *antirien* ; *anti-ceci ou cela* ;
i. *antiplouf* ; *anti-aïe aïe* ; *antigloups* ;
j. *anti-ceux-qui-défendent-apple-becs-et-ongles* ;
k. *antitriste*, *anti-obèse* ;
l. *antidéprimer* ; *anti-manger sain*

La deuxième question à laquelle je m'intéresse est l'existence de plusieurs sous-séries de dérivés en *anti-* dont les sens sont à première vue relativement distincts. Il y a d'une part les dérivés du type *antilimace* ou *antirides* habituellement paraphrasés au moyen de la préposition *contre* (un produit antilimace est un produit contre les limaces) et les dérivés du type *antihéros* ou *antiroman*, généralement paraphrasés en utilisant l'expression « tout le contraire de » (un antihéros est un personnage qui est tout le contraire d'un héros). Il existe en outre plusieurs sous-séries plus réduites de dérivés en *anti-*, notamment celles des dérivés qui ont un sens spatial comme *antisolaire* ou *Anti-Atlas*, celles des termes de la physique des particules comme *antiquark* ou *antibeau* et celles des termes mathématiques comme *antisymétrique* ou *antiparallèle*. Se pose ainsi la question de l'unicité du préfixe *anti-* : ces sous-séries relèvent-elles d'un seul préfixe *anti-* ou de plusieurs ? Et, dans le cas où il n'y en aurait qu'un, comment s'organisent les sens correspondant à ces sous-séries ? Quelques éléments de réponse à ces questions sont proposés dans la section 4.3. Suit en section 4.4 une discussion des aspects catégoriels.

Le troisième point abordé est la fameuse question des constructions dites parasynthétiques : l'existence de dérivés comme *antiparlementaire*, *antigrippal* ou *antipaternel* dont les radicaux sont des formes adjectivales mais dont le sens est le même que celui des dérivés qui ont comme radicaux les noms correspondants : *antiparlement*, *antigrippe* et *antipère*. Les dérivés *antiparlementaire*, *antigrippal* ou *antipaternel* sont traditionnellement analysés comme des mots construits par parasynthèse, c'est-à-dire par une opération qui comporte simultanément une préfixation et une suffixation. Par exemple, la parasynthèse qui permet de créer *antiparlementaire* se compose d'une préfixation en *anti-* et d'une suffixation adjectivale en *-aire*. Je propose, en section 4.5, une analyse de ces dérivés différente, formulée en termes d'emprunt de radicaux. Cette analyse s'inscrit dans le cadre du modèle théorique du chapitre 2, dans la lignée de Plénat (2000), Corbin (2001a), Roché (2009). Elle est illustrée en utilisant les données présentées dans Levkovych (2004) et Plénat et al. (2004a).

Sur le plan méthodologique, l'étude présentée ici se caractérise par son inscription dans la morphologie extensive (Plénat 2000, Plénat et al. 2002, Hathout et al. 2003, 2008, 2009a) ; voir aussi Durand (2009) pour une mise en perspective épistémologique. Elle s'appuie sur un corpus étendu de près de 3 000 mots constitué en complétant celui de Plénat et al. (2004a) et Levkovych (2004). Seule une description globale d'un corpus aussi grand que possible permet en effet de saisir l'ensemble des conditionnements d'une dérivation. Cette étude se

distingue aussi par le type d'analyse visé : il s'agit de déterminer la place des dérivés en *anti-* dans le lexique et l'ensemble des relations qu'ils entretiennent avec les autres mots et pas simplement trouver leurs bases et les relations qu'ils entretiennent avec ces dernières.

4.2 *Anti-* est un préfixe

Je rappelle tout d'abord qu'il existe en français quelques mots simples[2] dont l'initiale est *anti-* comme *antichrèse*, *antimoine* ou *antiphonaire*. Ces mots n'entrent naturellement pas dans le champ de cette étude. Je rappelle également que la formation des mots en *anti-* n'a pas toujours été considérée comme une préfixation. Dans la tradition (Darmesteter 1877)[3], les préfixes étaient considérés comme des prépositions ou des adverbes et les mots préfixés analysés comme des composés. Cette conception est par exemple adoptée par Zribi-Hertz (1973). Pour Rey (1968), *anti-* est un élément de composition qui a ensuite été préfixé à des bases autonomes. Ce ne serait donc ni un préfixe ni une préposition. Par ailleurs, il considère que les dérivés dont le radical est un nom sont des locutions adjectivales. À l'inverse, *anti-* est un préfixe pour Dell (1970), Corbin (1980) ou Fradin (1997b). Ce dernier observe que *anti-* se préfixe volontiers à des bases autochtones et qu'il dispose d'une grande productivité. Celle-ci témoigne de sa pleine intégration dans le système morphologique du français. Fradin en déduit que cette formation n'est pas une composition.

La proximité et les différences qui existent entre les préfixes et les prépositions ont fait l'objet de travaux très nombreux dont on peut trouver un compte-rendu complet dans Montermini (2002, 2008) ou Amiot et Montermini (2009). Je n'apporte pas ici d'éléments nouveaux par rapport à ces analyses. Ma contribution se limite à montrer l'étendue de la diversité des séquences qui comportent l'initiale *anti-* et à vérifier leur compatibilité avec la catégorisation de *anti-* comme un préfixe.

On peut distinguer parmi les dérivés en *anti-* ceux dont les radicaux ne peuvent pas fonctionner tels quels comme des compléments de prépositions (sections 4.2.1, 4.2.2, 4.2.3 et 4.2.4) et ceux qui le peuvent (sections 4.2.5, 4.2.6 et 4.2.7).

4.2.1 Radicaux nominaux, adjectivaux et verbaux

Le fait que *anti-* soit utilisé comme un préfixe dans la plupart des mots qui comportent cette initiale ne fait aucun doute. C'est le cas pour les plus prototypiques d'entre eux dont les radicaux sont des formes de citation, monolexicales, nominales ou adjectivales comme *antilimace*, *antihéros* (2), *antiparlementaire* (3), *antigélif*, *antitriste* (4) ou *antistérile* (5). Je signale que pour Dell (1970), Zribi-Hertz (1973), Durand (1982) entre autres, les dérivés en *anti-* ne peuvent pas avoir d'adjectifs simples comme bases. Les deux derniers exemples infirment cette hypothèse.

(2) *Dans le monde des jeux vidéo, on parle souvent des héros de jeux vidéo, mais rarement des* ***anti-héros****.* `www.jeuxvideo.fr/forum/forum-general/les-anti-heros-`

2. Par « mot simple », j'entends ici qu'il n'est pas possible de les analyser en synchronie comme des dérivés morphologiques.

3. Cité par Dal (2003b).

dans-les-jeux-video-id9654-page1.html

(3) *Pepito : le discours* ***anti-parlementaire*** *qui consiste à les considérer comme des paresseux relève d'un niveau zéro de la réflexion ou au moins de l'information.* societales.blogs.liberation.fr/alain_piriou/2008/05/pacs-et-ogm-qua.html

(4) *Trop rigolo, Julie est une potion* ***anti-triste*** *!* julieadore.blogspot.com/2009/06/betty-et-evirgo.html

(5) *SANTOFEROL : Préparation concentrée de vitamine E* ***anti-stérile****, dite de reproduction.* perso.wanadoo.fr/artese/animaux/prod/ferol.htm

On peut compléter ce premier groupe par les dérivés dont les radicaux sont des formes verbales comme *antitranspirant* (6), *anticoupant* (7) ou *antibasculant*.

(6) *j'ai une veste longue hiver Bogner et un maillot* ***anti-transpirant****, touchés en promo.* forum.velovert.com/index.php?showtopic=75533

(7) *une veste de sécurité de type « bûcheron » et une paire de bottes de type «* ***anti-coupant*** *» ...* www.chaudfontaine.be/Vac/101/02_conseils_communaux.pdf

L'initiale *anti-* apparaît plus rarement devant des infinitifs comme *antidéraper* (8), *antidéprimer* (9), *antimanger* (11), *antidormir* (12), *antimordre* (78) ou *antirire* (79). Les deux premiers exemples sont probablement des « rétroformations » créées à partir de *antidérapant* et *antidépresseur*. Ces infinitifs permettent de régulariser le paradigme dérivationnel de ces adjectifs et d'en faire des participes présents à part entière. Les premiers restent cependant très proches des seconds et peuvent dans ces contextes être paraphrasés par (10a) et (10b). Dans les exemples (11 ; 12), les radicaux *manger sain* et *dormir* pourraient aussi être interprétés comme des propositions infinitives.

(8) *— j'ai oublié de préciser que les semelles shopping and dancing de Scholl sont indispensables pour la survie de tes pieds...*
— ... mais qu'il ne faut pas leur faire une confiance exagérée non plus, sauf pour ***anti-déraper*** *dans tes chaussures si tu portes des bas.* forum.cosmopolitan.fr/cosmo/tendance-shopping-fashion-look/chaussures-sac-lunettes/existentielle-chaussures-question-sujet_391_2.htm

(9) *Xanax, toujours à votre service pour vous* ***antidéprimer!*** www.e-carabin.net/showthread.php?t=45675&page=3

(10) a. *... sauf comme* ***antidérapant*** *dans tes chaussures ...*

b. *... toujours à votre service comme votre* ***antidépresseur****.*

(11) *Les obèses, les fumeurs, les buveurs, les anti-sports, les* ***anti-manger sain****, les anti-"je veux mourir en bonne santé", voilà les nouveaux hérétiques.* vitanova.blogspot.com/2005_04_01_archive.html

(12) *mais quand on voit que les transporteurs ne veulent meme pas dépenser 1000€ pour une barre anti-ancastrement (:colere::evil:), je les vois mal équiper leur tracteur du systeme* ***anti-dormir****.* www.forum-auto.com/automobile-pratique/section16/sujet69821.htm

La découverte de ces dérivés découle directement de l'approche extensive. Ces exemples sont en effet très rares en comparaison de ceux dont les radicaux sont nominaux ou adjectivaux, à l'exception bien sûr des dérivés dont les radicaux sont des participes présents. Ils sont totalement absents des descriptions antérieures de la préfixation en *anti-*. Par exemple, Corbin (1980:194) considère l'existence de ces dérivés comme impossible :

> « Les verbes et adverbes en *anti-* sont issus, non d'une préfixation de *anti-*, mais d'une suffixation et/ou d'un changement de catégorie d'une base déjà préfixée par *anti-*. »

Les seuls dérivés qu'elle envisage sont ceux dont les bases sont des dérivés en *anti-* comme *antiparasiter*, *antiseptiser*, etc.

On peut d'autre part observer que les dérivés déverbaux de notre corpus tendent à être construits préférentiellement sur des verbes à agent (i.e. dont le sujet remplit le rôle d'agent comme *antimanger*, *antimordre*, *antirire*, etc.) plutôt que sur des verbes à patient (i.e. dont le sujet recevrait le rôle de patient comme *antiglisser*, *antifondre*, etc.). L'exemple de *antidormir* (12) est à ce titre exceptionnel. Cette contrainte apparaît également dans la sélection des dérivés dont le radical est un participe : si les participes présents sont nombreux (*antidérapant*, *antipolluant*, etc.), les participes passés sont rares. Je n'ai ainsi pas trouvé d'*antidérapé*, d'*antipollué*, etc. et un exemple comme *antivibré* (13) me semble être le participe passé d'un verbe *antivibrer* qui signifierait 'rendre antivibrant'.

(13) *Nos modèles à double lame sont bien équilibrés, efficaces et bien* ***anti-vibrés****.* `www.husqvarna.com/fr/landscape-and-groundcare/products/hedge-trimmers/325hd60x`

4.2.2 Radicaux pluriels

L'initiale *anti-* peut aussi apparaître devant des radicaux moins prototypiques que ceux des exemples (2 ; 3), notamment des formes nominales au pluriel, comme *antirides*, *antilimaces* (14) ou *anti-animaux* (15). Ces radicaux sont remarquables car les marques de pluriel n'apparaissent normalement qu'en discours, lorsque les mots sont utilisés dans des constructions syntaxiques. Notons que les dérivés en (14 ; 15) sont au masculin singulier. Leurs finales ne peuvent donc pas être des marques de flexion contextuelle (Booij 1996).

(14) *L'autopsie a décelé du méthiocarbe, un* ***antilimaces****, en trop forte quantité pour que cela soit accidentel.* `www.ornithomedia.com/phorum/read.php?6,71681`

(15) *Un lance pierre ou alors tu lui balances un répulsif* ***anti animaux*** *quand tu la vois.* `fr.answers.yahoo.com/question/index?qid=20090423220818AAhf6WS`

La présence de marques du pluriel sur les radicaux des mots en *anti-* est massive. Par exemple, il existe selon Google™ cinq fois plus de formes *anti*(-)*rides* que de formes *anti*(-)*ride* et trois fois plus d'*anti*(-)*animaux* que d'*anti*(-)*animal* (voir table 4.1).

Les marques de pluriel sur les radicaux dans des exemples comme (14 ; 15) ne permettent pas d'analyser ces séquences comme des groupes prépositionnels. Dans une grande majorité de cas, ces marques sont uniquement graphémiques. Elles n'apparaissent qu'à l'écrit et ne donnent pas lieu à une réalisation phonique, y compris dans des contextes de liaison comme dans *un antirides épatant* que personne ne prononce /œ̃ nɑ̃tiʁid zepatɑ̃/.

	singulier	pluriel
antiride	97 000	520 000
antianimal	470	1 600

TABLE 4.1 – Comparaison des fréquences des radicaux au singulier et au pluriel. Les chiffres sont les nombres de pages Web en français indexées par Google contenant ces mots, avec ou sans séparation entre le préfixe et le radical. Décompte réalisé le 10 octobre 2009.

Ces marques peuvent cependant apparaître quand la flexion est contextuelle, en l'occurrence dans un contexte pluriel. Par exemple, *des antirides épatants* peut être prononcé /de zɑ̃tiʁid zepatɑ̃/. Les locuteurs ont ainsi la possibilité d'utiliser comme radical la forme de liaison du pluriel obtenue en ajoutant [z] à la forme libre (Plénat 2008). On rencontre le même type de flexion contextuelle dans un syntagme comme *la lutte anti-sous-marine* (96). L'existence de radicaux plus clairement pluriels comme *anti-animaux* n'est pas davantage un argument en faveur d'une analyse prépositionnelle de ces séquences. La sélection de radicaux au pluriel se retrouve par exemple aussi dans des composés VN comme *tue-flics* (16) ou *garde-chevaux* (16). Le recours à une forme différente de celle qui est utilisée par défaut pour les dérivés préfixés en *anti-*, en l'occurrence, la forme de citation de la base, ne pose aucun problème dans le cadre du modèle théorique décrit en section 2.3. Les locuteurs ont la possibilité de sélectionner comme radical n'importe quelle forme capable d'évoquer la base du dérivé. Seule compte en effet l'optimalité du dérivé.

(16) *Les narcotrafiquants mexicains affectionnent ces pistolets « **tue-flics** » - comme on les appelle au Mexique - de luxe.* `www.lemonde.fr/web/recherche.../1,13-0,37-1058029,0.html`

(17) *La toiture du clocher et de l'église vole en éclats tombant sur les chevaux haut-le-pied cachés non loin de là et tuant un **garde-chevaux**.* `www.genemilassoc.fr/charge.php?f=MTIyOA==`

La sélection de ces radicaux est probablement une stratégie utilisée par les locuteurs pour marquer l'interprétation générique des bases de ces dérivés (Montermini 2002, 2008) : de même qu'il n'y a pas de différences entre (18a) et (18b), il n'y en a pas entre *antilimace* et *antilimaces* ou entre *antiride* et *antirides*. En d'autres termes, le pluriel en (14 ; 15) n'est pas une marque de nombre mais un indice facilitant l'interprétation générique. Cette analyse rejoint celle de Heyna (2008) qui considère que ces marques forcent l'interprétation du nom base comme un nom de classe (N_{Classe}).

(18) a. *Le lion est un mammifère*

b. *Les lions sont des mammifères*

Un autre argument contre l'analyse de *antirides* ou *anti-animaux* comme des groupes prépositionnels est l'absence de déterminant. Or, sans déterminant, ces radicaux ne peuvent pas fonctionner comme des syntagmes nominaux classiques. L'importance du déterminant

apparaît par exemple dans la paraphrase (19a) de *un répulsif anti-animaux*. Ce même syntagme ne peut pas être paraphrasé par (19b).

(19) a. *... un répulsif contre les animaux ...*

b. *... un répulsif contre animaux ...*

4.2.3 *Anti-*VN

Le préfixe *anti-* se rencontre aussi devant des composés VN comme *anti-porte-bagage* (20) ou *anti-tourne-tête* (21). Les composés étant des lexèmes comme les autres, on a simplement affaire à des dérivés légèrement moins prototypiques que ceux dont le radical est monolexical (i.e. une forme simple). Les VN peuvent également comporter des déterminants comme dans *tire-au-flan*, *crève-la-faim* ou *tue-l'amour*. Les radicaux des dérivés en (22 ; 23) relèvent de ce type.

(20) *Encore un argument dans le camp des* ***anti-porte-bagage****. Les modèles sans porte-bagage ont une roulette sur le garde-boue.* bromptonforum.conceptforum.net/trucs-et-astuces-f2/b-sans-garde-boue-ca-roule-plie-t1688.htm

(21) *mais par contre, ma tête fait le tour du monde et c'est ça qui me tourneboule le ventre après... bref, au moins, elle m'a donné de "l'****anti-tourne-tête****" et là, je recommence à marcher droit...* forum.aufeminin.com/world/communaute/forum/forum2__forum=fitness1&m=52117&whichpage=3.html

(22) *Et puis, celles qui ont essayé les vernis "****anti-ronge-tes-ongles****" pourront te dire que ça ne marche pas forcément !!* www.infobebes.com/htm/bebe/sevrage-impossible,d-169258.aspx

(23) *D*Lires: La cuillère* ***anti-brûle-la-langue*** *On en a tous fait la cruelle expérience. On réchauffe un bol au micro-ondes, il est tiède mais le liquide est brulant. Du coup, on se fait griller la langue.* cetnia.blogs.com/d_lires/2006/02/la_cuillre_anti.html

Ces exemples ne peuvent pas être considérés comme des syntagmes prépositionnels. Les VN peuvent notamment servir de radicaux de dérivés suffixés comme dans *tourne-disqueur* (24) 'animateur de discothèque', *tire-au-flantisme* ou *casse-têtesque*.

(24) *Dentiste pour chiens morts,* ***tourne-disqueuse*** *incontournable de la scène kitch.* www.la-machine-a-machins.com/machins.htm

Le fait que les radicaux de certains de ces VN (*tourne-tête* ; *ronge-tes-ongles* ; *brûle-la-langue*) ne soient pas lexicalisés ne peut pas non plus être invoqué comme argument en faveur d'une analyse prépositionnelle. On trouve en effet des configurations comparables dans des dérivés suffixés comme en (25). Le fait que ces radicaux soient créés spécialement pour ces dérivés, ne réduit en rien le statut morphologique de ces derniers. Leur seule différence avec les radicaux lexicalisés comme *porte-bagage* est le niveau de complexité plus élevé de leur construction qui comprend à la fois une préfixation et une composition. Une autre caractéristique de ces dérivés est qu'ils ne sont normalement pas destinés à entrer dans le lexique. Il existe cependant des exceptions comme *anti-monte-lait* (26).

(25) *je vois que je mettais trompé, et je vois surtout que Tsunia est une adepte du* ***gros-bourin-qui-détruit-tout-en-un-claquement-de-doigtisme****,* forumold.wakfu.com/topic.php?lang=fr&id=10115...

(26) *L'anti-monte-lait est un disque lourd, qui tombe au fond, là où se forment les bulles.* fr.wikipedia.org/wiki/Anti-monte-lait

4.2.4 *Anti-* devant des syntagmes nominaux sans déterminant

Le préfixe *anti-* peut aussi apparaître devant des syntagmes nominaux sans déterminants. Comme pour les VN, ces radicaux peuvent soit être lexicalisés comme dans *anti-maison de retraite* (27) ou dans (28 ; 29), soit ne pas l'être comme dans (30 ; 31) et dans le premier exemple en (32). Les dérivés dont les radicaux ne sont pas lexicalisés sont similaires à (25). Ils ne sont normalement pas destinés à entrer dans le lexique. Ces exemples sont, de ce point de vue, proches des dérivés de la section précédente dont les radicaux sont des composés morphologiques créés « à la demande » (21 ; 22 ; 23). Ils suggèrent que les locuteurs peuvent utiliser comme radicaux de constructions morphologiques des séquences construites par des moyens syntaxiques. Morphologie et syntaxe opèrent ainsi au même niveau, celui de la production et de la compréhension des énoncés : la morphologie permet d'exprimer du sens au moyen de mots et de comprendre le sens des mots ; la syntaxe permet de le faire au moyen de structures syntaxiques et de comprendre le sens de ces dernières.

(27) *C'est une utopie réaliste qui commence à s'essaimer un peu partout en France et c'est surtout une* ***anti-maison de retraite****.* www.millebabords.org/spip.php?article7615

(28) *anti-rond-de-cuir* ; *anti-jeune premier*

(29) *Le Québec et la Colombie-Britannique sont les seules provinces qui ont des dispositions* ***anti-briseurs de grève****, communément appelées « lois anti-scabs ». ...* tempsreel.nouvelobs.com/actualites/20080707.OBS1839/lsreactions00e5.html?l=7

(30) *Les* ***anti-produits-laitiers-à-l'âge-adulte*** *estime qu'en dehors de la croissance (bébé-enfant-ado), le lait n'est plus nécessaire, et peut provoquer (sans aller jusqu'au cancer) d'importants troubles digestifs.* fr.answers.yahoo.com/question/index?qid=20070917060124AAGXHru

(31) *J'ai bien compris que tu étais* ***anti-fil à pendre le linge*** *alors hop, enlevé!* www.vacanceo.com/albums_photos/voir-photo_396772.php

(32) *celui-ci est l'****anti "homme qui me sied bien"****, l'****anti "celui que je peux croiser dans un train et y penser pendant toute une journée"****.* mariealix.canalblog.com/.../14232858.html

4.2.5 *Anti-* devant des noms propres, des propositions, des interjections et des onomatopées

Pour des raisons sémantiques (voir section 4.3), les noms propres sont particulièrement fréquents après *anti-* : *anti-Bush*, *anti-Aznavour*, *anti-Besson* (33), *anti-Boeing*, *anti-Crédit*

Lyonnais, *anti-RPR*, *anti-FMI* (34), etc. La nature dérivationnelle de ces constructions ne fait aucun doute, même si les noms propres sont référentiellement autonomes et peuvent être des compléments dans des groupes prépositionnels. Les locuteurs n'ont par exemple aucun mal à les utiliser comme radicaux de dérivés suffixés (35). Ces radicaux peuvent être rapprochés des propositions utilisées de façon autonymique comme dans (36) ou dans le deuxième exemple de (11). Notons que l'on retrouve le même type de radical dans des dérivés suffixés comme (37).

(33) *Je vois déjà les* ***anti-Besson*** *brandir les pancartes et crier aux scandales !* www.uneporte.net/ONG-BAK,-l-edition-DVD-zone-1-de-Christal-Films_a813.html

(34) *La police disperse une manifestation* ***anti-FMI*** *à Istanbul.* www.latribune.fr/.../la-police-disperse-une-manifestation-anti-fmi-a-istanbul.html

(35) *moliéresque*, *chiraquien*, *gaullisme*, *jospiniste*, *juppette*

(36) *Je demande la création d'une brigade* ***anti "je veux transformer mon os en voiture de course sans payer un centime et sans rien y connaitre"*** www.forum-auto.com/pole.../sujet380039.htm

(37) *Au vieux réflexe stalino-****je-ne-veux-voir-qu'une-tetesque*** *répond une bordée de ...* cozop.com/frednetick/pim_paf_pouf_chocs_de_neurones_au_ps/proches

Les dérivés en *anti-* formés sur des onomatopées ou des interjections comme *anti-aïe aïe* (38), *anti-plouf* (39), ou *anti-snif* (40) sont habituellement ignorés par les études de ce préfixe. Ils illustrent l'absence de contraintes sur la catégorie des lexèmes dont les formes sont utilisées comme radicaux. Je signale que des radicaux similaires apparaissent dans des dérivés suffixés comme *vlantesque*, *pif-pafesque* (41), *clonguiste* ou *ploufiste* (42).

(38) *il ne faut pas resté trop longtemps sur le trône et évité de mangé trop pimenté, prend un bon bain chaud et mais une crème* ***anti aie aie****.* fr.answers.yahoo.com/question/index?qid=20070830075705AAKQZIC

(39) *ca c'est parce que tu connais pas le pq* ***anti-plouf****!!!mdr...quoique pour la courante ca marche pas!!!* forum.aufeminin.com/world/communaute/forum/forum2__forum=f382&m=59783&whichpage=6.html

(40) *Je suis* ***anti snif****, car c'est con de pleurer alors qu'on peut sourire.* smile.exprimetoi.net/fun-time-f13/je-suis-un-anti-t109-255.htm

(41) *J'ai même failli être perdu sur certaines portions trop "****pif-pafesque****".* www.melivelo.fr/forum/viewtopic.php?t=3000...

(42) *Est-il un* ***clonguiste*** *mou ou un* ***ploufiste*** *(clonguiste éclaboussé) ?* fr.groups.yahoo.com/group/Fulmedico2000/message/54607

4.2.6 *Anti-* devant des pronoms

L'analyse des dérivés dont les radicaux sont des pronoms comme *antirien* (43), *anti-ceci ou cela* (44), *antitout* (45) est *a priori* plus problématique. Les pronoms ont un sens procédural. Ils ne permettent pas de faire référence à des catégories et peuvent apparaître

directement après des prépositions (46). Ces caractéristiques ne permettent cependant pas de conclure que ces séquences sont des groupes prépositionnels. Rien n'impose en effet au radical d'un dérivé d'évoquer une catégorie ni d'être totalement spécifié. Cette liberté s'observe aussi dans des dérivés suffixés comme *rienisme* (47) ou *moiesque* (48) dont le caractère affixal ne peut pas être mis en doute. Les radicaux pronominaux des dérivés préfixés en *anti-* peuvent également comporter des propositions comme (49) ou dans le deuxième exemple de (32).

(43) *Et change vite d'antivirus, Avast est un* ***anti-rien*** *!* `fr.answers.yahoo.com/question/index?qid...`

(44) *Et puis être* ***anti-ceci ou cela****, c'est avoir des idées arrêtées et c'est reservé aux nazes qui ont besoin de se coller une étiquette.* `fr.answers.yahoo.com/question/index?qid...`

(45) ***Anti-tout****, anti-ghettos, anti-lobbies, anti-chapelles, pour résumer libertaire et anticons, probablement* ***anti-lui-même****, comme il l'écrit.* `www.lepost.fr/article/2008/08/27/1252204_apres-charlie-hebdo-sine-hebdo.html`

(46) *Ils ne sont* ***contre personne****, mais contre les erreurs qui ont conduit la FRANCE à la défaite* `www.cg47.fr/archives/recherche/.../pdf%5C1W141-001.pdf`

(47) *Toutefois, le refus d'entrer dans l'excès, c'est à dire dans le* ***rienisme****, conduit à s'interroger sans cesse sur la validité du processus* `fr.wikipedia.org/.../Société_anonyme_de_production_littérair`

(48) *Le roman est extrêmement descriptif et sa qualité principale se situe, d'un point de vue purement* ***moiesque*** *(le mien, donc), dans les comparaisons.* `effetlarsen.site.voila.fr/salammbo.htm`

(49) *Il y aurait donc du "racisme langagier", par exemple ; ou du racisme "****anti-ceux-qui-portent-des-casquettes****".* `dogmatique.kanak.fr/.../racisme-et-integration-t52.htm`

4.2.7 *Anti-* devant des syntagmes nominaux comportant un déterminant

Le dernier groupe de constructions comportant l'initiale *anti-* de cet inventaire de radicaux est celui des syntagmes nominaux introduits par un déterminant. Leur analyse est plus problématique que celle des précédentes. Plusieurs cas de figures peuvent être distingués. Le premier est celui où le déterminant fait partie d'une expression ou d'un titre comme *anti-la vie* (50) ou *anti-les ch'tis* (51). Dans le premier cas, le radical est une partie de l'expression usuelle *la vie quoi !*. Dans le second, il s'agit du titre du film *Les Ch'tis*.

(50) *je lis aussi à gauche à droite des posts anti enfants, anti mecs, anti animaux ,* ***anti la vie*** *quoi.* `jesuisunique.blogs.com/.../la-clibataire.html`

(51) *voilà* ***les anti-les ch'ti****!voilà la france mesquine,radine,paranoîaque et raciste!* `www.allocine.fr/video/player_gen_cmedia=18824509&cfilm=129327.htm`

Le deuxième cas est celui où le déterminant marque l'interprétation générique. Il joue alors le même rôle que les radicaux pluriels présentés en section 4.2.2. Ce type d'emploi peut être illustré par des dérivés comme *anti-les rêves de gonzesses* ou *anti-la poésie* (52)[4].

(52) ***Antie les demies mesures, anti les happy end, anti les reves de gonzesse, anti la poesie,*** *je vous offre des cliches.* `www.myspace.com/escone`

Le troisième cas est celui où le déterminant marque la nominalisation du radical comme dans *anti-les anti-corrida* (53), *anti-corrida* étant par défaut un adjectif.

(53) *Mais je suis plutot* ***Anti les Anti-corrida****. Vous m'horripilez.* `forum.doctissimo.fr/../manifestation-corrida-carcassonne-sujet_8368_1.htm`

Le dernier cas est celui où le déterminant sert de quantificateur comme en (54 ; 55 ; 56 ; 57) ou fait partie d'un quantificateur comme en (58).

(54) *Comme la plupart des sites* ***anti des chanteurs et chanteuses****, on y retrouve des chansons truquées, des photos modifiées ...* `echo.levillage.org/169/1087.cbb`

(55) *Que veut tu beaucoup de gens se définissent malheureusement plus* ***anti une équipe*** *que pro une autre.* `www.rmc.fr/.../diouf-dement-pour-gudjohnsen/?...`

(56) *Nous n'avons absolument pas une posture anti-Ordre (et tu le sais si tu nous lis) mais foncièrement* ***anti cet Ordre-là****.* `www.physiorama.com/forum/?k=aucunes`

(57) *mais trouver son nom associé à une diatribe* ***anti-ces-méchantes-molécules-qui-nous-font-du-mal*** *est au mieux ironique, plus vraisemblablement désespérant,* `www.irenedelse.com/.../du-bon-et-du-mauvais-usage-de-la-chimie/`

(58) *Etre anti-Bush de la manière dont vous l'exprimez c'est aujourd'hui être* ***anti-une majorité d'américain****.* `www.les4verites.com/.../comments.php?...id..`

L'analyse de ces séquences est rendue difficile par leur structure, identique à celle des groupes prépositionnels. Elle l'est aussi par le fait qu'il n'existe pas de moyen simple permettant de repérer des dérivés suffixés dont les radicaux sont introduits par un déterminant. Il est plus difficile d'exhiber des dérivés qui montrent clairement que ces syntagmes peuvent servir de radicaux. Notons également qu'il est difficile de trouver des syntagmes nominaux de ce type comme radicaux d'autres préfixes comme *post-*, *pré-*, *intra-*, *infra-*, etc., si l'on excepte *pro-* (59). La différence entre *anti-* et *pro-* d'un côté et ces autres préfixes est le niveau de langue des discours dans lesquels ils sont utilisés. Ce dernier peut être très bas pour les deux premiers, permettant ainsi une assez grande souplesse langagière.

(59) *Pour ma part, je suis* ***pro les principes d'Hadopi****.* `fallengalaxy.com/forum/viewtopic.php?id...`

On peut néanmoins avancer trois arguments en faveur d'une analyse dérivationnelle des séquences comme (54 ; 55 ; 56 ; 57 ; 58). Le premier est le fait que ces séquences sont très rares et relativement marquées. Si *anti* était une préposition, on devrait pouvoir trouver un nombre très important de configurations de ce type, or ce n'est pas le cas. Ces

4. Les exemples sont reproduits en respectant les graphies originales et notamment les fautes de frappe comme *Antie* dans cet exemple, d'orthographe, etc.

séquences sont par exemple totalement absentes dans la base Frantext catégorisée. Cette dernière contient pourtant 21 108 occurrences de mots préfixés en *anti-* (avec ou sans séparation entre le préfixe et le radical). En comparaison, on trouve dans cette base 41 487 occurrences de la préposition *contre* suivie d'un déterminant pour 7 256 formes de substantifs précédées par *contre-* (avec un trait d'union ou sans séparation entre le préfixe et le radical).

Le deuxième argument est que dans les exemples ci-dessus, *anti-* a le même sens que dans les dérivés présentés dans les sections précédentes. Or s'il existait une préposition *anti*, on pourrait s'attendre à ce qu'elle ait un sens qui lui soit propre, différent de celui du préfixe *anti-*. C'est ce qu'on observe par exemple avec le préfixe *contre-* et la préposition homonyme. Dal (2003b:189) montre notamment qu'avec les noms déverbaux ou assimilés, la préposition *contre* « active préférentiellement sa valeur adversative » et fonctionne comme un inverseur (De Mulder et Amiot 2003), tandis que le préfixe *contre-* « dit du référent du dérivé qu'il construit qu'il constitue une réplique du référent de la base à laquelle il s'applique ». Cette analyse rejoint le présupposé de secondarité posé par De Mulder et Amiot (2003:214).

L'identité sémantique qui vient d'être évoquée fournit un troisième argument contre l'analyse prépositionnelle de *anti*. Il existe en effet en français une préposition sémantiquement très proche du préfixe *anti-* (voir section 4.3.6), que l'on peut utiliser notamment pour paraphraser ses dérivés. Il s'agit naturellement de *contre* (Zribi-Hertz 1973, Dal 2003b). Une éventuelle préposition *anti* n'aurait de ce fait que peu d'utilité et l'on peut envisager que dans les exemples (54 ; 55 ; 56 ; 57 ; 58), l'initiale *anti* est utilisée comme une variante de la préposition *contre* qui renforce l'interprétation de ces syntagmes comme des nuisances (voir section 4.3.5). *Anti-* est aussi plus approprié pour certaines têtes. Par exemple, dans (54), le nom *site* facilite l'utilisation de *anti-* par rapport à celle de *contre*. On trouve ainsi 250 000 pages en français indexées par Google qui contiennent la séquence *site*(*s*) *anti* alors qu'il n'en existe que 75 000 qui contiennent *site*(*s*) *contre*. Notons que le premier chiffre ne prend pas en compte les dérivés dans lesquels le préfixe n'est pas séparé du radical.

Cette description confirme la classification par Amiot (2004), de *anti-* comme un préfixe de type 2, situé entre les préfixes les plus prototypiques, de type 1, comme *re-* et *dé-* et ceux de type 3 qui peuvent être utilisés à la fois comme préfixe et préposition. Le préfixe *anti-* est sémantiquement proche de la préposition *contre*, mais il n'est pas en revanche un concurrent du préfixe *contre-* qui ne permet pas de construire des adjectifs dénominaux (Rey 1968, Amiot 2004). Je signale que Fradin (1997b) réfute lui aussi la possibilité que *anti-* puisse être une préposition. Son argument principal est que les vraies prépositions se préfixent facilement à des verbes. Il affirme que cette possibilité est totalement exclue pour *anti-*. Plusieurs contre-exemples dans notre corpus (11 ; 12 ; 78 ; 79) infirment cette assertion.

4.3 La sémantique des dérivés en *anti-*

La seconde grande question abordée dans ce chapitre est l'organisation des sens des dérivés en *anti-* et les conclusions qu'il est possible d'en tirer concernant l'unicité de ce

préfixe. Les études antérieures de la préfixation en *anti-* identifient trois types d'interprétations pour ses dérivés (Rey 1968, Zribi-Hertz 1973, Durand 1982, Corbin 1987, Fradin 1996, 1997a, b) :

1. un sens locatif que l'on trouve dans des dérivés comme *antichambre*, et qui peut être paraphrasé par « qui est situé devant X » où X est la base ;
2. un sens antagoniste que l'on trouve dans *antiparlementaire* ou *antilimace*, et qui peut être paraphrasé par « qui s'oppose à X » ;
3. un sens antonymique que l'on trouve dans *antihéros*, et qui peut être paraphrasé par « qui est le contraire de X ». Les dérivés des domaines de la physique des particules et des mathématiques relèvent de ce sens (Zribi-Hertz 1973).

L'approche extensive permet de raffiner et d'étendre cette description grâce à la découverte d'exemples jusqu'ici inconnus ou ignorés. Cette extension a été présentée dans une forme préliminaire dans Plénat et al. (2004a). J'en présente ici une version plus complète qui organise les sens des dérivés en *anti-* selon deux critères qui définissent cinq configurations. Le premier critère distingue les dérivés endocentriques de ceux qui sont exocentriques. Il est utilisé notamment par Zribi-Hertz (1973). Les dérivés endocentriques désignent des notions (objets, qualités ou événements) de même nature que leurs bases. Ils correspondent aux altérations au sens de Roché (en préparation). Les exocentriques désignent des notions de nature différente de celle de leurs bases. Ils correspondent aux transferts au sens de Roché (en préparation). Le second critère distingue trois modes d'interprétation : spatial, logique et adversatif. Dans le premier, *anti-* sert à décrire des relations spatiales, dans le deuxième, des catégories et dans le dernier, des protections contre des nuisances. Ces trois modes reprennent simplement les trois types d'interprétation préalablement identifiés. La principale contribution de la présente étude est d'établir que ces deux critères sont indépendants, que leur croisement définit bien six possibilités et que l'on peut exhiber des exemplaires pour cinq d'entre d'elles. Dans les études antérieures, les deux critères sont considérés comme corrélés, comme l'illustrent les exemples présentés dans le tableau 4.2. Par ailleurs, la description proposée pour les interprétations adversatives est plus fine que celle que l'on trouve dans les études précédentes. Je n'ai pas réussi à trouver d'exemples clairs de dérivés dont l'interprétation est logique exocentrique. Cette interprétation est probablement impossible. Cette incompatibilité pourrait avoir pour origine le fait que, dans ses emplois logiques, *anti-* est avant tout un inverseur et qu'il induit donc obligatoirement une interprétation endocentrique.

Les interprétations endocentriques se caractérisent par l'identité des catégories du dérivé et sa base (60) [5] et par celle de leurs genres quand la base est nominale. En conséquence, le dérivé désigne un hyponyme de la base comme le notent Rey (1968), Durand (1982) et Heyna (2008) à propos des dérivés de la sous-série *antihéros*.

(60) a. *chauffeur* > *chauffard*
b. *réforme* > *réformette*
c. *pendre* > *pendouiller*

5. Ces exemples sont empruntés à Roché (en préparation).

	endocentrique	exocentrique
spatial	antichambre	?
logique	antihéros	?
adversatif	?	antiparlementaire

TABLE 4.2 – Le sens des dérivés en *anti-* est décrit au moyen de deux critères indépendants. Dans les études précédentes, ces critères étaient considérés comme corrélés, comme le fait apparaître la distribution des dérivés dans le tableau.

d. *célèbre* > *célébrissime*

L'alternance entre les interprétations endocentriques et exocentriques est en fait partagée par de nombreuses préfixations comme *pré-* ou *sous-* (61) (voir la section 2.2 de Aurnague et Plénat (2008) pour une discussion plus détaillée de cette question). Par exemple, un sous-préfet est un préfet de rang inférieur (dérivé endocentrique). À l'inverse, un sous-main est une entité qui n'est pas une main (dérivé exocentrique). Le genre du dérivé est dans ce cas différent de celui de la base. De même, une pré-révolution est une répétition générale d'une révolution à venir (dérivé endocentrique). La pré-Révolution décrit une période qui précède la Révolution. Ce n'est pas une révolution et le dérivé est par conséquent exocentrique.

(61) a. *un sous-préfet* ; *un sous-main*

b. *une pré-révolution* ; *la pré-Révolution*

Le reste de la section est organisé de la manière suivante. Je présente en 4.3.1 quelques notions générales utilisées pour décrire la sémantique de *anti-*. Les cinq sous-sections suivantes sont consacrées aux cinq types d'emplois présents dans notre corpus. Dans la sous-section 4.3.7, je présente différents arguments en faveur de l'unicité de la préfixation en *anti-*.

4.3.1 Notions générales

À la suite de Heyna (2008), j'adopte dans cette étude les notions de type, d'attracteur et de satellite pour décrire l'interprétation des bases des dérivés en *anti-*. J'utilise également la catégorie de la dynamique des forces pour affiner l'analyse des dérivés adversatifs comme *antilimace* ou *antiparlementaire*.

Types

Selon Berrendonner (1995, 2002), les types sont des noms d'objets intensionnels ou génériques. Leur définition repose sur l'hypothèse que les univers d'interprétation construits par les discours sont composés de deux sous-ensembles d'entités : un sous-ensemble R d'objets extensionnels et un sous-ensemble I d'objets intensionnels ou types. Les éléments

de R sont des entités « pensées comme existant *in re* », c'est-à-dire dans la réalité. Les éléments de I sont « des êtres conçus comme subsistant *in intellectu* », i.e. dans l'esprit des locuteurs. Les types sont des individus comme les autres, aptes à recevoir des prédicats, et à servir de cibles aux opérations de référence, et tout spécialement aux références génériques ; ce ne sont pas des fonctions applicables sur les choses. Ils servent par ailleurs d'instrument de catégorisation des objets extensionnels. La notion de type est impliquée dans toutes les dérivations morphologiques dans lesquelles la base peut être exprimée par un nom. Par exemple, dans *terrain piscinable*, la base évoquée par le radical *piscine* n'est pas utilisée pour faire référence à une piscine particulière, ni à un quelconque élément de la classe des piscines, mais à un type, à une piscine abstraite, générique.

Attracteurs et satellites

Les types peuvent aussi donner lieu à une interprétation métonymique et permettre ainsi de viser un autre objet. Cet objet doit entretenir avec le type une relation de contiguïté référentielle suffisamment typique (Berrendonner 1990). L'interprétation métonymique exploite la procédure d'implication discursive qui consiste à placer un objet de connaissance x dans le savoir commun au locuteur et à son interlocuteur puis à conclure à l'existence d'un autre objet y en mettant à profit l'existence d'un lieu commun ou d'une relation de cause à effet reliant x à y ; x doit être un objet de connaissance fortement implanté, servant de support à des engagements axiologiques intenses et controversés. Les arguments et les assertions tenues à son propos, de façon répétitive, acquièrent un statut de lieux communs. Le type x condense ainsi l'ensemble des actes d'énonciation typiques ou symptomatiques tenus à son propos (Berrendonner 1990).

C'est ce qui se passe par exemple pour un nom propre comme *Sarkozy* qui est un objet de discussion fortement implanté dans les connaissances partagées des locuteurs et auquel sont associés des lieux communs nombreux. Les objets y dont l'existence est impliquée par x = Sarkozy sont dans ce cas : les idées de Sarkozy ; la politique de Sarkozy ; le discours de Sarkozy ; etc. Il en va de même pour les virus informatiques (x), eux aussi bien connus, et faisant l'objet d'un grand nombre de lieux communs relatifs à des y comme les symptômes qui les caractérisent, les dysfonctionnements qu'ils causent, les pertes qu'ils induisent, les façons dont on peut s'en prémunir, etc.

À la suite de Heyna (2008), j'adopte ici la terminologie de Berrendonner qui propose d'appeler *attracteurs* les entités x et *satellites* les y correspondants. La préfixation en *anti-* exploite les relations qui s'établissent entre les attracteurs et les satellites pour les dérivés du type *antiparlementaire* et *antilimace*.

La dynamique des forces

La catégorie de la dynamique des forces proposée par Talmy (2000) est un niveau de description adapté à la sémantique des dérivés adversatifs en *anti-*. Dans ce cadre, les entités sont appréhendées à travers les forces qu'elles exercent les unes sur les autres. Par exemple, le sens d'un énoncé comme (62) peut être décrit comme une interaction entre deux entités, l'eau et le barrage, exerçant des forces contraires : l'eau exerce une force qui tend à la faire

s'écouler vers l'aval tandis que le barrage exerce une force qui tend à sa propre immobilité et par suite à celle de l'eau. Les deux forces sont ainsi opposées. De plus, la force exercée par le barrage l'emporte sur celle de l'eau ce qui a pour conséquence l'immobilité de l'eau. En d'autres termes, le barrage manifeste sa tendance à l'immobilité aux dépens de l'eau. Dans la terminologie de Talmy, l'élément qui correspond à la cible [6] est appelé « agoniste » (l'eau dans l'exemple (62)) et celui qui correspond au site, « antagoniste » (le barrage dans l'exemple (62)).

(62) *Le barrage retient l'eau.*

Talmy utilise des représentations schématiques pour décrire les forces et les rapports qui s'établissent entre elles. Ces schémas permettent de représenter :

1. la force agoniste (cercle),
2. la force antagoniste (figure concave),
3. le fait qu'une force tend au déplacement (>) ou à l'immobilité (•),
4. le fait qu'une force soit supérieure (+) ou inférieure à l'autre, et
5. le fait que la résultante (ligne inférieure) soit le déplacement de la cible (>) ou son immobilité (•).

La figure 4.1 présente la configuration qui correspond à la phrase (62).

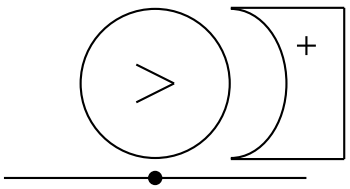


FIGURE 4.1 – Représentation schématique qui décrit les forces impliquées dans l'interprétation de la phrase *Le barrage retient l'eau*. La force de l'eau est représentée par le cercle et celle du barrage par la figure concave. L'eau exerce une force qui tend à son déplacement (>) et le barrage une force qui tend à son immobilité. La force exercée par le barrage est supérieure à celle (+) de l'eau. Le barrage manifeste sa tendance à l'immobilité aux dépens de l'eau (• sur la ligne inférieure).

L'intérêt majeur du cadre de la dynamique des forces est qu'il impose l'identification précise des participants et des interactions qui existent entre eux. Dans le cas d'un dérivé adversatif comme *antilimace*, il met en évidence les limaces, l'antilimace et, au second plan, un agent qui s'en sert. Les limaces exercent une force qui tend à la consommation des plantes. Cette force nuit aux intérêts d'un agent intentionnel. Pour s'en protéger, ce dernier utilise un antilimace qui exerce une force opposée à celle des limaces. La limace exerce ainsi une

6. Les termes de *cible* et de *site* (Vandeloise 1986) correspondent à ceux de *figure* et de *ground* (Talmy 2000).

force qui est une nuisance et l'antilimace une force de protection des intérêts de l'agent[7]. Le mécanisme mis en place par l'agent intentionnel peut aboutir à la destruction de l'entité agoniste, si l'antilimace est par exemple un poison, ou prendre d'autres formes comme des ondes, un répulsif ou d'autres dispositifs qui éloigneraient les limaces ou les paralyseraient momentanément. Dans ce dernier cas, on reste très proche du scénario des forces puisque l'entité agoniste n'est pas détruite. Seule la nuisance, la force qu'elle exerce, est neutralisée. De manière similaire, un plateau antivibrant permet de prémunir les objets qui se trouvent sur ce dernier des vibrations, considérées comme des nuisances. Les vibrations exercent ainsi une force contraire aux intérêts d'un agent qui n'est pas nommé et un plateau antivibrant une force antagoniste qui préserve ces intérêts. Dans le cas de *anti-Sarkozy*, le dérivé a pour référence l'agent. Il exerce lui-même une force destinée à le protéger de ce qu'il considère comme des nuisances que Nicolas Sarkozy cause à ses intérêts.

4.3.2 Interprétations spatiales endocentriques

L'interprétation spatiale endocentrique concerne les dérivés à sens locatifs relativement anciens comme en (63), signalés par Rey (1968).

(63) *antibust*[8] ; *antichambre* ; *antichapelle* ; *anticour* ; *anticuisine* ; *antiposte* ; *antisalle*

Ces dérivés sont tous endocentriques. Ils décrivent des lieux qui se trouvent avant, devant ou qui donnent accès à un autre lieu, désigné par la base. Ces emplois d'*anti-* sont généralement analysés comme des variantes du préfixe *anté-* qui forme *antécambrien*, *antéposer*, *antépénultième* ou *antédiluvien*. C'est notamment le cas dans l'article du TLF consacré à *anté-*. Il existe par ailleurs un dérivé à sens temporel, *antidater*, dans lequel *anti-* est clairement une variante de *anté-* (voir section 2.10.1). Étant le seul à avoir ce sens, il n'est pas utile de l'intégrer au reste des dérivés préfixés en *anti-*.

La série de dérivés ci-dessus peut être complétée par les noms de montagne (64), le nom de ville *Antibari*[9] et des termes d'anatomie comme (65) ou comme *antitrochanter*. Tous ces dérivés désignent des entités qui se trouvent en face de celles qui sont désignées par leurs bases. Par exemple, l'Anti-Liban est une montagne qui fait face au Mont Liban. Le trochanter et l'anti-trochanter sont des faces articulaires en contact. Elles se font face, sont opposées du point de vue de leur orientation et exercent l'une sur l'autre des forces opposées.

(64) *Anti-Atlas*, *Anti-Liban*, *Anti-Taurus*

(65) *antitragus* ; *antilobe* ; *antithénar* ; *antiprostate*

(66) *L'articulation fémoro-pelvienne comprend l'acétabulum sur la ceinture pelvienne et la tête fémorale* (caput femoris*) (figure 21), mais aussi le trochanter (*trochanter femoris*), et plus précisément sa surface articulaire (*facies articularis antitrochanterica*)*

7. On pourrait aussi considérer que les participants au schéma sont les limaces et l'agent intentionnel. L'antilimace devient alors un instrument utilisé par l'agent pour produire sa force antagoniste.

8. La poitrine, prise dans le sens de 'ce qui est à l'avant du buste'.

9. Je remercie Fabio Montermini de m'avoir signalé cet exemple. *Antibari* est l'ancien nom de la ville de Bar qui se trouve sur la côté adriatique, sur le territoire du Monténégro, en face de la ville de Bari. Bari est une ville italienne, située elle aussi sur le bord de l'Adriatique.

*avec l'**antitrochanter** de l'os coxal.* wwwbibli.vet-nantes.fr/theses/2003/guillet3-13/part1.pdf

Ces exemples permettent de connecter l'interprétation spatiales endocentriques aux autres interprétations de la préfixation en *anti-*, en faisant ressortir l'idée d'opposition qui les caractérise. Le fait de se trouver devant ou de faire face constitue en effet une forme d'opposition induite par l'orientation en miroir des entités (Vandeloise 1986). La configuration est ainsi associée à l'idée de confrontation que l'on retrouve dans les dérivés comme *antiparlementaire*. D'autre part, l'orientation en miroir est aussi associée à une inversion qui rapproche cette interprétation de celle de dérivés comme *antihéros*. Cette dérivation n'est plus aujourd'hui productive ce qui explique qu'elle soit ignorée dans la plupart des études de la préfixation en *anti-*.

4.3.3 Interprétations spatiales exocentriques

L'interprétation spatiale exocentrique est à l'inverse de la précédente, à la fois productive et récente. Les dérivés de notre corpus qui relèvent de cette interprétation appartiennent aux terminologies des satellites, de la médecine et de la géologie (67). Par exemple, la face anti-Terre d'un satellite (68) et celle qui se trouve de l'autre côté de la face Terre, c'est-à-dire de celle qui est tournée vers la Terre. De même, le bord anti-mésentérique (69) est le bord des anses de l'intestin grêle qui se situe du côté opposé au bord mésentérique, i.e. au bord sur lequel est fixé le mésentère. L'interprétation est légèrement différente pour la direction anti-apenninique qui est perpendiculaire à la direction des Appennins.

(67) *anti-Soleil* ; *antisolaire* ; *antispéculaire* ; *antitumoral* ; *antipéristaltique* ; *anti-apenninique*

(68) *Le second est de type "tube", pour les satellites disposant d'une interface sur leur face **anti-Terre**.* www.air-defense.net/index.php?index=20&num_art=203

(69) *L'anse est détubulisée en sectionnant longitudinalement le tube digestif sur le bord **antimésentérique** au bistouri électrique.* linkinghub.elsevier.com/retrieve/pii/S0003440104000774

Ces dérivés sont exocentriques. Ce sont des adjectifs décrivant des directions dont les bases sont des noms qui désignent des entités (Terre, Soleil, tumeur, mésentère, Apennins) ou des activités (mouvement péristaltique). Tous présentent une idée d'opposition qui légitime l'utilisation du préfixe *anti-*. Ces exemples font par ailleurs apparaître une configuration des orientations qui n'est ni en tandem, ni en miroir. Ainsi, la face Terre et la face anti-Terre d'un satellite sont situées de part et d'autre de ce dernier et présentent une orientation que l'on pourrait appeler « en dos-à-dos ». Il en va de même du bord mésentérique et du bord antimésentérique.

Les exemples présentés dans cette section sont relativement rares. Ils sont absents des études antérieures de la préfixation en *anti-*. Leur découverte est directement liée à l'approche extensive adoptée dans la présente étude.

4.3.4 Interprétations logiques endocentriques

L'interprétation logique endocentrique concerne l'une des deux grandes sous-séries de dérivés en *anti-*, à savoir celle qui contient des exemples comme en (70) ou *anti-Sarkozy* en (71). Ces dérivés sont endocentriques. Un antihéros est une sorte de héros, personnage important d'un œuvre qui est présenté par un critique ou un commentateur comme n'étant pas un héros comme il faut. Ce n'est pas un personnage attachant, intelligent, courageux, beau, etc. Un antihéros est donc une entité qui dispose d'un trait classifiant essentiel de la catégorie des héros, mais dont les traits caractérisants sont inversés. Il en va de même dans le cas de *anti-Sarkozy* (71). Le nom propre *Sarkozy* est converti par antonomase en un type (voir section 4.3.1). Juppé est présenté comme un Sarkozy, en l'occurrence un homme politique de la droite moderniste, mais un Sarkozy inversé : il a fait « des études brillantissimes » ; il a « une formation littéraire d'une rare qualité » ; il est « discret, pudique, avare de ses sentiments » ; il « n'[étale] jamais sa vie privée » ; etc. [10]. Notons que ces dérivés ont la même catégorie grammaticale que leur base et le même genre lorsque celle-ci est nominale.

(70) *antihéros*, *antiroman*, *antijeu*, *antipape*, *antipédagogie*, *antipsychiatrie*

(71) *Juppé, c'est l'**anti-Sarkozy**.* `pascaldecaillet.blog.tdg.ch/archive/2008/03/10/juppe-l-anti-sarkozy.html`

L'inversion des propriétés caractérisantes est bien identifiée dans les études antérieures, notamment par Rey (1968), Durand (1982), Fradin (1997a, b) ou Heyna (2008). Ces propriétés sont identifiés comme des propriétés stéréotypiques dans Fradin (1997a, b) et comme des propriétés prototypiques dans Heyna (2008) [11]. Par ailleurs, les propriétés inversées sont spécifiques à chaque locuteur et à chaque instance de chaque dérivé. Heyna (2008) note ainsi que *anti-* ne précise pas les propriétés qui sont annulées.

Cette interprétation de *anti-* est très proche de celle du préfixe *contre-*, lui aussi endocentrique (72) [12]. Comme *anti-*, *contre-* sert d'inverseur (De Mulder et Amiot 2003) mais il est plus spécialisé que ce dernier. *Contre-* inverse l'orientation de la base (orientation argumentative, axiologique, orientation de la courbure, du don, etc.) là où *anti-* inverse de façon globale l'ensemble de ses propriétés caractérisantes.

(72) *manifestation* > *contremanifestation*
courant > *contrecourant*
courbe > *contrecourbe*
hermine > *contrehermine*
valeur > *contrevaleur*

Les dérivés endocentriques en *anti-* dont l'interprétation est logique peuvent aussi être utilisés pour indiquer l'absence de la base ou l'incompatibilité de ce que désigne le dérivé avec cette dernière. Par exemple en (73), *antipédagogique* exprime que le nom recteur est

10. Ces qualités sont extraites de la description de Pascal Décaillet, auteur de l'article dont (71) est le titre.
11. Heyna (2008) considère que dans la série *antihéros*, les propriétés prototypiques sont annulées et non inversées.
12. Ces exemples sont empruntés à De Mulder et Amiot (2003).

dépourvu de toute pédagogie et non pas qu'il relève de l'antipédagogie. Cet emploi de *anti-* est très proche du préfixe *a-*. On le retrouve également dans les dérivés de la terminologie des mathématiques comme en (80). L'exemple (74) présente une nuance sémantique légèrement différente. L'adjectif *antipédagogique* y est utilisé pour exprimer que le nom recteur est contraire à la pédagogie.

(73) *Ce qui au passage est* ***anti-pédagogique*** *puisque cela revient à laisser ces individus dans leur démarche peu respectueuses des dispositions d'autrui.* `amitie-entre-les-peuples.org/spip.php?article673`

(74) *Le ministre de l'Éducation nationale, dénonce le caractère* ***antipédagogique*** *de l'offre de devoirs proposée par le site « Faismesdevoirs.com »* `www.education.gouv.fr/cid23983/xavier-darcos-denonce-la-mise-en-place-d-un-marche-du-moindre-effort.html`

L'emploi de *antihéroïque* en (75) présente la même interprétation. Ce dérivé est en effet utilisé pour évoquer les personnages qui ne sont pas les héros dans les histoires et notamment les méchants et les traîtres auxquels s'opposent ces derniers.

(75) *Je tenais ce discours l'autre jour à un ami, que la moustache est un attribut profondément* ***anti-héroïque****, dans la mesure où les héros moustachus sont sous-représentés dans les oeuvres de fictions; en outre, les moustachus sont sur-représentés dans les rôles de traîtres.* `finpoil.wordpress.com/2009/03/16/la-moustache-est-anti-heroique/`

Il en va de même pour un adjectif comme *anti-obèse* (76) qui qualifie des rats qui n'engraissent pas lorsqu'ils mangent trop. C'est aussi un contre-exemple supplémentaire à l'hypothèse que les dérivés en *anti-* ne peuvent pas avoir comme bases des adjectifs simples (Dell 1970, Zribi-Hertz 1973, Durand 1982).

(76) *Régulation du stockage et de l'oxydation des acides gras dans un modèle de rat* ***anti-obèse****: implications du muscle et du complexe adipocyte-foie.* `cjc.univ-lyon2.fr/158-Dossiers-acceptes-dans-cadre-20e-Appel-d-Offres-Centre-Jacques-Cartier.html`

L'interprétation logique endocentrique impose peu de contraintes sur les bases des dérivés en *anti-*. Ces dernières peuvent être des noms communs (*antiroman*), des noms propres (*anti-Sarkozy*), des syntagmes nominaux sans déterminant comme *anti-maison de retraite* (27) ou comme le premier exemple de (32), des pronoms suivis d'une complétive comme le deuxième exemple de (32). J'ai également trouvé un exemple remarquable (77) dans lequel le préfixe n'est pas placé devant le radical mais à l'intérieur de ce dernier. Le dérivé *journal anti-intime* est clairement formé sur l'expression *journal intime*. Ce mode de marquage se retrouve dans les constructions du type *pêcheurs sous la glace* étudiés en détail dans Kerleroux (2007). Dans *pêcheurs sous la glace*, ce n'est pas le verbe *pêcher* qui est suffixé, mais le syntagme nominal *pêche sous la glace* : les pêcheurs sous la glace sont les personnes qui pratiquent la pêche sous la glace. L'exemple (77) peut également être rapproché d'un adjectif comme *peinable de mort* construit sur *peine de mort* (Hathout et al. 2003) ou de noms comme *dépendeur d'andouilles* ou *donneur de leçons* respectivement construits sur *dépendre les andouilles* et *donner des leçons* (Roché en préparation).

(77) *Non, ceci ne sera pas le journal* ***anti intime*** *de ma vie trépidante et passionante.* `mistermisfit.canalblog.com/archives/2006/08/06/2418577.html`

L'inversion des propriétés caractérisantes peut également concerner de façon exceptionnelle des activités comme dans *antimordre* (78) ou *antirire* (79). On peut noter dans ces deux exemples la présence d'une explication du sens de ces dérivés.

(78) *J'ajoute quand même que j'aime aussi "****anti-mordre****".*
On prend un caramel dur, on le croque vivement jusqu'à avoir une poussiere de caramel dur en bouche, puis cette poussiere, si vous avez de la salive va fondre puis se reconstituer au fond des molaires. Le plus gros du travail est fait, il suffit ensuite de mordre tres fort (ah on y revient !) et on attend en mordant. Apres 1 min, on essaie d'ouvrir la bouche et là on a la sensation de soulever ses dents, de les déchausser : c'est l'anti-morsure ! `condylure.mesdiscussions.net/condylure/DTC/Trucsserieux/aimez-mordre-dents-sujet_2663_1.htm`

(79) *AH AH AH. Vous voulez* ***antirire*** *(le contraire de rire) ? Alors lisez ces... enfin de ces machins là, en-dessous.* `tapecenomsuperlong.free.fr/blagues.html`

L'utilisation de radicaux verbaux est probablement une solution de dernier recours. Dans l'exemple (78), le locuteur préfère la forme marquée *antimordre* à une solution plus neutre comme *antimorsure* car cette dernière serait ambiguë : on ne saurait s'il veut dire qu'il aime antimordre ou être antimordu.

Relèvent également de l'interprétation logique endocentrique, les termes de mathématique comme (80) et les termes de la physique des particules qui nomment les particules qui appartiennent au monde de l'antimatière (81) et leur propriétés (82).

(80) *anticommutatif* ; *antiréflexif* ; *antisymétrique* ; *antiparallèle* ; *antichaîne*

(81) *antimatière* ; *antiparticule* ; *antiunivers* ; *antiprotons* ; *antiquark* ; *antigluon*

(82) *antirouge* ; *antivert* ; *antibleu* ; *antibeau* ; *anticharmé* ; *antiétrange*

Chaque particule de matière se trouve en correspondance avec une antiparticule. Cette correspondance est étendue à leurs propriétés : à chaque propriété de particules matérielles correspond une antipropriété des antiparticules correspondantes.

4.3.5 Interprétations adversatives exocentriques

Les dérivés en *anti-* dont l'interprétation est adversative exocentrique sont à la fois les plus nombreux et les plus centraux. Ils présentent en outre une grande homogénéité du point de vue sémantique : (*i*) leurs bases sont interprétées soit comme des nuisances (83) soit comme des causes de nuisances (84) ; (*ii*) ces dérivés décrivent tous la qualité de s'opposer à ces nuisances. C'est en ce sens qu'ils sont adversatifs. Ils sont aussi exocentriques. Par exemple, un antivol permet de se protéger contre le vol et ne peut pas être considéré comme une forme de vol. Un antilimace empêche les limaces de consommer les plantations et ne peut être conçu comme une sorte de limace, etc. Cette interprétation est confirmée par le fait que les dérivés de cette sous-série sont des adjectifs alors que leur bases peuvent être nominales (*panne*), verbales (*dormir*) ou adjectivales (*triste*).

(83) *antivol* ; *antipanne* ; *antinoyade* ; *anti-marée noire* ; *antilèpre* ; *antidérapant* ; *anticoupant* ; *antistérile* ; *antitriste*

(84) *antiparlementaire* ; *antilimace* ; *antipublicité* ; *anti-Bush* ; *anti-Sarkozy* ; *anti-Belgique* ; *antidormir* ; *antisec*

Pour les dérivés dont les bases désignent des nuisances, le sens des premiers est directement lié à celui des secondes. À l'inverse, pour des dérivés comme en (84), l'identification de la nuisance met en jeu un mécanisme métonymique : la base fonctionne dans ce cas comme un attracteur et les nuisances sont l'un de ses satellites (voir section 4.3.1). Deux cas peuvent être distingués : les nuisances peuvent soit être identifiables soit ne pas l'être. Le dérivé *antilimace* relève du premier cas. Le satellite visé est le fait que les limaces mangent les plantations. La base est dans ce cas l'agent d'un événement qui est considéré comme une nuisance. Mais il n'est pas toujours possible d'identifier précisément la nuisance visée et cette dernière peut même ne pas l'être par le locuteur qui utilise le mot. Par exemple pour *antipublicité* les sources de nuisances possibles sont nombreuses : interférence avec les contenus éditoriaux, pollution visuelle, encombrement des boîtes aux lettres, gaspillage de papier, etc. Cependant, leur identification précise n'est pas essentielle à la bonne interprétation du mot. Il suffit seulement d'admettre que la publicité est une source de nuisances ou qu'elle peut être considérée comme telle.

L'interprétation de la base comme une nuisance ou une source de nuisances est l'une des contributions de cette étude. Elle affine celle de Heyna (2008) qui considère elle aussi que les bases des dérivés de la sous-série *antilimace* sont interprétées métonymiquement. Pour Heyna, les bases de ces dérivés évoquent « confusément » l'un quelconque de leurs satellites et réfèrent « de façon nébuleuse ». Je propose à l'inverse que *anti-* impose une restriction sur les satellites associés à l'attracteur et que l'accès soit limité aux seules nuisances. On peut également comparer la présente étude à celle de Fradin (1997a) qui s'intéresse aux adjectifs dont le radical est un participe présent comme *antidérapant*. Fradin propose une analyse dans laquelle la fonction sémantique de *anti-* est d'introduire un causateur destiné à contrarier l'évenance exprimée par le lexème de base. Cette idée générale est proche de la description proposée ici. La principale différence entre cette analyse et la mienne est que j'identifie une nuisance et un agent intentionnel là où Fradin (1997a) introduit une évenance. Son analyse ne peut, de ce fait, pas être généralisée à des dérivés comme *des sentiments anti-arabes* ou *une opinion anti-Sarkozy* pour lesquels il n'y a aucune évenance particulière. Dans le premier cas, une personne anti-arabe considère que les arabes sont des nuisances, de par leur simple présence ou leur simple existence. Dans le second, c'est la politique, les idées, les propos, le comportement de Nicolas Sarkozy qui sont considérés comme des nuisances, sans qu'il n'existe d'évenance particulière qui leur soit associée. L'analyse de Fradin ne permet pas non plus de rendre compte des dérivés qui désignent une protection passive comme *antifroid* (85). Des gants antifroid n'agissent pas contre le froid. Notons cependant que le causateur pourrait être dans ce cas l'agent intentionnel qui se sert des gants.

(85) *mikel, ça a l'air super tes gants* ***antifroid****, ça me tente bien, hop je vais sur le site!*
`www.lerepairedesmotards.com/forum/read.php?f=3&i=284680&t=284680`

Une analyse en termes de nuisance est ainsi à la fois plus simple et plus efficace. Elle rend compte d'autre part de façon triviale des connotations axiologiques dépréciatives des dérivés en *anti-* comme *antiparlementaire* et *antilimace* (Rey 1968, Zribi-Hertz 1973, Heyna 2008). Notons que dans certains exemples (86), ces valeurs peuvent être explicites. Cette valeur se retrouve aussi dans la préfixation en *contre-* (De Mulder et Amiot 2003, Dal 2003b).

(86) *Moi j'ai mon amoureuse, rené char, du vin blanc, bcp de haine et de rage que je mets au service de mon temps militant* ***anti ces 2 zouaves abjectes****. honte à la france, honte à eux, honte à ceux qui les ont élus.j'ai honte.* `www.libe.com/php/pages/pageReactionsList.php?rubId=14&docId=290109&s2=&pp=&next=30`

Les emplois adversatifs exocentriques concernent aussi un groupe relativement homogène de dérivés dont les radicaux sont des adjectifs en *-able*. Ces dérivés sont bien attestés comme le montre l'importante liste présentée en (87). Ils sont sémantiquement proches des participes passés, ce qui explique leur fréquence réduite (voir section 4.2.1).

(87) *anti-arrachable* ; *antiblocable* ; *antibrisable* ; *anticoagulable* ; *anticombustible* [13] ; *anticonsommable* ; *anticopiable* ; *anticrevable* ; *anticrochetable* ; *antidéchirable* ; *antidécrochable* ; *antidéformable* ; *antidégondable* ; *antidégradable* ; *antidéjantable* ; *antidéjaugeable* ; *anti-démontable* ; *anti-déprimable* ; *antidévissable* ; *antidiluable* ; *antiéjectable* ; *antifermentiscible* ; *antifroissable* ; *anti-inflammable* ; *anti-irritable* ; *antimanipulable* ; *anti-orientable* ; *anti-oxydable* ; *antiperçable* ; *antiperméable* ; *antipiratable* (23) ; *antiprévisible* ; *antiputréscible* ; *antirenversable* ; *antispammable* ; *antivibrable* (124h)

Ces dérivés (87) servent à exprimer que leur nom recteur est protégé contre une nuisance donnée comme l'arrachage, le blocage, la coagulation, la consommation, la copie, etc. Ce sens se distingue nettement de celui des dérivés de la forme *in-Xable* qui expriment que leurs noms recteurs ne sont pas *Xables* et sont eux endocentriques.

4.3.6 Interprétations adversatives endocentriques

Les dérivés adversatifs en *anti-* peuvent également avoir des emplois endocentriques. Ces emplois n'ont à ma connaissance jamais encore été décrits. Le premier s'observe avec des dérivés comme (88) lorsqu'ils sont utilisés comme des synonymes de leurs noms bases. L'antidémêlant est un produit qui s'oppose à l'emmêlement des cheveux, l'antidésherbant à la prolifération des mauvaises herbes, l'antidésinfectant aux infections, etc. L'interprétation de ces exemples est endocentrique comme le montrent les exemples (89 ; 90) et (91 ; 92) : la synonymie entre le dérivé et sa base correspond à un niveau maximal d'endocentricité. Dans tous ces exemples, l'ajout du préfixe *anti-* sert à renforcer l'idée que le produit protège contre une nuisance, exactement comme pour les dérivés dont l'interprétation est exocentrique.

(88) *antidécongestionnant*, *antidémêlant*, *antidésherbant*, *antidéperlant*, *antidésinfectant*, *antidésodorisant*

13. Rey (1968) signale la présence de *anticombustible* et *antifermentescible* dans l'édition de 1877 du *Littré*.

(89) *Cette casquette souple et légère, en microfibres* ***anti-déperlantes****, protège aussi bien du soleil que de la pluie sous les climats les plus extrêmes.* www.randosbalades.fr/article/2007/12/12/20071212-Nouveautes-Chantons-sous-la-pluie.php?PHPSESSID=03a920a3eb88c8ab

(90) *C'est pourquoi il ne faut jamais laver les vestes et surtout pas à la machine ou au pressing mais l'usure aidant les mcrofibres* ***déperlantes*** *disparaissent.* club.discount-marine.com/forum/viewtopic.php?p=12347&sid=5d65620ac9af8d70218a761b87ebd7bf

(91) *les vignes avec tous leurs produits nocifs, ben j'ose pas imaginer ce que je dois avoir comme dose d'****anti-désherbant*** *dans le corps !* www.vulgaris-medical.com/forum/s11917-guaifenesin-fibromyalgie-medicament-miracle-svp.html

(92) *moi aussi j'ai de longs cheveux et je n'utilise jamais de sèche-cheveux, trop la flemme, par contre je suis obligée de mettre un* ***anti démêlant*** *sinon c'est la galère à les coiffer, j'en ai pour des heures* www.forumfr.com/sujet90410-post10-avons-nous-ici-des-accros-du-seche-cheveux.html

4.3.7 Unité de famille

La présentation séparée des cinq interprétations possibles des dérivés en *anti-* ne doit pas cacher l'unité de cette préfixation. Cette unité globale est d'une nature familiale plutôt que hiérarchique. Les différentes interprétations des dérivés en *anti-* ne s'organisent pas en une hiérarchie ou une arborescence, avec un sens abstrait général permettant de dériver tous les autres, mais comme un groupe de sous-sens connexes permettant de passer de proche en proche de l'un à l'autre, par ressemblance de famille.

L'existence d'un seul préfixe *anti-* ou de plusieurs est discutée dans différentes études. Par exemple, Corbin (1980) et Durand (1982) considèrent que les sous-séries *antilimace* et *antihéros* relèvent de deux préfixes différents. De même, Corbin (1987) propose quatre règles de construction de mots distinctes dont les dérivés se superposent partiellement. À l'inverse, Rey (1968), Zribi-Hertz (1973) et Heyna (2008) défendent l'idée que tous les dérivés en *anti-* sont formés par le même préfixe. Pour Heyna (2008), la différence entre les dérivés des deux sous-séries provient de la nature référentielle de leurs bases et non du préfixe lui-même. Les arguments avancés par Zribi-Hertz (1973) sont l'unité phonologique et sémantique du préfixe qui dans tous les cas réfère à une opposition malgré ses deux valeurs possibles. Elle note également que « les deux valeurs en question ne se distinguent pas toujours de manière très tranchée, et peuvent même se mêler dans l'interprétation ».

Je me rallie à la position de Zribi-Hertz. Ses arguments peuvent être complétés par quelques autres. On peut ainsi noter que les dérivés dont l'interprétation est exocentrique sont tous des adjectifs. Un deuxième argument est la proximité des interprétations endocentriques et exocentriques au sein de chaque ensemble d'emplois (spatiaux, logiques et adversatifs) :

(*i*) On observe ainsi que le sens des dérivés adversatifs exocentriques et endocentriques comporte la même valeur axiologique dépréciative et qu'à l'inverse, les dérivés dont

l'interprétation est spatiale ou logique en sont dépourvus, et cela indépendamment du fait qu'ils soient endocentriques ou exocentriques.

(*ii*) Les dérivés endocentriques et exocentriques au sein de chaque ensemble d'emplois partagent les mêmes types de bases. Les bases des dérivés dont l'interprétation est spatiale évoquent la position d'entités matérielles, celles des dérivés dont l'interprétation est adversative évoquent des nuisances ou des sources de nuisance. Les dérivés dont l'interprétation est logique partagent même une partie de leurs radicaux (*pédagogie*, *héros*).

(*iii*) Un troisième argument est l'existence de dérivés polysémiques dont l'interprétation peut être soit logique, soit adversative comme *anti-Sarkozy* (71 ; 93). Il existe ainsi une proximité, au niveau des positions lexicales qui correspondent à ces mots. La proximité lexicale est renforcée par l'existence de dérivés susceptibles des deux interprétations en même temps comme *antipape* [14] (Rey 1968), *antipsychiatrie* [15] (Durand 1982) ou *antipédagogie*. Par exemple, un antipape est un pape dont les propriétés caractérisantes sont inversées puisqu'il n'appartient pas à la papauté catholique romaine. Mais c'est aussi une personne qui considère cette papauté comme une nuisance, qui s'oppose à elle et en conteste la légitimité. De même, l'antipédagogie est une pédagogie basée sur des principes différents de ceux de la pédagogie traditionnelle. Elle considère aussi cette dernière comme une nuisance, la combat et cherche à s'imposer à sa place.

(93) *La mobilisation* ***anti-Sarkozy*** *continue sur la Toile.* `www.lemonde.fr/cgi-bin/ACHATS/acheter.cgi?offre=ARCHIVES&type_item=ART_ARCH_30J&objet_id=987148&clef=ARC-TRK-NC_01`

L'unité de la préfixation en *anti-* se manifeste aussi dans la possibilité de rattacher ses principaux emplois au sens originel du préfixe grec *anti-* que l'on peut traduire par 'en face' :

1. On peut tout d'abord être en face dans le face-à-face. Ce sens correspond à l'affrontement, à l'opposition d'une force agoniste et d'une force antagoniste que l'on trouve dans les dérivés dont l'interprétation est adversative comme *antiparlementaire* ou *antilimace*.
2. Le face-à-face peut aussi être spatial comme dans le cas de deux maisons qui se font face, des deux côtés de la rue. Ce sens se retrouve dans les dérivés à sens positionnel comme *Anti-Liban* qui fait face au Mont Liban, symétriquement par rapport à la plaine de la Békaa.
3. On peut être en face d'un miroir. On est alors en face d'une image qui est identique mais inversée. Cette inversion peut être rapprochée de celle que l'on observe dans le sens des dérivés dont l'interprétation est logique endocentrique comme *antihéros*.

14. Je rappelle qu'un antipape est une personne qui a exercé la fonction et porté le titre de pape mais dont l'avènement à cette charge n'est pas ou plus reconnu aujourd'hui comme régulier et valable par l'Église catholique romaine (Wikipédia). *Antipape* figure dans la nomenclature du TLF.

15. L'antipsychiatrie est une théorie psychiatrique qui tend à interpréter la maladie mentale dans une perspective sociologique (d'après Wikipédia). *Antipsychiatrie* est une entrée secondaire dans le TLF.

4. Ce qui est en face se trouve devant. Cette relation spatiale est exploitée par les dérivés anciens comme *antichambre* [16].

Il ne s'agit naturellement pas ici de reconstituer l'histoire du préfixe *anti-*, mais seulement de montrer que 'en face' peut aussi servir de sens de base à *anti-* et qu'il couvre une bonne partie de la diversité de ses emplois.

Il se dégage ainsi une unité du préfixe *anti-* similaire à celle des dérivations en *-isme* et en *-iste* défendue dans Roché (2010a) et à celle de la préposition *contre* étudiée par De Mulder et Amiot (2003). De Mulder et Amiot décrivent notamment le sens de cette préposition comme un réseau sémantique qui comporte un sens abstrait qui correspond à l'idée d'inversion et quatre sens contextuels.

4.4 Caractérisation catégorielle

La question catégorielle est au centre de nombreuses études de la préfixation en *anti-* du fait des difficultés d'analyse posées par les constructions parasynthétiques comme *antiparlementaire*. L'importance accordée à cette question est une conséquence directe de la conception traditionnelle de la préfixation comme homocatégorielle par nature. Par exemple, *in-* ne formerait que des adjectifs dont les bases sont des adjectifs et *re-* que des verbes dont les bases sont des verbes. C'est la position adoptée entre autres par Rey (1968), Dell (1970), Zribi-Hertz (1973), Voir (1982). Ainsi, *anti-* ne pourrait pas former seul *antiparlementaire* à partir de *parlement* ; cette formation serait le résultat d'une opération qui comporte simultanément une préfixation en *anti-* et une suffixation en *-aire*. Cette analyse a été combattue vigoureusement par Corbin (1980, 1987, 2001b) qui défend l'idée que la finale *-aire* est un intégrateur paradigmatique et non un affixe dérivationnel. Une autre solution permettant de contourner cette difficulté est proposée, notamment par Durand (1982) et Fradin (1997b). Elle repose sur l'observation que les noms et leurs adjectifs de relation sont quasi identiques d'un point de vue sémantique (Roché 2009). Cette identité permet d'analyser les dérivés dont les radicaux sont des adjectifs comme étant simplement des lexèmes construits sur ces adjectifs. Ce débat semble aujourd'hui dépassé et il est communément admis que les propriétés catégorielles des affixes sont d'abord définies par leurs propriétés sémantiques (Corbin 2001b, Hathout et al. 2003, Amiot 2008).

La description sémantique élaborée en section 4.3 distingue d'un côté des dérivés endocentriques et par suite homocatégoriels et de l'autre des dérivés exocentriques qui peuvent être transcatégoriels. Dans le premier groupe on trouve des couples base : dérivé [17] comme en (94) dont les deux membres sont de la même catégorie. Les trois catégories majeures, nom, verbe et adjectif, sont représentées. Dans le second groupe (95), on trouve aussi des bases qui appartiennent aux trois catégories majeures.

(94) *Atlas* : *anti-Atlas*, *héros* : *antihéros*, *obèse* : *anti-obèse*, *réfléxif* : *antiréflexif*, *mordre* : *antimordre*, *désherbant* : *antidésherbant*

16. Cette relation pourrait être étendue à la dimension temporelle pour rendre compte du sens du verbe *antidater*.

17. Je rappelle que les instances des relations lexicales sont notés x : y.

(95) *mésentère* : *antimésentérique*, *héros* : *antihéroïque*, *parlement* : *antiparlementaire*, *triste* : *antitriste*, *stérile* : *antistérile*, *dormir* : *antidormir*

À l'exception de quelques noms de la terminologie des mathématiques, les dérivés dont l'interprétation est exocentrique sont à l'origine des adjectifs, y compris ceux qui sont employés nominalement. Ces noms permettent en effet de nommer des entités d'après leurs propriétés. Par exemple, un antilimace est produit désigné d'après sa qualité qui est de protéger des limaces. Plusieurs arguments en faveur de cette analyse peuvent être avancés :

1. Le genre des noms comme *antirides* ou *antilimace* est toujours le masculin (excepté lorsqu'ils désignent une femme, comme dans *une anti-Sarkozy*, ou un animal femelle), même lorsque la base est un nom féminin ou lorsque le nom désigne une réalité que l'on peut décrire par un nom féminin. On dit par exemple *un antirides* même quand il s'agit d'une crème.
2. Les marques adjectivales sont présentes lorsque les mots sont utilisés comme des noms. On dit ainsi *Jean est un antiparlementaire* et non *Jean est un antiparlement*. Or le recours au radical adjectival *parlementaire* serait inutile si ce nom était formé directement par préfixation en *anti-* et non par nominalisation de l'adjectif *antiparlementaire*.
3. Comme le notent Zribi-Hertz (1973) et Durand (1982), les dérivés de cette sous-série qui sont utilisés comme des noms peuvent être paraphrasés au moyen d'un syntagme nominal dans lequel ils fonctionnent comme des adjectifs. Par exemple, *un antirouille* peut se paraphraser par *un produit antirouille*.

On ne peut donc pas considérer comme le fait Heyna (2008) que les dérivés de la sous-série *antilimace* appartiennent à une catégorie grammaticale qui regroupe les substantifs et les adjectifs.

On peut aussi noter que les dérivés adjectivaux en *anti-* s'accordent avec leur nom recteur selon les schémas flexionnels propres à cette catégorie, et ceci même lorsqu'il n'existe pas de correspondant féminin à leurs radicaux nominaux. Fradin (1997b:336) et Levkovych (2004:60) relèvent par exemple l'accord de l'adjectif dans le syntagme *la lutte anti-sous-marine* (96). Cette situation est tout à fait normale : *anti-sous-marin* est un adjectif à part entière et se fléchit à ce titre de la même manière que n'importe quel autre mot de sa catégorie.

(96) *Les premiers navires chargés de la lutte* ***anti-sous-marine*** *emportent deux charges, qui sont larguées par la poupe.* `fr.wikipedia.org/wiki/Grenade_anti-sous-marine`

4.5 Radicaux d'emprunt

Le modèle proposé dans le chapitre 2 rend compte très simplement des principaux phénomènes qui apparaissent dans la dérivation en *anti-*. Ses capacités sont confirmées par l'analyse des données présentées dans Levkovych (2004) et Plénat et al. (2004a). Je rappelle que ces données ont été obtenues par une collecte extensive de dérivés en *anti-* réalisée notamment au moyen de Webaffix (Hathout et Tanguy 2006) et par une exploration manuelle du Web.

Le préfixe *anti-* ayant une grande autonomie formelle (voir section 2.10), les contraintes des sections 2.9 et 2.11, exceptée la contrainte de transparence des séries dérivationnelles (C2.7, page 51), n'affectent que la sélection des radicaux des dérivés construits par cette préfixation. Les données confirment que ces radicaux sont empruntés à des adjectifs déjà présents dans le lexique (contrainte de recyclage (C2.3), page 47) et morphologiquement proches de la base (contrainte de transparence de la base (C2.6), page 50). Par suite, ces adjectifs appartiennent aux familles lexicales de ces dérivés et de leurs bases. Les données montrent aussi que tous les adjectifs de relation peuvent être utilisés et illustrent la variété des stratégies permettant de satisfaire au mieux le système de contraintes.

4.5.1 Parasynthèse

L'une des questions incontournables que toute étude la préfixation en *anti-* se doit de traiter est celle des dérivés dont le radical est une forme adjectivale comme *anticancéreux* ou *antiparlementaire*. Clairement, dans *un médicament anticancéreux* l'adjectif peut être glosé par 'qui s'attaque au cancer' mais non par 'qui s'attaque aux cancéreux' ; de même, dans *une ligue antiparlementaire*, le dérivé signifie 'opposée au parlement' et non 'qui combat les parlementaires'. L'analyse de ces dérivés présente donc une difficulté : leurs sens sont construits sur ceux des lexèmes *cancer* et *parlement* alors que leurs radicaux sont *cancéreux* et *parlementaire*. Cet écart est en partie à l'origine de l'intérêt que les morphologues portent à la préfixation en *anti-*. En réalité, ces constructions ne sont en rien particulières et ne sont intéressantes que parce qu'elles mettent en défaut les conceptions trop simplistes de la morphologie dérivationnelle, et parce qu'elles permettent d'observer les effets de quelques-unes des contraintes morpholexicales qui pèsent sur la formation des mots.

Ces constructions sont dites parasynthétiques pour des raisons historiques : elles ont été initialement analysées comme des mots construits par parasynthèse. L'analyse traditionnelle, formulée dans un cadre morphématique, consiste à ajouter au dispositif de la morphologie dérivationnelle un procédé supplémentaire destiné spécifiquement au traitement de ces constructions. Par exemple, *antigrippal* est analysé comme un mot construit en ajoutant en même temps à *grippe* le préfixe *anti-* et le suffixe *-al*. Ainsi, en plus de la dérivation affixale (préfixation et suffixation), de la conversion et de la composition, il existerait un procédé nommé **parasynthèse** qui réalise simultanément une préfixation et une suffixation. Je ne reviendrai pas ici sur les motivations de cette proposition et ni sur les problèmes qu'elle engendre. Ils sont discutés en détail par Corbin (1980). L'analyse parasynthétique est essentiellement utilisée dans les travaux les plus anciens sur la préfixation en *anti-*, notamment Rey (1968), Dell (1970), Zribi-Hertz (1973), Voir (1982). Dell (1970) propose par ailleurs de compléter le procédé parasynthétique par un **principe de copie** qui soumet la sélection des suffixes à la pression du lexique existant. Ce principe de nature analogique rend compte de carrés comme (97a) ou (97b). Il est également invoqué pour les dérivés en *anti-* comme en (98) même si dans ce cas l'analogie est seulement formelle, et trompeuse.

(97) a. *approuver* : *approbation* :: *désapprouver* : *désapprobation*

b. *fidèle* : *fidélité* :: *infidèle* : *infidélité*

(98) *parlement* : *parlementaire* :: *anti-parlement* : *anti-parlementaire*

Corbin (1980, 1987) rejette l'analyse parasynthétique et réfute l'existence d'un tel procédé. Elle affirme que les constructions traditionnellement considérées comme parasynthétiques sont en réalité seulement préfixées, et que le phénomène observé découle d'un ajustement formel postérieur à la préfixation, induit par le principe de copie. Ce principe serait à l'origine de la finale adjectivale des dérivés comme *anti-parlementaire* (Fradin 1996). Corbin (1987) propose en outre une version étendue du principe de copie permettant de rendre compte de cas comme (99) dans lesquels l'adjectif de relation *ovulatoire* n'est pas formé sur le nom *ovulation* mais sur un autre mot, en l'occurrence le verbe *ovuler*. Dans des travaux ultérieurs, Corbin (2001a, b) développe une analyse proche de celle que je propose ici, dans laquelle le principe de copie est remplacé par une contrainte de conformité catégorielle (voir section 2.11.3).

(99) *ovulation* : *ovulatoire* :: *anti-ovulation* : *anti-ovulatoire*

Je signale qu'il existe aussi entre ces deux extrêmes théoriques (la tradition d'un côté et l'analyse corbinienne de l'autre), une troisième voie proposée par Roché (2009) dans laquelle une dérivation se compose d'un ensemble d'opérations sémantique, catégorielle et formelle. Chaque opération peut donner lieu à un marquage visible (*overt*) ou ne pas avoir de réalisation phonologique (*covert*). Par exemple, *dératiser* est formé par une dérivation dans laquelle l'opération sémantique est réalisée par le préfixe *dé-* et l'opération catégorielle par le suffixe verbal *-iser*. À l'inverse, dans *déverminer*, l'opération catégorielle n'a pas de réalisation.

La position que j'adopte dans la présente étude est que les formes des dérivés en *anti-*, traditionnellement analysées comme parasynthétiques, sont en réalité construites en empruntant leurs radicaux à d'autres mots du lexique.

4.5.2 Radicaux suffixés

Le premier point mis en évidence par les données de notre corpus est qu'il n'existe aucune restriction sur les finales des adjectifs dont les formes sont utilisées comme radicaux. Tout adjectif de relation dérivé d'un nom peut servir de radical à un dérivé en *anti-* dont la base est ce nom. Les contraintes de transparence de la base et de recyclage rendent compte de cette régularité. La table 4.3 reprend l'inventaire proposé par Levkovych (2004) et Plénat et al. (2004a).

On pourrait ajouter à cette liste un dérivé comme *antipiratable* construit sur *piratage* (voir (23), page 47), qui fait apparaître que la classe des adjectifs dont les formes sont utilisées comme radicaux n'est pas limitée aux seuls adjectifs de relation. Tout adjectif suffisamment saillant et morphologiquement proche de la base peut être sollicité, y compris un adjectif en *-able*. En d'autres termes, on utilise généralement une forme de l'adjectif relationnel car c'est celui qui est le plus proche, mais dans le cas où le lexique mémorisé n'en contient pas de convenable, la forme d'un autre adjectif proche et suffisamment bien attesté peut servir de radical de secours. Les contraintes de transparence rendent compte très simplement de cette possibilité. Elles sont ainsi nettement supérieures à une règle qui stipulerait que les radicaux des dérivés en *anti-* sont sélectionnés parmi les formes des adjectifs de relation de leurs bases. Le modèle théorique rend compte également du fait que

finale	base	radical nominal	radical adjectival	2e base
-ain	*Rome*	*anti-Rome*	*antiromain*	•
-aire	*particule*	*antiparticule*	*antiparticulaire*	
-ais	*Finlande*	*anti-Finlande*	*antifinlandais*	•
-al	*famille*	*antifamille*	*antifamiliale*	
-an	*Perse*	*anti-Perse*	*antipersan*	•
-ant	*givre*	*antigivre*	*antigivrant*	
-aque	*insomnie*	*anti-insomnie*	*anti-insomniaque*	•
-ard	*la Cagoule*	*anti-Cagoule*	*anticagoulard*	•
-éen	*Europe*	*anti-Europe*	*anti-européen*	•
-el	*carence*	*anticarence*	*anticarenciel*	
-esque	*amiante*	*anti-amiante*	*anti-amiantesque*	
-estre	*terre*	*antiterre*	*antiterrestre*	
-eux	*poussière*	*antipoussière*	*antipoussiéreux*	
-ien	*Napoléon*	*anti-Napoléon*	*antinapoléonien*	•
-ier	*émeute*	*anti-émeute*	*anti-émeutier*	•
-if	*gel*	*antigel*	*antigélif*	
-ile	*fièvre*	*antifièvre*	*antifébrile*	
-in	*Bysance*	*anti-Bysance*	*antibysantin*	•
-ique	*hémorragie*	*antihémorragie*	*antihémorragique*	
-iste	*Bonaparte*	*anti-Bonaparte*	*antibonapartiste*	•
-oire	*aléa*	*anti-aléa*	*anti-aléatoire*	
-on	*Bourgogne*	*anti-Bourgogne*	*antibourguignon*	•
-ois	*Chine*	*anti-Chine*	*antichinois*	•
-ote	*patrie*	*antipatrie*	*antipatriote*	•

TABLE 4.3 – Inventaire partiel des finales des dérivés en *anti-* dont les radicaux sont des adjectifs de relation. La présence d'un point dans la dernière colonne indique que le radical adjectival peut aussi s'utiliser comme un nom. Il exprime, dans ce cas, une deuxième base possible pour le dérivé.

les dérivés dont le radical est une forme adjectivale ont un sens strictement identique à ceux qui ont un radical nominal. Ces formes sont simplement associées à une seule et même position lexicale.

4.5.3 Allomorphie

La contrainte de recyclage intervient aussi dans l'analyse des dérivés dont les radicaux présentent des allomorphies [18] comme en (100a). Ces allomorphies sont identiques à celles

18. Dans le modèle à quatre niveau, les allomorphies se distinguent des supplétions par le fait que les premières concernent des mots qui appartiennent à la même famille morphologique et que les secondes s'éta-

des adjectifs de relation dérivés par suffixation (100b). D'autres exemples comme *Dieu* > *antidivin* ou *asthme* > *anti-asthmatique* présentent des allomorphies semblables.

(100) a. *anticarenciel* ; *antibyzantin* ; *antifébrile* ; *anti-aléatoire* ; *antibourguignon*

b. *carence* > *carenciel* ; *Byzance* > *byzantin* ; *fièvre* > *fébrile* ; *aléa* > *aléatoire* ; *Bourgogne* > *bourguignon*

4.5.4 Supplétion

Les radicaux adjectivaux des dérivés en *anti-* ne sont pas tous des adjectifs formés par suffixation du nom qui exprime leur base. Des formes supplétives peuvent aussi servir de radicaux comme en (101). Ces exemples montrent que l'adjectif dont les formes servent de radical n'est pas nécessairement un membre de la famille morphologique de la base et qu'il peut être recherché, si nécessaire, dans la famille lexicale de cette dernière.

(101) *champignon* > *antifongique* ; *champignon* > *antimycosique* ; *oreillons* > *anti-ourlien* ; *fièvre* > *antipyrétique*

4.5.5 Substitution de suffixes

Les adjectifs de relation des noms peuvent aussi être formés par substitution de suffixes comme en (102a). Ainsi, *vibratoire* est lexicalement l'adjectif relationnel de *vibration* même si son radical est le thème *vibrat-* du verbe *vibrer*. Plus généralement, les adjectifs de relation des noms en *-ion* sont souvent formés en substituant à cette finale un suffixe comme *-oire*, *-if* et *-ile*. Naturellement, les formes des adjectifs de relation des noms en *-ion* peuvent aussi être construites par suffixation simple comme en (102b). La substitution de suffixes n'est pas limitée aux noms en *-ion*. Elle concerne aussi les dérivés en *-isme*, *-iste*, *-ique* ou *-ien* comme en (102c) (Roché 2007).

(102) a. *vibration* > *antivibratoire* ; *inflammation* > *anti-inflammatoire* ; *abrasion* > *anti-abrasif* ; *érection* > *anti-érectile*

b. *gestation* > *antigestationnel* ; *gravitation* > *antigravitationnel*

c. *jansénisme* > *antijanséniste* ; *magnétisme* > *antimagnétique* ; *palludisme* > *antipalludéen*

4.5.6 Surcharge

On rencontre également des cas où l'adjectif de relation d'un nom utilise la forme d'un autre adjectif. Par exemple, *électoral* est l'adjectif de relation de *électeur*. Son radical est issu d'une forme de ce nom. Mais *électoral* est aussi l'adjectif de relation de *élection*. Ce dernier utilise la forme d'un adjectif proche qui se trouve déjà dans le lexique (Roché 2009). Cette surcharge se retrouve naturellement aussi au niveau des dérivés en *anti-* où *anti-électoral* signifie 'qui s'oppose aux élections' (103). Pour des raisons conceptuelles, le sens

blissent entre des mots qui appartiennent à la même famille lexicale mais à des familles morphologiques différentes.

'qui s'oppose aux électeurs' est moins facile à utiliser car il est plus difficile de percevoir les électeurs comme une nuisance.

(103) *Meeting* ***anti-électoral*** *"désertons les urnes, occupons la rue" du 21 mars 2007 voir ...* farennes.free.fr/reunionspubliques.htm

4.5.7 Identité de forme

On rencontre une situation similaire à la surcharge pour des formes comme *paysan*, *acarien*, *froid* ou *parallèle* qui peuvent s'utiliser indifféremment comme nom et comme adjectif [19]. Les radicaux des dérivés en *anti-* qui ont ces noms comme bases (*antipaysan* (104) ; *anti-acarien* (105) ; *antifroid* (106) ; *antiparallèle*) satisfont donc trivialement la contrainte de transparence catégorielle (C2.9) puisque ce sont aussi des formes adjectivales. Pour *antifroid*, c'est la spécificité des marques partagées avec des mots de la catégorie des adjectifs, en l'occurrence avec l'adjectif *froid*, qui permet la pleine satisfaction de la contrainte.

(104) *La consolidation de la politique* ***anti-paysans*** *en Bolivie.* www.alencontre.org/Bolivie/bolivie19.html

(105) *ACAR-HOUSSES housses* ***anti acariens*** *pour matelas et literie.* www.acarhousses.com/

(106) *Pour les pieds, la CREME* ***ANTI FROID*** *AKILEINE avec ses composants permettant de garder les pieds au sec, d'activer la micro-circulation et de les réchauffer.* www.tepaseul.fr/category/prends-soin-de-toi/ma-beaute

Les adjectifs précédents ont des formes distinctes au féminin (107 ; 108 ; 109). Parmi ces dérivés, *antifroide* (109) est remarquable parce que son radical est une forme de l'adjectif simple *froid*, or ces adjectifs sont réputés ne pas pouvoir engendrer de dérivés en *anti-* (Dell 1970, Zribi-Hertz 1973, Durand 1982). Elle est également intéressante parce que les locuteurs la ressentent comme étant légèrement plus marquée que *antifroid*. *Antifroid* est en effet totalement satisfaisant pour l'ensemble des contraintes qui contrôlent la préfixation en *anti-*, notamment pour la contrainte de transparence catégorielle. À l'inverse, le radical de *antifroide* n'étant pas une forme du nom *froid*, il est légèrement pénalisé par la contrainte de transparence de la base.

(107) *La politique* ***anti-paysanne*** *du gouvernement Mesa (novembre 2003).* www.alencontre.org/dossier/Bolivie.html

(108) *Recouvrez votre matelas et vos oreillers avec des housses* ***anti-acariennes****.* www.astrazeneca.fr/AZFR/VotreSante/Pneumologie/DossierPratique/82471.htm

(109) *Crème protectice* ***anti-froide****.* www.bingotrade.ch/pdf/catalogue_2003.pdf

19. Le nom et l'adjectif correspondant sont dans une relation de conversion (Roché 2005). Pour Corbin (2001a), il s'agit d'une suffixation dont le suffixe a été inhibé.

4.5.8 Élision

La façon la plus simple de construire un dérivé en *anti-* consiste à utiliser comme radical la forme de l'adjectif de relation de sa base. Mais d'autres stratégies existent. L'une d'elles est l'élision que l'on rencontre pour les dérivés dont la base peut être exprimée par un composé nominal constitué d'un nom et d'un adjectif (voir section 4.2.4). Le nom est dans ce cas simplement élidé afin de ne conserver que l'adjectif et satisfaire ainsi parfaitement la contrainte de transparence catégorielle. Par exemple, *radiation atomique* permet de former *anti-radiation atomique* (110) et *anti-atomique* (111), *attaque aérienne* de former *anti-attaque aérienne* (112) et *anti-aérien* (113), et *politique sécuritaire* de former *anti-politique sécuritaire* et *antisécuritaire*. Dans tous les cas, la forme qui a un radical adjectival est beaucoup plus fréquente que celle qui contient le nom et l'adjectif (table 4.4). La seconde forme est en effet plus marquée parce qu'elle ne satisfait ni la contrainte de recyclage (son radical n'est pas une entrée lexicale classique) ni la contrainte de transparence catégorielle (une séquence nom-adjectif ne ressemble pas suffisamment à un adjectif). À l'inverse celle dont le radical est une forme adjectivale satisfait ces deux contraintes, mais pas celle de transparence de la base (C2.6). Les chiffres de la table 4.4 font apparaître que la contrainte de recyclage (C2.3) pénalise fortement les radicaux qui ne sont pas des mots simples.

N Adj	*anti-N Adj*	*anti-Adj*
attaque aérienne	20	120 000
radiation atomique	6	240 000
politique sécuritaire	3	100 000

TABLE 4.4 – Comparaison du nombre des formes ayant pour radical l'expression nominale et l'adjectif. Les chiffres correspondent aux nombres de pages indexées par Google qui contiennent ces séquences (décompte réalisé le 17 octobre 2009). Ils font clairement apparaître que les expressions nominales sont fortement défavorisées face aux formes monolexicales, plus prototypiques.

(110) *Le 2e choix celui du nucléaire permettait de rencontrer un agent immoblier offrant un catalogue d'abris anti-atomique, du plus luxueux ou plus simple, de combinaisons* ***anti-radiations atomiques*** *et autre dosimètres très pratiques pour mesurer soi-même son taux de radioactivité.* `www.greenpeace.org/france/groupes-locaux/bordeaux/actus/bdx-240207-eprourevolutionenergetique?mode=send`

(111) *convoqué par l'organisme du don de sang, j'ai appris tout ça !!!!! pas d'hépatite C à l'époque, ils m'ont acceuillis avec des masques et combinaisons* ***anti atomiques*** *!!!!!!!!! ils voulaient même pas me faire de prise de sang !!!!!!* `www.hepatites.net/index.php?name=PNphpBB2&file=printview&t=1233&start=0`

(112) *En même temps quand tu vois les journalistes TV français aux 20 heures qui poussent des cris d'orfraie en présentant des batteries de missiles (soit-disant "****anti attaques***

***aériennes**") autour des sites olympiques alors que ce sont des canons atmosmphériques pour dissiper les nuages...* www.planetargonautes.eu/2008/07/les-promesses-n.html

(113) *L'armée du Bélarus a constaté que des pièces de batteries de missiles **anti-aériens** S-300, de production russe, avaient été volées dans des garnisons du pays.* tf1.lci.fr/infos/monde/2005/0,,3202137,00-pieces-missiles-anti-aeriens-disparaissent-belarus-.html

4.5.9 Suffixation d'expressions nominales

Une stratégie plus exceptionnelle pour les dérivés dont les bases peuvent être exprimées par une séquence adjectif-nom est l'adjectivation de ces derniers par simple suffixation comme dans *grand vicaire* > *anti-grand vicarial* (114) ou *Grande Comore* > *anti-grand comorien* (115). Ces dérivés satisfont à la fois la contrainte de transparence catégorielle et celle de transparence de la base.

(114) *Ses attaques étaient principalement dirigées contre les anciens administrateurs de la province, qu'il avait si violemment insultés dans ses Lettres Violettes et Noires ou anti-épiscopales et **anti-grand vicariales**, touchant les administrations ...* clap.jac.free.fr/livre/bernardines.html

(115) *Cette ligne éditoriale est d'autant plus manifeste que le journal se fait constamment l'écho d'une conspiration **anti-grand comorienne**.* http://www.malango-actualite.com/kashkazi/archives/57-2006-10-12.pdf

4.5.10 Bases désadjectivales

Outre *antifroid* (106 ; 109), notre corpus contient plusieurs dérivés dont les radicaux sont des adjectifs simples (116a). Les bases de ces exemples peuvent être exprimées par des noms de propriété dérivés de ces adjectifs, comme en témoignent l'existence des formes en (116b).

(116) a. *antitriste* (4) ; *antistérile* (5) ; *antisec* (120)

b. *antistérilité* (118) ; *antitristesse* (117) ; *antisécheresse* (119)

(117) *Déjà à l'époque, on avait parlé pour ce médicament de "pilule du bonheur" ou de "molécule **anti-tristesse**" (cf. le Figaro du 27/06/85).* www.medcost.fr/html/economie_sante_eco/eco_110699d.htm

(118) *Les cacahuètes donnent la diarrhée, pas très aphrodisiaque ; toutefois, elles fournissent de la vitamine E, la vitamine **anti-stérilité**...* forum.doctissimo.fr/doctissimo/aphrodisiaques/cacahuetes-augmente-sperme-sujet_143244_1.htm

(119) *bon, le beurre, ça peut faire 1) un anti-froid pas cher pour les lèvres 2) un masque **anti-sècheresse** pas cher pour les cheveux...* fr.answers.yahoo.com/question/index?qid=20081110115241AAlEA7H

(120) *Et en plus j'utilise qq fois le baume seul comme un masque **anti-sec**.* forum.aufeminin.com/forum/beaute1/__f4825_beaute1-Existe-t-il-de-teintures-naturelles.html

Les exemples (4 ; 5 ; 120) n'ont en fait rien d'exceptionnel même s'ils sont ressentis comme plus marqués que les dérivés à radical nominal ou ceux dont le radical est un adjectif de relation dérivé comme *antiparlementaire*. Leurs radicaux sont des formes d'adjectifs suffisamment proches de leurs bases et le fait qu'ils soient les formes des bases de ces dernières n'a pas d'importance. La principale différence entre les dérivés (4 ; 5 ; 120) et ceux dont le radical est la forme d'un adjectif dérivé de la base est la distance morphologique légèrement plus grande pour les premiers (voir section 2.4). Le même type de relation se retrouve dans des dérivés dont la base est un nom de pays lui-même dérivé d'un nom de peuple comme *Russie* > *antirusse* ou *Turquie* > *antiturque* (Levkovych 2004, Plénat et al. 2004a, Roché 2005, 2008, 2009).

4.5.11 Périmètre du voisinage morphologique

La table 4.3 montre qu'à côté des dérivés dont les radicaux sont des formes d'adjectifs de relation relativement neutres obtenus par suffixation en *-al*, *-el*, *-ien*, *-ique*, *-esque* (121) ou de gentilés en *-ain*, *-ais*, *-an*, *-ois*, *-ote*, etc., on trouve des formes d'adjectifs sémantiquement plus éloignés de la base. Par exemple, *piratable* dans *antipiratable* (voir (23), page 47) qualifie le patient, i.e. l'objet du piratage ; *insomniaque* dans *anti-insomniaque* désigne la personne qui souffre d'insomnie ; *poussiéreux* dans *antipoussiéreux* désigne ce qui est recouvert de poussière. À l'inverse, dans *anticagoulard* (122), *anti-émeutier* ou *anti-musicien*, le radical est un nom qui qualifie l'agent, celui qui participe aux activités de la Cagoule, à une émeute ou qui joue de la musique. Ces radicaux sont tous suffisamment proches de leurs bases et suffisamment saillants pour permettre l'identification de ces dernières (contrainte C2.6).

(121) a. *Orient* > *anti-oriental*,
b. *ministère* > *antiministériel*
c. *Wagner* > *antiwagnérien*
d. *ergonomie* > *anti-ergonomique*
e. *titan* > *antititanesque*

(122) *... il fut nommé sous-chef de bureau au ministère de l'Intérieur et maintenu à ce poste par le ministre de l'Intérieur Roger SALENGRO, qui lui aurait confié la responsabilité du fichier central de la Sûreté nationale en raison de ses convictions républicaines et* ***anticagoulardes*** *(19).* `www.crdp-reims.fr/memoire/enseigner/rene_bousquet/01_carriere.htm`

Notre corpus fait également apparaître que l'essentiel des adjectifs non relationnels dont les formes sont utilisées comme radicaux servent à désigner des agents. La première raison qui explique cette préférence est que le sens du préfixe *anti-* décrit une opposition à une action (*antiblocage*) ou à l'agent qui la produit (*antibloquant*, *antibloqueur*). La seconde est que cette opposition est intentionnelle. L'utilisation d'un suffixe agentif permet de rapprocher l'entité désignée par le dérivé en *anti-* d'une série de noms d'agents : *antibloquant* est proche de la série des adjectifs en *-ant* et *antibloqueur* de celle des adjectifs en

-eur. On a ainsi une forme de réanalyse, ou plus exactement de superposition de deux séries : celle des dérivés en *anti-* au premier plan et celle des dérivés agentifs au second. Cette stratégie se retrouve dans les dérivés *antidéperlant* (89), *antidésherbant* (91) ou *antidémêlant* (92) présentés en section 4.3.6.

Par ailleurs, on observe une grande variabilité entre les locuteurs dans les voisinages d'un même mot, la saillance des positions lexicales qui l'entourent et les distances qui les séparent. Il existe par exemple au moins neuf réalisations possibles pour un dérivé en *anti-* dont le sens est 'qui réduit les vibrations', 'qui empêche de vibrer' (123)[20] et (124). Le dernier radical semble obtenu par conversion du verbe *vibrer* mais il pourrait être également analysé comme une forme apocopée de *vibration*.

(123) *antivibration* ; *antivibrant* ; *antivibratoire* ; *antivibratif* ; *antivibratile* ; *antivibrateur* ; *antivibreur* ; *antivibrable* ; *antivibre*

(124) a. *il existe des patins* ***anti-vibrations*** *pour machine à laver.* `astrosurf.com/astropratique/013324.html`

b. *Elles obligent également à monter le moteur sur des blocs* ***anti-vibrants*** *qui dissipent de l'énergie ...* `sycomoreen.free.fr/syco_francais/moteurs_pompes_problematique_archihybrid.html`

c. *Objet du marché : Table optique* ***anti-vibratoire****.* `www.espci.fr/mpa/appels/2008/a083fkp/index.html`

d. *Le faux plafond en placo phonique avec suspentes* ***anti vibratives*** *(assez cher quand même, ...* `http://www.systemed.fr/Forum/forum_posts.asp?TID=24488&get=last`

e. *Les plots* ***anti-vibratiles*** *fournis avec la baignoire ne doivent pas être déposés, ni modifiés, ni scellés.* `www.jacobdelafon.com/fr/pro/pdf/pose-balneo.pdf`

f. *Plots* ***anti-vibrateur*** *poids 150 Kg AAM108 (lot de 4).* `www.sem-angles.fr/climatisation/plots-anti-vibrateur-poids-150-kg-aam108-lot-de-4-A1540271p67S4R4c9216`

g. *... de nouvelles consignes de maintenance de leur dispositif* ***antivibreur*** *ont été édictées.* `www.senat.fr/questions/base/1993/qSEQ930501225.html`

h. *Toulouse. Theâtre LABEDA - 800 m2 plafond phonique avec plâtre Lafarges avec tampon* ***antivibrable*** *(plafond 3 couches)* `www.cyberbtp.com/pmp/page4/default.htm`

i. *je vais voir pour un rack* ***anti vibre*** *car dans mon boitier je peux mettre env 6 DD (ca devrait paser !)* `forum.zebulon.fr/aide-achat-hdd-raptor-t143169.html`

4.6 Conclusion

L'étude proposée dans ce chapitre a permis de répondre aux principales questions posées dans l'introduction. Elle a également permis d'atteindre d'autres objectifs. Le premier

20. Je rappelle que *antivibré* (13) est un participe construit sur un dérivé en *anti-* (voir section 4.2.1).

est de présenter un ensemble important de données originales permettant de renouveler l'analyse de la préfixation en *anti-*. Le deuxième est d'illustrer la rentabilité de la morphologie extensive et les gains qu'elle permet de réaliser, y compris sur des questions qui ont fait l'objet d'études nombreuses et détaillées. L'analyse proposée dans ce chapitre constitue également un exemple-type d'étude DUMALienne, à la fois sur le plan méthodologique et sur celui de l'approche pragmatique de la morphologie. La préfixation en *anti-* a ainsi été abordée sans *a priori* théorique, en utilisant toutes les notions et les outils qui permettent de démêler les conditionnements qui règlent cette construction. Le quatrième objectif est d'illustrer la souplesse et l'efficacité du modèle théorique présenté dans le chapitre 2.

Les grandes lignes de l'analyse de la préfixation en *anti-* qui vient d'être développée sont les suivantes :

1. *Anti-* est un préfixe. Les données ne permettent pas de conclure qu'il existe une préposition *anti*. Cependant, certains locuteurs utilisent marginalement *anti* comme une préposition synonyme de *contre*.
2. La sémantique des dérivés en *anti-* peut être décrite selon deux dimensions. La pre-

	endocentrique	**exocentrique** [+ adjectif]
spatial [– axiologique]	*Anti-Atlas* ; *antitragus* Y se trouve en face de X	*antimésentérique* ; *antisolaire* ; *anti-Terre* l'orientation Y est l'opposée de celle de X
logique [– axiologique]	*antihéros* ; *antiparticule* ; *anti-chaîne* ; *antibeau* ; *anti-obèse* ; *antimordre* Y est l'inverse de X ; Y est un X qui appartient au monde de l'antimatière ; Y est l'équivalent de X pour les antiparticules ; Y est la qualité d'être dépourvu de X, incompatible avec X ou contraire à X	
adversatif [+ axiologique]	*antidéperlant* Y est un synonyme de X	*antiparlementaire* ; *antilimace* ; *antidormir* Y est la qualité de s'opposer aux nuisances X ou aux nuisances causées par X

TABLE 4.5 – Récapitulation des interprétations des dérivés préfixés en *anti-*. Les cellules présentent quelques exemples représentatifs et une glose dans laquelle Y représente le dérivé et X sa base.

mière distingue les dérivés dont le sens est endocentrique (i.e. qui désignent des notions de même nature que leurs bases) de ceux dont le sens est exocentrique (i.e. qui désignent des notions de nature différente de celles de leurs bases). La seconde permet de séparer trois groupes de sens : spatiaux, logiques et adversatifs. Les sens de ces dérivés sont résumés dans le tableau 4.5. Ces sens ne peuvent pas être réunis sous un invariant unique mais ils sont connectés. Ils présentent une unité de famille (i.e. de type ressemblance de famille) et on peut passer de proche en proche d'un sens à l'autre.

3. Les formes des dérivés traditionnellement analysés comme parasynthétiques ou en faisant appel au principe de copie sont en réalité construites en empruntant leurs radicaux à des voisins morphologiques de leurs bases. Ces emprunts sont destinés à satisfaire la contrainte de transparence catégorielle.

Conclusion

J'ai présenté dans ce mémoire les grandes lignes de mon parcours scientifique et quelques-uns de mes chantiers en cours. Ma recherche se caractérise par le fait qu'elle se distribue dans deux domaines distincts de la linguistique : la morphologie et le traitement automatique des langues. Mes thématiques de recherche sont cependant relativement concentrées. Elles s'organisent autour de deux questions essentielles : la découverte de la structure morphologique du lexique et l'approche extensive en morphologie. Les chantiers que je viens de présenter délimitent un programme qui permet d'aborder un grand nombre de questions ouvertes sur la nature de la morphologie, ses relations avec la sémantique et la phonologie, ou ses interactions avec la syntaxe.

Le premier de mes chantiers est un modèle informatique de la morphologie dérivationnelle établi autour d'une mesure de la proximité morphologique et de l'analogie formelle. Ce modèle, radicalement lexématique, permet de réaliser des analyses morphologiques sans avoir recours aux morphèmes, aux affixes, aux exposants morphologiques ni à aucune représentation de ces notions. La mesure morphologique que j'ai proposée est calculée en représentant le lexique sous la forme d'un bigraphe et en l'explorant au moyen d'un algorithme basé sur les marches aléatoires. Elle est utilisée pour identifier des voisinages morphologiques mais elle pourrait aussi servir à apparier les lexiques de langues proches comme le français et l'occitan ou de différents états d'une même langue (à partir, par exemple, de textes originaux en ancien ou en moyen français et d'éditions modernes des mêmes œuvres). L'analogie formelle permet d'identifier des quadruplets de mots que l'on peut ensuite étendre en familles et en séries morphologiques. Les premiers résultats obtenus sont encourageants. Ce modèle constitue un changement de paradigme pour l'analyse morphologique automatique dont la finalité n'est plus de découvrir la structure morphologique des mots mais celle du lexique dans son entier. Ce modèle informatique est cependant loin d'être complet. Il reste à séparer les familles des séries, à améliorer la qualité des traits sémantiques en procédant à l'analyse des définitions à partir desquelles ils sont extraits, à proposer un traitement permettant de découvrir les allomorphies, à traiter les conversions et les compositions, à intégrer des informations relatives à la fréquence des mots et à leur distribution, à utiliser des formes fléchies, etc. L'amélioration des descriptions sémantiques

passera aussi par la sélection, parmi ces traits, des meilleurs prédicteurs de la proximité morphologique.

Ce premier chantier devrait permettre de créer à court terme une base de données morphologiques à large couverture en utilisant une méthode de *bootstrap*. La réalisation de la maquette d'un système de création de bases de données d'exemples morphologiques devrait suivre rapidement. Il s'agit d'intégrer la base de données morphologiques dans un système Wiki permettant à ses utilisateurs d'enregistrer et d'exploiter leurs exemples au fur et à mesure de leur collecte et de leur analyse. À moyen terme, je souhaiterais adapter la mesure à d'autres types de langues et notamment aux langues gabaritiques comme l'arabe ou l'hébreu et aux langues agglutinantes comme le finnois ou le turc. La comparaison des structures morphologiques des langues est aussi une question sur laquelle je souhaite me pencher. Une autre application possible de ce travail est l'aide à la révision des définitions des dérivés morphologiques dans les dictionnaires. On pourra en effet définir une mesure de la cohérence et de l'homogénéité de ces définitions. Cette mesure permettra d'identifier les définitions qu'il faut réviser et de vérifier l'amélioration apportée par leur révision.

Mon deuxième chantier est la description morphologique dans le cadre de l'approche extensive proposée par Marc Plénat (2000). J'ai présenté dans le chapitre 4 une étude de la préfixation en *anti-* réalisée en utilisant un corpus de 3 000 exemples collectés en grande partie sur le Web. La sémantique des dérivés en *anti-* y est analysée en fonction de deux critères, l'alternance des interprétations endocentriques et exocentriques et les types d'emplois, spatiaux, logiques et adversatifs. La caractérisation de ces derniers est affinée davantage en s'appuyant sur la catégorie de la dynamique des forces de Talmy (2000). Selon l'analyse proposée, les dérivés décrivent la qualité de s'opposer à la base vue comme une nuisance ou une source de nuisances aux intérêts d'un agent qui n'est pas nommé. Le chapitre présente un nombre important d'exemples et montre comment la forme des dérivés parasynthétiques découle de contraintes de recyclage et de transparence qui portent sur l'organisation morphologique du lexique. Ces contraintes sont décrites dans le chapitre 2. J'entends poursuivre ce travail de description de la sémantique des procédés morphologiques, notamment en reprenant l'analyse de la suffixation en *-able* puis en m'intéressant à celles en *-eur* et en *-if*. L'analyse que j'envisage pour *-able* est esquissée en section 1.3.3. Je prévois de travailler également sur la conversion, phénomène qui, comme les constructions parasynthétiques, nous éclaire sur certains types de déformations dans les correspondances qui s'établissent entre les niveaux sémantique, lexical et formel décrits en section 2.3.

Le troisième chantier est le plus récent. Il vise à la définition d'un modèle théorique de la morphologie dérivationnelle qui permette d'articuler les notions de proximité et de paradigmes utilisées dans le modèle informatique et les analyses linguistiques formulées en termes de contraintes lexicales concurrentes. Cette plate-forme est pour l'instant une ébauche dont un grand nombre d'éléments devront être précisés au fur et à mesure du développement du modèle informatique et de la description de nouveaux procédés. Parmi les questions en suspens, il y a le contenu des positions, la représentation des paradigmes, l'articulation des paradigmes flexionnels et dérivationnels et la nature des objets sémantiques mis en correspondance avec les positions lexicales. Un autre point qui reste à expliciter est la place et la représentation des formes et des sens mémorisés. La manière dont ces

représentations interviennent dans la construction et l'analyse des mots devra également être définie. En d'autres termes, il s'agit de préciser la place du lexique mémorisé dans le lexique mental. L'acquisition de la morphologie est une autre voie de recherche ouverte par ce modèle théorique. Le fait que la structure morphologique du lexique émerge à partir des données auxquelles les locuteurs sont exposés a notamment comme conséquence que celle-ci ne peut se mettre en place que lorsque le nombre des mots mémorisés est suffisamment important (Calderone et al. 2007).

Bibliographie

Amiot, D. (2004). Préfixes ou prépositions ? Le cas de *sur*(-), *sans*(-), *contre*(-) et les autres. *Lexique*, **16**:67–83.

Amiot, D. (2008). La catégorie de la base dans la préfixation en *dé-*. In B. Fradin, éd., *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, pp. 1–15. John Benjamins Publishing Company, Amsterdam / Philadelphia.

Amiot, D. et F. Montermini (2009). Affixes et mots grammaticaux. In B. Fradin, F. Kerleroux, et M. Plénat, éds., *Aperçus de morphologie du français*, pp. 127–141. Presses universitaires de Vincennes, Saint-Denis.

Amsili, P. et N. Hathout (1996). Computational semantics of time/negation interaction. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*. Copenhagen.

Amsili, P. et N. Hathout (1998). Systèmes de types pour la (λ-)DRT ascendante. In *Actes de la 5e conférence annuelle sur le traitement automatique des langues naturelles (TALN-98)*, pp. 92–101. Paris.

Anderson, S. R. (1992). *A-Morphous Morphology*. Cambridge University Press, Cambridge.

Aronoff, M. (1976). *Word Formation in Generative Grammar*. Linguistic Inquiry Monographs. MIT Press, Cambridge, Mass.

Aronoff, M. (1994). *Morphology by Itself. Stem and Inflectional Classes*. Linguistic Inquiry Monographs. MIT Press, Cambridge, Mass.

Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Kluwer Academic Publisher, Dordrecht.

Aurnague, M. et M. Plénat (2008). Sémantique de l'espace et morphologie : le cas de la préfixation en é-. *Bulletin de la Société de Linguistique de Paris*, **CIII**(1):201–236.

Baayen, H. (1992). Quantitative aspects of morphological productivity. In G. E. Booij et J. van Marle, éds., *Yearbook of Morphology 1991*, pp. 109–149. Kluwer Academic Publishers, Dordrecht.

Baayen, R. H., R. Piepenbrock, et L. Gulikers (1995). The CELEX lexical database (release 2). CD-ROM. Linguistic Data Consortium, Philadelphia, Penn.

Baroni, M., J. Matiasek, et H. Trost (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In *Proceedings of the Workshop on Morphological and Phonological Learning of ACL-2002*, pp. 48–57. ACL, Philadelphia, Penn.

Becker, T. (1994). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In G. E. Booij et J. van Marle, éds., *Yearbook of Morphology 1993*, pp. 1–25. Kluwer Academic Publishers, Dordrecht.

Benamara, F., N. Hathout, P. Muller, et S. Ozdowska, éds. (2007). *Actes de la 14e conférence sur le traitement automatique des langues naturelles (TALN-2007) et de la 11e rencontre des étudiants chercheurs en informatique pour le traitement automatique des langues (RECITAL-2007)*. ATALA, CLLE-ERSS & IRIT, Toulouse.

Bernhard, D. (2006a). *Apprentissage de connaissances morphologiques pour l'acquisition automatique de ressources lexicales*. Thèse de doctorat, Université Joseph Fourier, Grenoble.

Bernhard, D. (2006b). Automatic acquisition of semantic relationships from morphological relatedness. In *Advances in Natural Language Processing, Proceedings of the 5th International Conference on NLP, FinTAL 2006*, volume 4139 de *Lecture Notes in Computer Science*, pp. 121–132. Springer Verlag, Berlin / Heidelberg.

Berrendonner, A. (1990). Attracteurs. *Cahiers de linguistique française*, **11**:149–158. Actes du 4e colloque de pragmatique de Genève.

Berrendonner, A. (1995). Quelques notions utiles à la sémantique des descripteurs nominaux. *Travaux neuchâtelois de linguistique (TRANEL)*, **23**(9-39).

Berrendonner, A. (2002). Types. In M. Carel, éd., *Les facettes du dire : hommage à Oswald Ducrot*, pp. 39–53. Kimé, Paris.

Bertram, R., M. Laine, et K. Karvinen (1999). The interplay of word formation type, affixal homonymy, and productivity in lexical processing: Evidence from a morphologically rich language. *Journal of Psycholinguistic Research*, **28**(3):213–226.

Bertram, R., R. Schreuder, et R. H. Baayen (2000). The balance of storage and computation in morphological processing: The role of word formation type, affixal homonymy, and productivity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26**(2):489–511.

Blevins, J. P. (2003). Stems and paradigms. *Language*, **79**(4):737–767.

Blevins, J. P. (2006). Word-based morphology. *Journal of Linguistics*, **42**(3):531–573.

Bonami, O. et G. Boyé (2003). Supplétion et classes flexionnelles dans la conjugaison du français. *Langages*, **152**:102–126.

Bonami, O., G. Boyé, et F. Kerleroux (2009). L'allomorphie radicale et la relation flexion-construction. In B. Fradin, F. Kerleroux, et M. Plénat, éds., *Aperçus de morphologie du français*, pp. 103–125. Presses universitaires de Vincennes, Saint-Denis.

Booij, G. E. (1996). Inherent versus contextual inflection and the split morphology hypothesis. In G. E. Booij et J. van Marle, éds., *Yearbook of Morphology 1995*, pp. 1–16. Kluwer Academic Publishers, Dordrecht.

Booij, G. E. (2002). *The Morphology of Dutch*. Oxford University Press, Oxford.

Booij, G. E. (2007). Construction morphology and the lexicon. In F. Montermini, G. Boyé, et N. Hathout, éds., *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, pp. 34–44. Cascadilla Proceedings Project, Somerville, MA.

Booij, G. E. (2008). Paradigmatic morphology. In B. Fradin, éd., *La raison morphologique. Hommage à la mémoire de Danielle Corbin*, pp. 29–38. John Benjamins Publishing Company, Amsterdam / Philadelphia.

Boullier, P., L. Clément, B. Sagot, et E. Villemonte de La Clergerie (2005). Chaînes de traitement syntaxique. In *Actes de la 12e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2005)*, pp. 103–112. ATALA, Dourdan.

Bourigault, D. (1994). *LEXTER, un Logiciel d'EXtraction de TERminologie. Application à l'acquisition de connaissances à partir de textes*. Thèse de doctorat, École des Hautes Études en Sciences Sociales, Paris.

Bourigault, D. (2007). *Un analyseur syntaxique opérationnel : SYNTEX*. Habilitation à diriger des recherches, Université Toulouse II-Le Mirail, Toulouse.

Bourigault, D. et C. Fabre (2000). Approche linguistique pour l'analyse syntaxique des corpus. *Cahiers de grammaire*, **25**:131–151.

Boyé, G. (2010). Régularité et classes flexionnelles dans la conjugaison du français. In M. Roché, G. Boyé, N. Hathout, S. Lignon, et M. Plénat, *Des unités morphologiques au lexique*. Manuscrit.

Burzio, L. (2002). Surface-to-surface morphology: when your representations turn into constraints. In P. Boucher, éd., *Many Morphologies*, p. 142–177. Cascadilla Press, Somerville, Mass.

Bybee, J. L. (1985). *Morphology. A Study of the Relation between Meaning and Form*, volume 9 de *Typological Studies in Language*. John Benjamins Publishing Company, Amsterdam.

Bybee, J. L. (1988). Morphology as lexical organization. In M. Hammond et M. Noonan, éds., *Theoretical Morphology. Approaches in Modern Linguistics*, pp. 119–141. Academic Press, San Diego, CA.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and cognitive processes*, **10**(5):425–455.

Béchet, F. (2001). Lia_phon : un système complet de phonétisation de textes. *Traitement automatique des langues*, **42**(1):47–67.

Calderone, B., I. Herreros, et V. Pirrelli (2007). Learning inflection: The importance of starting big. *Lingue & Linguaggio*, **2**:175–200.

Chomsky, N. (1981). *Lectures on Government and Binding*, volume 9 de *Studies in Generative Grammar*. Foris, Dordrecht.

Chomsky, N. (1986). *Barriers*, volume 13 de *Linguistic Inquiry Monographs*. MIT Press, Cambridge, Mass.

Claveau, V. et M.-C. L'Homme (2005). Apprentissage par analogie pour la structuration de terminologie. Utilisation comparée de ressources endogènes et exogènes. In *Actes de la conférence terminologie et intelligence artificielle (TIA-2005)*. Rouen.

Colmerauer, A. (1990). An introduction to Prolog III. *Communications of the ACM*, **33**(7):69–90.

Corbin, D. (1980). Contradictions et inadéquation de l'analyse parasynthétique en morphologie dérivationnelle. In A.-M. Dessaux-Berthonneau, éd., *Linguistiques et traditions grammaticales*, pp. 181–224. Presses universitaires de Lille, Villeneuve-d'Asq.

Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*. Max Niemeyer Verlag, Tübingen.

Corbin, D. (1991). Introduction. La formation des mots : structure et interprétation. *Lexique*, **10**:7–30.

Corbin, D. (1997). La représentation d'une « famille » de mots dans le *Dictionnaire dérivationnel du français* et ses corrélats théoriques, méthodologiques et descriptifs. *Recherches linguistiques de Vincennes*, **26**:5–38.

Corbin, D. (2001a). Du nouveau sur *beurre laitier*. Note sur une fausse conversion. In C. Buridant, G. Kleiber, et J.-C. Pellat, éds., *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*, volume 45 de *Bibliothèque de l'information grammaticale*, pp. 127–143. Éditions Peeters, Leuven.

Corbin, D. (2001b). Préfixe et suffixes : du sens aux catégories. *Journal of French Language Studies*, **11**(1):41–69.

Corbin, D. et M. Temple (1994). Le monde des mots et des sens construits : catégories sémantiques, catégories référentielles. *Cahiers de lexicologie*, **65**(2):5–28.

Creutz, M. et K. Lagus (2002). Unsupervised discovery of morphemes. In *Proceedings of the ACL Workshop on Morphological and Phonological Learning*, pp. 21–30. ACL, Philadelphia, Penn.

Creutz, M. et K. Lagus (2005). *Unsupervised Morpheme Segmentation and Morphology Induction from Text Corpora Using Morfessor 1.0*. Rapport technique A81, Helsinki University of Technology.

Dal, G. (2003a). Analogie et lexique construit : La preuve par les textes. *Cahiers de grammaire*, **28**.

Dal, G. (2003b). Arguments pour un préfixe *contre-*. *Recherches linguistiques*, **26**:173–201.

Dal, G. (2007). Les adverbes de manière en *-ment* du français : Dérivation ou flexion. In N. Hathout et F. Montermini, éds., *Morphologie à Toulouse*, numéro 37 in LINCOM Studies in Theoretical Linguistics, pp. 121–147. Lincom Europa, Munich.

Dal, G. (2008). Analogie et lexique construit : Un retour ? In J. Durand, B. Habert, et B. Laks, éds., *Actes du Congrès Mondial de Linguistique Française (CMLF-2008)*, pp. 1587–1599. ILF, Paris.

Dal, G., N. Hathout, et F. Namer (1999). Construire un lexique dérivationnel : Théorie et réalisation. In P. Amsili, éd., *Actes de la 6^e conférence sur le traitement automatique des langues naturelles (TALN-99)*, pp. 115–124. ATALA, Cargèse.

Dal, G., N. Hathout, et F. Namer (2004). Morphologie constructionnelle et traitement automatique des langues : Le projet MorTAL. *Lexique*, **16**:199–229.

Dal, G. et F. Namer (2005). L'exception infirme-t-elle la notion de règle ? Ou le lexique construit et la théorie de l'optimalité. *Faits de langue*, **25**:123–130.

Darmesteter, A. (1877). *De la création actuelle de mots nouveaux dans la langue française*. Vieweg, Paris.

Davis, E. (1988). A logical framework for commonsense predictions of solid objects behavior. *Artificial Intelligence in Engineering*, **3**(3):125–140.

De Mulder, W. et D. Amiot (2003). Préposition *contre* préfixe. *Recherches linguistiques*, **26**:203–232.

Dell, F. (1970). *Les règles phonologiques tardives et la morphologie dérivationnelle du français*. Thèse de PhD, MIT.

Dell, F. (1995). Consonant clusters and phonological syllabes in French. *Lingua*, **95**:5–26.

Durand, J. (1982). A propos du préfixe *anti-* et de la parasynthèse en français. In *Occasional Papers from the Language Center*, volume 25, pp. 1–35. University of Essex.

Durand, J. (2009). On the scope of linguistics: data, intuitions, corpora. In Y. Kawaguchi, M. Minegishi, et J. Durand, éds., *Corpus and Variation in Linguistic Description and Language Education*, pp. 25–52. John Benjamins Publishing Company, Amsterdam/Philadelphia.

Durand, J., B. Laks, et C. Lyche (2002). La phonologie du français contemporain : usages, variétés et structure. In C. D. Pusch et W. Raible, éds., *Romanistische Korpuslinguistik – Korpora und gesprochene Sprache / Romance Corpus Linguistics – Corpora and Spoken Language*, pp. 93–106. Gunter Narr Verlag, Tübingen.

Durand, J., B. Laks, et C. Lyche (2005). Un corpus numérisé pour la phonologie du français. In G. Williams, éd., *La linguistique de corpus*, pp. 205–217. Presses universitaires de Rennes, Rennes.

Déjean, H. (1998a). *Concepts et algorithmes pour la découverte des structures formelles des langues*. Thèse de doctorat, Université de Caen.

Déjean, H. (1998b). Morphemes as necessary concept for structures discovery from untagged corpora. In *Proceedings of the Workshop on Paradigms and Grounding in Natural Language Learning*, pp. 295–299. Adelaide.

Fabre, C., F. Floricic, et N. Hathout (2004). Collecte outillée pour l'analyse des emplois discordants des déverbaux en *-eur*. Communication aux journées d'étude sur *La place des méthodes quantitatives dans le travail du linguiste*. ERSS, Université de Toulouse II-Le Mirail.

Fellbaum, C., éd. (1999). *WordNet: an Electronic Lexical Database*. MIT Press, Cambridge, Mass.

Fradin, B. (1996). On the morphological entities and the copy principle. *Acta Linguistica Hungarica*, **43**(1–2):111–151.

Fradin, B. (1997a). Esquisse d'une sémantique de la préfixation en *anti-*. *Recherches linguistiques de Vincennes*, **26**:87–112.

Fradin, B. (1997b). Une préfixation complexe : le cas de *anti-*. *Neuphilologische Mitteilungen*, **XCVIII**(4):333–349.

Fradin, B., G. Dal, N. Hathout, F. Kerleroux, M. Plénat, et M. Roché, éds. (2003a). *Les unités morphologiques*, volume 3 de *Silexicales*. Silex, Université de Lille 3, Villeneuve-d'Asq.

Fradin, B., N. Hathout, et F. Meunier (2003b). La suffixation en *-ET* et la question de la productivité. *Langue Française*, **140**:56–78.

Fradin, B., F. Kerleroux, et M. Plénat, éds. (2009). *Aperçus de morphologie du français*. Presses universitaires de Vincennes, Saint-Denis.

Galy, E. et D. Bourigault (2005). Analyse distributionnelle de corpus de langue générale et synonymie. In G. Williams, éd., *Texte et corpus : Actes des 4es journées de la linguistique de corpus*, pp. 163–174. Lorient.

Gaume, B. (2003). Analogie et proxémie dans les réseaux petits mondes. *Revue d'Intelligence Artificielle*, **17**(5/6):935–951. « Regards croisés sur l'analogie ».

Gaume, B. (2004). Balades aléatoires dans les petits mondes lexicaux. *I3. Information - Interaction - Intelligence*, **4**(2).

Gaume, B., K. Duvigneau, O. Gasquet, et M.-D. Gineste (2002). Forms of meaning, meaning of forms. *Journal of Experimental and Theoretical Artificial Intelligence*, **14**(1):61–74.

Gaume, B., N. Hathout, et P. Muller (2004a). Désambiguisation par proximité structurelle. In *Actes de la 11e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2004)*, pp. 205–214. ATALA, Fez.

Gaume, B., N. Hathout, et P. Muller (2004b). Word sense disambiguation using a dictionary for sense similarity measure. In *Proceedings of the 20th International Conference on Computational Linguistics (Coling 2004)*, volume II, pp. 1194–1200. ACL, Genève.

Gaume, B., F. Venant, et B. Victorri (2006). Hierarchy in lexical organization of natural language. In D. Pumain, éd., *Hierarchy in natural and social sciences*, Methodos series, pp. 121–143. Springer, Dordrecht.

Gaussier, E. (1999). Unsupervised learning of derivational morphology from inflectional lexicons. In *Proceedings of the ACL Workshop on Unsupervised Learning in Natural Language Processing*. College Park, MD.

Gawelko, M. (1975). *Évolution des suffixes adjectivaux en français*. Polska Akademia Nauk Komitet Neofilologiczny, Wrocław.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, **7**(2):155–170.

Goldsmith, J. (2001). Unsupervised learning of the morphology of natural language. *Computational Linguistics*, **27**(2):153–198.

Grabar, N., G. Dal, B. Fradin, N. Hathout, S. Lignon, F. Namer, C. Plancq, D. Tribout, F. Yvon, et P. Zweigenbaum (2006a). Productivité quantitative de la suffixation par *-Able* dans un corpus journalistique du français. In *Actes des 8es Journées internationales d'analyse statistique des données textuelles (JADT 2006)*, pp. 473–485. Presses universitaires de Franche-Comté, Besançon.

Grabar, N., G. Dal, B. Fradin, N. Hathout, S. Lignon, F. Namer, C. Plancq, D. Tribout, F. Yvon, et P. Zweigenbaum (2006b). Productivité quantitative des suffixation par *-ité* et *-Able* dans un corpus journalistique du français. In *Actes de la 13e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2006)*, pp. 167–175. ATALA, Presses universitaires de Louvain, Louvain.

Grabar, N. et P. Zweigenbaum (1999). Acquisition automatique de connaissances morphologiques sur le vocabulaire médical. In *Actes de la 6e conférence annuelle sur le traitement automatique des langues naturelles (TALN-99)*, pp. 175–184. ATALA, Cargèse.

Gruaz, C. (1997a). Catégories lexématiques et catégories morphémiques dans les familles synchroniques dérivationnelles du français. *Recherches linguistiques de Vincennes*, **26**:39–59.

Gruaz, C. (1997b). La stratification dérivationnelle dans les familles synchroniques des mots français contemporains. In *III Coloquio Internacional de Lingüística Francesa*. Salamanca.

Harris, Z. (1955). From phoneme to morpheme. *Language*, **31**(2):190–222.

Hathout, N. (1996a). Les marqueurs superficiels dans les définitions spécialisées du *Trésor de la Langue Française*. In *Actes de colloque Informatique et Langue Naturelle (ILN-96)*. Nantes.

Hathout, N. (1996b). Pour la construction d'une base de connaissances lexicologiques à partir du *Trésor de la Langue Française*. Les marqueurs superficiels dans les définitions spécialisées. *Cahiers de Lexicologie*, **68**(1):137–173.

Hathout, N. (2000). Morphological pairing based on the Network Model. In *Proceedings of the Workshop on Computational Lexicography and Multimedia Dictionaries (COMLEX 2000)*, pp. 35–38. Pyrgos.

Hathout, N. (2001). Analogies morpho-synonymiques. Une méthode d'acquisition automatique de liens morphologiques à partir d'un dictionnaire de synonymes. In D. Maurel, éd., *Actes de la 8e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2001)*, pp. 223–232. ATALA, Tours.

Hathout, N. (2002). From WordNet to CELEX: Acquiring morphological links from dictionaries of synonyms. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pp. 1478–1484. ELRA, Las Palmas de Gran Canaria.

Hathout, N. (2003). L'analogie, un moyen de croiser les contraintes et les paradigmes. Acquisition de connaissances morphologiques à partir de dictionnaires de synonymes. *Revue d'Intelligence Artificielle*, **17**(5-6):923–934.

Hathout, N. (2005). Exploiter la structure analogique du lexique construit : Une approche computationnelle. *Cahiers de lexicologie*, **87**(2):5–28.

Hathout, N. (2008). Acquisition of the morphological structure of the lexicon based on lexical similarity and formal analogy. In *Proceedings of the Coling workshop Textgraphs-3*, pp. 1–8. ACL, Manchester.

Hathout, N. (2009a). Acquisition morphologique à partir d'un dictionnaire informatisé. In *Actes de la 16e conférence sur le traitement automatique des langues naturelles (TALN-2009)*. ATALA, Senlis.

Hathout, N. (2009b). Acquisition of morphological families and derivational series from a machine readable dictionary. In F. Montermini, G. Boyé, et J. Tseng, éds., *Selected Proceedings of the 6th Décembrettes: Morphology in Bordeaux*. Cascadilla Proceedings Project, Somerville, Mass.

Hathout, N. et C. Fabre (2002). Constitution et exploitation de lexiques de formes déverbales. Communication aux *Journées d'études sur les noms déverbaux*. Silex, Université Lille 3, Villeneuve-d'Asq.

Hathout, N. et F. Montermini, éds. (2007). *Morphologie à Toulouse. Actes du colloque international de morphologie 4es Décembrettes*, Lincom Studies in Theoretical Linguistics 37. Lincom Europa, Munich.

Hathout, N., F. Montermini, et L. Tanguy (2008). Extensive data for morphology: Using the World Wide Web. *Journal of French Language Studies*, **18**(1):67–85.

Hathout, N. et F. Namer (1997a). Génération (semi)-automatique de ressources lexicales réutilisables à grande échelle. Conversion des tables du LADL. In *Actes des 1ères JST FRANCIL*. AUPELF-UREF, Avignon.

Hathout, N. et F. Namer (1997b). (Semi-)automatic generation of ALEP analysis lexicon. In *Proceedings of the 3rd ALEP User Group Workshop*. Saarbrücken.

Hathout, N. et F. Namer (1998). Automatic construction and validation of French large lexical resources: Reuse of verb theoretical linguistic descriptions. In *Proceedings of the First International Conference on Language Resources and Evaluation*, pp. 627–636. ELRA, Granada.

Hathout, N., F. Namer, et G. Dal (2002). An experimental constructional database: The MorTAL project. In P. Boucher, éd., *Many Morphologies*, pp. 178–209. Cascadilla, Somerville, Mass.

Hathout, N., F. Namer, M. Plénat, et L. Tanguy (2009a). La collecte et l'utilisation des données en morphologie. In B. Fradin, F. Kerleroux, et M. Plénat, éds., *Aperçus de morphologie du français*, pp. 267–287. Presses universitaires de Vincennes, Saint-Denis.

Hathout, N., M. Plénat, et L. Tanguy (2003). Enquête sur les dérivés en *-able*. *Cahiers de grammaire*, **28**:49–90.

Hathout, N. et M. Roché, éds. (2003). *Cahiers de grammaire*, **28**. ERSS, Toulouse. « Morphologie et lexique ».

Hathout, N., F. Sajous, et L. Tanguy (2009b). Looking for French deverbal nouns in an evolving Web (a short history of WAC). In *Proceedings of Web as Corpus 2009 (WAC5)*. San Sebastian.

Hathout, N. et L. Tanguy (2002a). Autodétection des Webnéologismes. Communication au *Colloque TALN, Corpus et Web*. Maison des Sciences de l'Homme, Saint-Denis.

Hathout, N. et L. Tanguy (2002b). Webaffix : Finding and validating morphological links on the WWW. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pp. 1799–1804. ELRA, Las Palmas de Gran Canaria.

Hathout, N. et L. Tanguy (2006). Webaffix : Une boîte à outils d'acquisition lexicale à partir du Web. *Revue québéquoise de linguistique*, **32**(1):61–84.

Hay, J. B. (2000). *Causes and Consequences of Word Structure*. Thèse de PhD, Northwestern University, Evanston, IL.

Hay, J. B. et R. H. Baayen (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences*, **9**(7):342–348.

Hayes, P. J. (1985). Naive physics I: Ontology for liquids. In J. R. Hobbs et R. C. Moore, éds., *Formal Theories of the Commonsense World*. Ablex Publishing Corporation, Norwood, NJ.

Herskovits, A. (1982). *Space and the Prepositions in English: Regularities and Irregularities in a Complex Domain*. Thèse de PhD, Stanford University, Stanford, CA.

Heyna, F. (2008). Sémantisme et potentiel argumentatif des dérivés dénominaux en *anti-*. *Discours*, **2**. `discours.revues.org/index2022.html`.

Jackendoff, R. (1975). Morphological and semantic regularities in the lexicon. *Language*, **51**(3):639–671.

Jacquemin, C. (1997). Guessing morphology from terms and corpora. In *Proceedings of 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-97)*, pp. 156–167. ACM, Philadelphia, PA.

Jacquemin, C. et E. Tzoukermann (1999). NLP for term variant extraction: Synergy between morphology, lexicon, ans syntax. In T. Strzalkowski, éd., *Natural Language Information Retrieval*, pp. 25–74. Kluwer Academic Publishers, Dordrecht.

Jaffar, J. et J.-L. Lassez (1987). Constraint logic programming. In *Proceedings of the Fourteenth ACM Symposium of the Principles of Programming Languages*, pp. 111–119. Munich.

Jaffar, J. et M. J. Maher (1994). Constraint logic programming: A survey. *Journal of Logic Programming*, **19 & 20**:503–581.

Jing, H. et E. Tzoukermann (1999). Information retrieval based on context distance and morphology. In *Proceedings of 22th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR-99)*, pp. 90–96. ACM, Berkeley, CA.

Jurafsky, D. et J. H. Martin (2000). *Speech and language processing*. Prentice-Hall, Upper Saddle River, NJ.

Kamp, H. (1981). A theory of truth and semantics representation. In J. Groenendijk, T. Jansen, et M. Stokhof, éds., *Formal Methods in the Study of Language*. Mathematical Centre Tracts, Amsterdam.

Kerleroux, F. (2007). Pêcheurs sous la glace. In E. Delais-Roussarie et L. Labrune, éds., *Des sons et des sens : données et modèles en phonologie et en morphologie*, pp. 141–154. Hermes-Lavoisier, Paris.

Kohonen, T. (1995). *Self-Organizing Maps*. Springer Verlag, Berlin / Heidelberg.

Koskiennemi, K. (1983). Two-level model for morphological analysis. In *Proceedings of 8th IJCAI*, pp. 683–685. Karlsruhe.

Kurimo, M., M. Creutz, et M. Varjokallio (2007). Unsupervised morpheme analysis evaluation by a comparison to a linguistic gold standard – Morpho Challenge 2007. In *Working Notes for the CLEF 2007 Workshop*. `clef.isti.cnr.it/2007/working_notes/kurimo_varjokallioCLEF2007_MC1.pdf`.

Kurimo, M. et M. Varjokallio (2008). Unsupervised morpheme analysis evaluation by a comparison to a linguistic gold standard – Morpho Challenge 2008. In *Working Notes for the CLEF 2008 Workshop*. `www.clef-campaign.org/2008/working_notes/kurimo1-paperCLEF2008.pdf`.

Langlais, P. (2009). Étude quantitative de liens entre l'analogie formelle et la morphologie constructionnelle. In *Actes de la 16e conférence sur le traitement automatique des langues naturelles (TALN-2009)*. ATALA, Senlis.

Langlais, P. et A. Patry (2007). Translating unknown words by analogical learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and the Conference on Computational Natural Language Learning, EMNLP-CoNLL 2007*, pp. 877–886. ACL, Prague.

Langlais, P. et A. Patry (2008). Enrichissement d'un lexique bilingue par apprentissage analogique. *Traitement automatique des langues*, **49**(1):13–40.

Langlais, P. et F. Yvon (2008). Scaling up analogical learning. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, p. 51–54. Manchester.

Langlais, P., F. Yvon, et P. Zweigenbaum (2009). Improvements in analogical learning: Application to translating multi-terms of the medical domain. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL-09)*, p. 487–495. Athène.

Lebart, L. et A. Salem (1994). *Statistique textuelle*. Dunod, Paris.

Lebart, L., A. Salem, et L. Berry (1998). *Exploring textual data*. Kluwer Academic Publishers, Dordrecht.

Lepage, Y. (1998). Solving analogies on words: An algorithm. In *Proceedings of COLING-ACL-98*, volume 2, pp. 728–735. Montréal.

Lepage, Y. (2000). Languages of analogical strings. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING-2000)*, pp. 488–494. ACL, Saarbrücken.

Lepage, Y. (2003). *De l'analogie rendant compte de la commutation en linguistique*. Habilitation à diriger des recherches, Université Joseph Fourier, Grenoble.

Levkovych, O. (2004). *Les mots français en* anti-. Mémoire de DEA, Université de Toulouse II-Le Mirail, Toulouse.

Lignon, S. (2000). *La suffixation en* -ien*. Aspects sémantiques et phonologiques*. Thèse de doctorat, Université de Toulouse II-Le Mirail, Toulouse.

Lignon, S. (2010). Entre histoire et morphophonologie, quelle distribution pour *-éen* vs *-ien*. In M. Roché, G. Boyé, N. Hathout, S. Lignon, et M. Plénat, *Des unités morphologiques au lexique*. Manuscrit.

Lignon, S. et M. Plénat (2009). Échangisme suffixal et contraintes phonologiques. In B. Fradin, F. Kerleroux, et M. Plénat, éds., *Aperçus de morphologie du français*, pp. 65–81. Presses universitaires de Vincennes, Saint-Denis.

Lombard, A. (1964). *Le rôle des semi-voyelles et leur concurrence avec les voyelles correspondantes dans la prononciation parisienne*. CWK Gleerup, Lund.

McCarthy, J. et A. Prince (1993). *Prosodic Morphology: Constraint Interaction and Satisfaction*. Technical report 3, Rutgers University Center for Cognitive Science, New Brunswick, NJ.

Meunier, F., G. Dal, B. Fradin, et N. Hathout (2007). Rôle de la productivité des procédés morphologiques dans la reconnaissance de lexèmes complexes. In N. Hathout et F. Montermini, éds., *Morphologie à Toulouse. Actes du colloque international de morphologie 4es Décembrettes*, numéro 37 in LINCOM Studies in Theoretical Linguistics, pp. 291–312. Lincom Europa, Munich.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, et K. J. Miller (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, **3**(4):335–391.

Montermini, F. (2002). *Le système préfixal en italien comtemporain*. Thèse de doctorat, Université de Paris X - Nanterre & Università degli Studi di Bologna, Paris / Bologne.

Montermini, F. (2008). *Il lato sinistro della morfologia. La prefissazione in italiano e nelle lingue del mondo*. FrancoAngeli, Milan.

Montermini, F., G. Boyé, et N. Hathout, éds. (2007). *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, Cascadilla Proceedings Project. Cascadilla Press, Somerville, MA.

Moreau, F., V. Claveau, et P. Sébillot (2007). Automatic morphological query expansion using analogy-based machine learning. In *Proceedings of the 29th European Conference on Information Retrieval (ECIR 2007)*, pp. 222–233. Rome.

Morin, Y.-C. (1971). *Computer experiments in generative phonology, low-level French phonology*. Natural language studies 11, Department of Computer and Communication Science, Universty of Michigan, Ann Arbor, MI.

Muller, P., N. Hathout, et B. Gaume (2006). Synonym extraction using a semantic distance on a dictionary. In D. Radev et R. Mihalcea, éds., *Proceedings of the HLT/NAACL Workshop Textgraphs*, pp. 65–72. ACL, New York, NY.

Namer, F. (2003a). Acquisition automatique de sens à partir d'opérations morphologiques en français. *Cahiers de grammaire*, **28**:31–48.

Namer, F. (2003b). WaliM: Valider les unités morphologiques complexes par le Web. In *Les unités morphologiques. Actes du 3e forum de morphologie*, numéro 3 in Silexicales, pp. 142–150. Presses universitaires de Lille, Villeneuve d'Ascq.

Namer, F. (2005). *La morphologie constructionnelle du français et les propriétés sémantiques du lexique : Traitement automatique et modélisation*. Habilitation à diriger des recherches, Université de Nancy 2, Nancy.

Neuvel, S. et S. A. Fulop (2002). Unsupervised learning of morphology without morphemes. In *Proceedings of the ACL Workshop on Morphological and Phonological Learning 2002*. ACL, Philadelphia, Penn.

Neuvel, S. et R. Singh (2001). Vive la différence ! What morphology is about. *Folia linguistica*, **35**(3-4):313–320.

New, B., M. Brysbaert, J. Segui, L. Ferrand, et K. Rastle (2004). The processing of singular and plural nouns in French and English. *Journal of Memory and Language*, **51**(4):568–585.

Perruchet, P. et A. Vinter (1998). Parser: A model for word segmentation. *Journal of Memory and Language*, **39**:246–263.

Perruchet, P. et A. Vinter (2002). The self organising consciousness: A framework for implicit learning. In R. M. French et A. Cleeremans, éds., *Implicit Learning and Consciousness. An empirical, philosophical and computational consensus in the making*, pp. 41–67. Psychology Press, London, UK.

Pirrelli, V. (2007). On the cognitive autonomy of language morphology. In N. Hathout et F. Montermini, éds., *Morphologie à Toulouse. Actes du colloque international de morphologie 4es Décembrettes*, numéro 37 in LINCOM Studies in Theoretical Linguistics, pp. 245–269. Lincom Europa, Munich.

Pirrelli, V., B. Calderone, I. Herreros, et M. Virgilio (2004). Non-locality all the way through: Emergent global constraints in the Italian morphological lexicon. In *Proceedings of the Seventh Meeting of the ACL Special Interest Group in Computational Phonology*, pp. 11–19. ACL, Barcelone.

Pirrelli, V. et S. Federici (1994). "Derivational" paradigms in morphonology. In *Proceedings of the 15th International Conference on Computational Linguistics (Coling 1994)*. Kyoto, Japon.

Pirrelli, V. et I. Herreros (2007). Learning morphology by itself. In G. E. Booij, L. Ducceschi, B. Fradin, E. Guevara, A. Ralli, et S. Scalise, éds., *Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*, pp. 269–290. Università degli Studi di Bologna, Bologne.

Pirrelli, V. et F. Yvon (1999). The hidden dimension: A paradigmatic view of data-driven NLP. *Journal of Experimental & Theoretical Artificial Intelligence*, **11**(3):391–408.

Plaut, D. C. et L. M. Gonnerman (2000). Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing? *Language and Cognitive Processes*, **15**(4/5):445–485.

Ploux, S. et B. Victorri (1998). Constructions d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement automatique des langues*, **39**(1):161–182.

Plénat, M. (1988). Morphologie des adjectifs en *-able*. *Cahiers de grammaire*, **13**:101–132.

Plénat, M. (2000). Quelques thèmes de recherche actuels en morphophonologie française. *Cahiers de lexicologie*, **77**:27–62.

Plénat, M. (2005). Brèves remarques sur les déverbaux en *-ette*. In F. Lambert et H. Nølke, éds., *La syntaxe au coeur de la grammaire. Recueil offert en hommage pour le 60^{e} anniversaire de Claude Muller*, pp. 245–258. Presses universitaires de Rennes, Rennes.

Plénat, M. (2008). La liaison « obligatoire » avec et sans enchaînement. In J. Durand, B. Habert, et B. Laks, éds., *Actes du Congrès Mondial de Linguistique Française (CMLF-2008)*, pp. 1657–1667. ILF, Paris.

Plénat, M. (2009). Les contraintes de taille. In B. Fradin, F. Kerleroux, et M. Plénat, éds., *Aperçus de morphologie du français*, pp. 47–63. Presses universitaires de Vincennes, Saint-Denis.

Plénat, M., N. Hathout, et O. Levkovych (2004a). Morphologie et recyclage : Le cas de *anti-*. Communication aux 3es Décembrettes. Toulouse.

Plénat, M., N. Hathout, et A. Ruttik (2004b). Les noms déverbaux en *-ette* du français. Communication au *Colloque international sur les noms déverbaux*. Silex, Université de Lille 3, Villeneuve-d'Asq.

Plénat, M. et M. Roché (2003). Prosodic constraints on suffixation in French. In G. E. Booij, J. DeCesaris, A. Ralli, et S. Scalise, éds., *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*, pp. 285–299. Universitat Pompeu Fabra, Barcelone.

Plénat, M., L. Tanguy, S. Lignon, et N. Serna (2002). La conjecture de Pichon. *Corpus*, **1**:105–150.

Prince, A. et P. Smolensky (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical report 2, Rutgers University Center for Cognitive Science, New Brunswick, NJ.

Prévot, L., B. Gaume, C.-R. Huang, C.-J. Chen, et S.-K. Hsieh (2008). Building and aligning Chinese and French paradigmatic graphs. In *Proceedings of the 18th International Congress of Linguists (CIL 18)*. Séoul.

Rajman, M., J. Lecomte, et P. Paroubek (1997). *Format de description lexicale pour le français. Partie 2 : Description morpho-syntaxique*. Rapport technique, EPFL & INaLF. GRACE GTR-3-2.1.

Rey, A. (1968). Un champs préfixal : Les mots français en *anti-*. *Cahiers de lexicologie*, **12**(1):37–57.

Roché, M. (1997). *Briard*, *bougeoir* et *camionneur*. Dérivés aberrants, dérivés possibles. In D. Corbin, B. Fradin, B. Habert, F. Kerleroux, et M. Plénat, éds., *Mots possibles et mots existants. 1res rencontres du forum de morphologie*, volume 1 de *Silexicales*, pp. 241–250. Villeneuve-d'Asq.

Roché, M. (2003a). De la « bonne formation » des dérivés en *-on*. *Cahiers de grammaire*, **28**:91–112.

Roché, M. (2003b). L'interfixe est-il une unité morphologique ? In B. Fradin, G. Dal, N. Hathout, F. Kerleroux, M. Plénat, et M. Roché, éds., *Les unités morphologiques. Actes du 3e forum de morphologie*, pp. 169–178. Université de Lille 3, Villeneuve d'Ascq.

Roché, M. (2005). Sur une classe d'adjectifs par conversion. In I. Choi-Jonin, M. Bras, A. Dagnac, et M. Rouquier, éds., *Questions de classification en linguistique : Méthodes et descriptions. Mélanges offerts au Professeur Christian Molinier*, pp. 321–347. Peter Lang, Berne.

Roché, M. (2007). Logique lexicale et morphologie : La dérivation en *-isme*. In F. Montermini, G. Boyé, et N. Hathout, éds., *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse*, pp. 45–58. Cascadilla Press, Somerville, Mass. `www.lingref.com/cpp/decemb/5/paper1614.pdf`.

Roché, M. (2008). Structuration du lexique et principe d'économie : Le cas des ethniques. In J. Durand, B. Habert, et B. Laks, éds., *Actes du Congrès Mondial de Linguistique Française (CMLF-2008)*, pp. 1571–1585. ILF, Paris.

Roché, M. (2009). Pour une morphologie *lexicale*. In *La morphologie lexicale est-elle possible ?*, volume 17 de *Mémoires de la Société de Linguistique, Nouvelle Série*. Éditions Peeters, Leuven.

Roché, M. (2010a). Quel traitement unifié pour les dérivations en *-isme* et en *-iste* ? In M. Roché, G. Boyé, N. Hathout, S. Lignon, et M. Plénat, *Des unités morphologiques au lexique*. Manuscrit.

Roché, M. (2010b). Quelle morphologie ? In M. Roché, G. Boyé, N. Hathout, S. Lignon, et M. Plénat, *Des unités morphologiques au lexique*. Manuscrit.

Roché, M. (en préparation). *Propositions en morphologie lexicale. Carnets de grammaire*, CLLE-ERSS, Toulouse.

Roché, M., G. Boyé, N. Hathout, S. Lignon, et M. Plénat (2010). *Des unités morphologiques au lexique*. Manuscrit.

Rumelhart, D. E. et J. L. MacClelland (1986). On learning the past tenses of English verbs. Implicit rules or parallele distributed processing? In J. L. MacClelland et D. E. Rumelhart, éds., *Parallele Distributed Processing: Explorations in the Microstructure of Cognition*, volume 2, pp. 216–261. MIT Press.

Sajous, F. et L. Tanguy (2006). Repérage de créations lexicales sur le Web francophone. Communication à la *Journée d'étude de l'ATALA sur « Le Web comme ressource pour le TAL »*. Paris.

Schone, P. et D. S. Jurafsky (2000). Knowledge-free induction of morphology using latent semantic analysis. In *Proceedings of the Conference on Natural Language Learning 2000 (CoNLL-2000)*, pp. 67–72. Lisbon.

Seidenberg, M. S. et L. M. Gonnerman (2000). Explaining derivational morphology as the convergence of codes. *Trends in Cognitive Sciences*, **4**:353–361.

Skousen, R. (1989). *Analogical Modeling of Language*. Kluwer Academic Publishers, Dordrecht.

Skousen, R. (1992). *Analogy and Structure*. Kluwer Academic Publishers, Dordrecht.

Stroppa, N. (2005). *Définitions et caractérisations de modèles à base d'analogies pour l'apprentissage automatique des langues naturelles*. Thèse de doctorat, École nationale supérieure des télécommunications, Paris.

Stroppa, N. et F. Yvon (2005). An analogical learner for morphological analysis. In *Proceedings of the 9th Conference on Computational Natural Language Learning (CoNLL-2005)*, pp. 120–127. ACL, Ann Arbor, MI.

Stroppa, N. et F. Yvon (2006). Du quatrième de proportion comme principe inductif : une proposition et son application à l'apprentissage de la morphologie. *Traitement automatique des langues*, **47**(2):33–59.

Talmy, L. (2000). *Toward a Cognitive Semantics*. MIT Press, Cambridge, Mass.

Tanguy, L. et N. Hathout (2002a). Webaffix : Un outil d'acquisition morphologique dérivationnelle à partir du Web. In J.-M. Pierrel, éd., *Actes de la 9e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2002)*, pp. 245–254. ATALA, Nancy.

Tanguy, L. et N. Hathout (2002b). Webaffix, un outil pour l'étude de la morphologie sur le web. Communication aux *Journées de Morphologie de l'ERSS*. Toulouse.

Tanguy, L. et N. Hathout (2003). Le français sur internet : recherche automatique de néologismes. Communication à *Multimédia, Internet et Etudes Françaises 2*. Vancouver.

Tanguy, L. et N. Hathout (2007). *Perl pour les linguistes. Programmes en perl pour exploiter les données langagières*. Hermès science-Lavoisier, Paris.

Toussaint, Y., F. Namer, C. Jacquemin, B. Daille, J. Royauté, et N. Hathout (1998). Une approche linguistique et statistique pour l'analyse de l'information en corpus. In *Actes de la 5e conférence annuelle sur le traitement automatique des langues naturelles (TALN-98)*, pp. 182–191. Paris.

Van den Bosch, A. et W. Daelemans (1999). Memory-based morphological analysis. In *Proceedings of the 37th Annual Meeting of the Association of Computational Linguistics*. College Park, MD.

Vandeloise, C. (1986). *L'espace en français : sémantique des prépositions spatiales*. Seuil, Paris.

Victorri, B. (1994). The use of continuity in modeling semantic phenomena. In C. Fuchs et B. Victorri, éds., *Continuity in linguistic semantic*, pp. 241–251. John Benjamins Publishing Company, Amsterdam / Philadelphia.

Voir, M. (1982). Les préfixes transcatégoriels. *Cahiers de lexicologie*, **42**(2):31–46.

Xu, J. et W. B. Croft (1998). Corpus-based stemming using co-occurrence of word variants. *ACM Transaction on Information Systems*, **16**(1):61–81.

Yarowsky, D. et R. Wicentowski (2000). Minimally supervised morphological analysis by multimodal alignment. In *Proceedings of the Association of Computational Linguistics (ACL-2000)*, pp. 207–216. Hong Kong.

Yvon, F. (2006). *Des apprentis pour le traitement automatique des langues*. Habilitation à diriger des recherches, Université Pierre et Marie Curie, Paris.

Yvon, F., N. Stroppa, A. Delhay, et L. Miclet (2004). *Solving analogical equations on words*. Rapport interne D005, École Nationale Supérieure des Télécommunications, Paris.

Zribi-Hertz, A. (1973). *Recherches sur la préfixation productive en français moderne*. Thèse de doctorat de 3e cycle, Université Paris 8.

Zweigenbaum, P. et N. Grabar (1999). Automatic acquisition of morphological knowledge for medical language processing. In W. Horn, Y. Shahar, G. Lindberg, S. Andreassen, et J. Wyatt, éds., *Artificial Intelligence in Medicine*, Lecture Notes in Artificial Intelligence, pp. 416–420. Springer-Verlag, Berlin.

Zweigenbaum, P. et N. Grabar (2003a). Corpus-based associations provide additional morphological variants to medical terminologies. In M. Musen, éd., *AMIA Annual Fall Symposium 2003*, pp. 768–772. Washington, DC.

Zweigenbaum, P. et N. Grabar (2003b). Learning derived words from medical corpora. In *9th Conference on Artificial Intelligence in Medicine Europe*, pp. 189–198. Chypre.

Zweigenbaum, P., F. Hadouche, et N. Grabar (2003). Apprentissage de relations morphologiques en corpus. In *Actes de la 10e conférence annuelle sur le traitement automatique des langues naturelles (TALN-2003)*, pp. 285–294. ATALA, Batz-sur-mer.