



HAL
open science

Dynamique d'évolution de graphes de cooccurrences lexicales : application à l'analyse de comptes rendus en prévention spécialisée entre 1972 et 2010

Dominique Dion

► To cite this version:

Dominique Dion. Dynamique d'évolution de graphes de cooccurrences lexicales : application à l'analyse de comptes rendus en prévention spécialisée entre 1972 et 2010. Informatique. Université Victor Segalen - Bordeaux II, 2012. Français. NNT : 2005 . tel-00842790

HAL Id: tel-00842790

<https://theses.hal.science/tel-00842790>

Submitted on 9 Jul 2013

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

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

Université Bordeaux Segalen
ÉCOLE DOCTORALE : SOCIÉTÉS, POLITIQUE ET SANTÉ PUBLIQUE

Année 2012

Thèse n° 2005

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ BORDEAUX SEGALEN

Mention : Sociétés, Politique, Santé Publique

Spécialité : Sciences cognitives et Ergonomie

Option : Sciences cognitives

Présentée et soutenue publiquement

le 19 décembre 2012

Par DION Dominique

DYNAMIQUE D'ÉVOLUTION DE GRAPHES DE COOCCURRENCES LEXICALES :
APPLICATION À L'ANALYSE DE COMPTES RENDUS EN PRÉVENTION SPÉCIALISÉE
ENTRE 1972 ET 2010

Composition du Jury :

M. Alain LELU, Professeur, Université de Franche-Comté, Rapporteur

M. Michel DURAMPART, Professeur, Université du Sud Toulon-Var, Rapporteur

M. François DANIELLOU, Professeur, Institut Polytechnique de Bordeaux,
Examineur

M. Nicolas HANUSSE, Directeur de Recherches (CNRS), Examineur

M. Benoît LE BLANC, Maître de conférences HDR, Institut Polytechnique de
Bordeaux , Directeur de thèse

Dynamique d'évolution de graphes de
cooccurrences lexicales : application à
l'analyse de comptes rendus en prévention
spécialisée entre 1972 et 2010

Résumé

Ces dix dernières années, l'étude des réseaux petits-mondes a montré une grande stabilité de certaines métriques issues de la théorie des graphes formels. Elle porte sur l'analyse de réseaux traduisant des activités de l'homme : réseaux d'échanges téléphoniques, de connexions aériennes, de navigation sur le Web, de structure des lexiques linguistiques, mais également de réseaux de diffusion des épidémies ou de réseaux de relations sociales. En revanche, peu d'études ont porté sur l'analyse de la dynamique de ces graphes et leur évolution au cours du temps. C'est cette approche que nous développons dans ce document. Nous nous intéressons ici à des graphes obtenus à partir de comptes rendus professionnels de travailleurs sociaux. Ces graphes modélisent les cooccurrences des mots au sein des phrases. Notre analyse porte sur un lexique professionnel et sur l'évolution de son usage sur une période de près de 40 ans. Après avoir constitué notre matériel à partir des textes écrits (environ 1500 pages dactylographiées), ce corpus d'étude a été quantifié, vérifié, homogénéisé et traité orthographiquement sur un mode semi-automatique. Puis ce corpus normalisé a donné lieu à la constitution d'un graphe global pour la période complète, et d'une quarantaine de graphes pour chacune des sous-périodes étudiées. C'est sur ceux-ci que porte l'analyse de la dynamique d'évolution de graphe issu d'un lexique professionnel. Au final, notre travail permet de pointer le paradoxe existant, entre d'une part la contrainte d'écriture liée à un objet qui ne change que très peu dans la nature de sa mission (le travail d'un éducateur de rue) et d'autre part la permissivité du langage oral qui ne cesse d'évoluer (le langage professionnel). Par ailleurs notre approche propose une certaine automatisation pour dégager l'essentiel d'un "dire professionnel" dans un corpus de comptes rendus.

Remerciements

Bien que je sache que je ne vais pas pouvoir remercier et nommer toutes les personnes qui m'ont accompagné à un moment ou à un autre, dans le cadre de l'université pendant cette période de thèse, je leur sais gré de tous les temps qu'elles ont bien voulu m'accorder pour me permettre d'en arriver là.

Je remercie tout d'abord Benoît Le Blanc, pour sa patience, son aide et sa sérénité. Je lui suis très reconnaissant d'avoir soutenu mes efforts dans mes plus grands moments de doute, d'avoir cru en mes capacités dans les moments où je pensais ne plus pouvoir y arriver, et surtout de m'avoir accompagné sans faille jusqu'à ce terme.

Alain Lelu et Michel Durampart ont bien voulu se pencher sur l'appréciation de cette thèse. C'est respectueusement que je leur adresse mes remerciements pour m'avoir accordé de leur temps autour de cette recherche en fin d'année civile et universitaire. Les travaux de Michel Durampart et d'Alain Lelu font profondément écho à des questionnements personnels sur les enjeux communicationnels et sur la diffusion des connaissances dans le monde social.

François Daniellou m'a toujours manifesté sa bienveillance tout au long des études que j'ai suivies en sciences cognitives. Il a été mon tuteur pendant l'ensemble de ma période de thèse, et je l'en remercie. De même, j'adresse ma reconnaissance à l'équipe du laboratoire d'ergonomie des systèmes complexes de l'ENSC-IPB pour son accueil lors de mes demandes de recherches.

J'ai rencontré Nicolas Hanusse au cours de mes apprentissages scientifiques dans le cadre d'échanges thématiques centrés sur les caractéristiques « petit monde » des graphes. Je lui suis extrêmement reconnaissant de sa présence parmi les membres du jury.

J'ai travaillé en tant qu'éducateur spécialisé et pendant vingt-quatre ans dans la structure associative dont les documents ont servi de base à cette étude. Je veux ici remercier toutes les personnes qui y sont intervenues à un moment ou à un autre, quelque'aient été leurs rôles dans l'association. Ce sont toutes ces rencontres et ces échanges qui ont fortifié ma démarche et

qui l'ont surtout construite. Aujourd'hui, c'est l'ensemble actuel de l'équipe que je veux remercier, pour m'avoir soutenu avec souplesse et intelligence à travers les difficultés d'exercice d'une profession en phase de changement organisationnel.

Pour finir, c'est à ma proche famille que je pense, pour tout le temps que je n'ai pas pu lui consacrer. Je remercie tout particulièrement ma femme et mes enfants pour leur soutien, leur tolérance et leurs encouragements.

Sommaire

Table des matières

I	Partie théorique	12
1	Nature et propriétés d'un corpus d'écrits de travail	12
1.1	Introduction	12
1.2	Les signes et les mots	14
1.2.1	Définitions	14
1.2.2	Une organisation des signes	15
1.2.3	Les signes linguistiques	17
1.3	Une archive, un corpus, un discours	18
1.3.1	Langage et travail	19
1.3.2	Travail prescrit et travail réel	20
1.4	Des calculs automatiques sur le langage	21
1.4.1	Les fréquences lexicales	23
1.4.2	La lemmatisation	25
1.4.3	Le lexique et les représentations sociales	26
2	La conceptualisation par graphe	27
2.1	Introduction	27
2.2	La formalisation par un graphe	28
2.2.1	Arêtes et sommets	28
2.2.2	Connexité et parcours	30
2.2.3	Les graphes petit-monde	32
2.3	Les graphes lexicaux	34
II	Partie pratique	36
3	Le matériel d'étude	36
3.1	L'association de prévention spécialisée de Bègles	36
3.2	Les documents étudiés	39

3.3	Le prétraitement des documents	41
3.4	La construction des graphes lexicaux	43
3.4.1	L'étiquetage morpho-syntaxique et la lemmatisation . .	43
3.4.2	Les cooccurrences lexicales séquentielles	46
3.4.3	Le mode de construction du graphe	47
4	La méthodologie du traitement	49
4.1	La triennalisation	49
4.2	Le corpus de comparaison et le test statistique	52
4.3	Les deux enquêtes auprès des professionnels	58
5	Les résultats	60
5.1	Les métriques sur les ensembles triennaux	60
5.2	Les graphes triennaux restreints	67
5.3	La dynamique des graphes triennaux	76
5.3.1	Les constantes	76
5.4	La frise chronologie de la période	84
5.5	Les enquêtes auprès des professionnels	86
5.6	Investigations complémentaires	91
5.7	Un graphe global construit par concaténation d'arêtes issues de triennaux restreints	93
6	Discussion	97
6.1	Rappel synthétique de la démarche	97
6.2	Le langage professionnel	98
6.3	Les graphes	102
6.4	Perspectives	104

Table des figures

1	Graphes connexes ou non, orientés ou non	28
2	Graphes isomorphes	30
3	Zone d'intervention de l'équipe éducative	39
4	Présentation de l'ensemble des documents	40
5	Relation entre le nombre d'arêtes, le nombre de sommets, et le temps dans les ensembles annuels	48
6	Relation entre le nombre d'arêtes, de sommets, dans les trien- naux et le temps	51
7	Relation dans le temps entre le nombre de sommets des en- sembles triennaux et le nombre de sommets des ensembles triennaux restreints	57
8	Nombre d'arêtes par sommets pour les graphes triennaux . . .	62
9	Pourcentage temporel de réduction du nombre de sommets des graphes triennaux par application du test exact de Fisher . . .	68
10	Visualisation du graphe triennal restreint TR_01	74
11	Visualisation du graphe triennal restreint TR_23	74
12	Détail du graphe TR_01, centré sur le mot « équipe »	75
13	Partage lexical entre chaque triennal restreint et le 36ème triennal	80
14	Moments d'apparition des mots du TR36	81
15	Distribution du vocabulaire en pourcentage	82
16	Nombre de mots différents par triennal restreint	83
17	Frise chronologique	84
18	Relation en pourcentage entre le temps et les arités de lemmes de la thématique travail de rue	89
19	Relation en pourcentage entre le temps et les arités de lemmes de la thématique scolarité	90
20	Distribution temporelle du degré de connection au voisinage des termes affichés	91
21	Histogramme des lemmes « Client » / « Usager » dans la pé- riode étudiée	92

22	Représentation globale du graphe des TR concaténés	94
23	Détail du voisinage du mot « prison » dans un extrait du graphe des TR concaténés	95
24	Détail du voisinage du mot « prison » dans le graphe des TR concaténés	96

Liste des tableaux

1	Caractéristiques structurelles des graphes de terrain	34
2	Tableau de contingence	55
3	Réduction par sur-présence dans le Triennal 1	56
4	Réduction par sur-présence dans le Triennal 23	56
5	Signification des noms de variables employés dans les tableaux de résultats 6 et 7.	61
6	Caractéristiques générales des graphes triennaux T01 à T18 .	64
7	Caractéristiques générales des graphes triennaux T19 à T36 .	65
8	Métriques petit-monde de la composante connexe principale des triennaux 1 à 36	66
9	Tableau récapitulatif des graphes triennaux restreints 1-18 . .	70
10	Tableau récapitulatif des graphes triennaux restreints 19-36 . .	71
11	Métriques des graphes triennaux restreints 1-36	72
12	Substantifs et adjectifs permanents significatifs	77
13	Verbes permanents significatifs	77
14	Bruit permanent significatif	78
15	Arêtes permanentes significatives	79
16	Connexions aux premiers voisins de l'aréopage des réponses par thème	88
17	Graphe global concaténé	93
18	Métriques du graphe global concaténé	94

Introduction

Cette recherche explore sur une longue période, une récurrence formelle et homogène de productions écrites annuelles. Ces documents sont destinés à faire le bilan d'une activité professionnelle auprès d'un commanditaire légalement agréé. Ces écrits s'étendent sur une période de presque quarante ans.

Ces productions sont selon les périodes appelés bilans, rapports ou comptes rendus. Ils utilisent des phrases et des mots pour tenter d'exprimer, approcher la matière et l'inventaire d'une telle activité professionnelle. La langue française contient plusieurs dizaines de milliers de mots [52], et le choix d'un mot parmi plusieurs autres est presque toujours une réalité pour chaque scripteur au moment de la rédaction. Nous avons exploré certaines des propriétés de ce choix de mots dans chaque document. Puis, nous avons analysé à travers leurs modalités statistiques les propriétés structurelles de l'ensemble des liens qui les unissent.

Les mots de chaque document, ont été extraits par phrase et associés par paires, à l'aide d'un traitement automatique. L'ensemble des productions écrites a ainsi été formalisé sur chaque année, par la production de graphes à partir des mots utilisés dans chaque bilan. Cette opération a été étendue sur des périodes de trois ans consécutifs pour composer des graphes triennaux.

Cette méthodologie de formalisation nous amène à interroger la nature et la dynamique fondamentale des graphes composés de cette façon. Ces graphes présentent-ils des caractéristiques déjà connues ailleurs pour des graphes représentatifs d'autres activités de l'homme ?

Un graphe est un objet qui dépasse rapidement les capacités cognitives d'un être humain dès lors qu'il a suffisamment de sommets et d'arêtes. Il est donc intéressant d'en dégager sa structure pour le rendre davantage compréhensible. Dispose-t-il du caractère petit-monde ? Puisque ces graphes s'étendent sur une durée conséquente, quelle en est leur dynamique ?

Par dynamique d'un graphe, nous entendons la comparaison de certaines parties d'un graphe d'une période donnée à celles de la période suivante. Ces dynamiques sont toujours en lien avec les exigences imposées par la langue

et celles imposées par la manière de construire le graphe.

Pour étudier ces multiples dimensions, techniques, particulières ou professionnelles, du langage écrit au sein des bilans, nous avons mesuré les fréquences des mots présents pour chaque document. Puis, nous avons comparé les fréquences relatives propres à chaque mot avec les fréquences des mêmes mots mais appartenant à un corpus extérieur. Ce corpus extérieur a été choisi pour être neutre et pour être externe au monde professionnel concerné par les sujets abordés dans les bilans.

La présentation de cette recherche se structure en deux grandes parties.

La première partie est une partie théorique. Elle est composée d'une part des éléments liés à une expression langagière professionnelle exercée dans un cadre de travail et capturée à travers un ensemble d'écrits. C'est le chapitre 1 qui présente les concepts invoqués et positionne les caractéristiques linguistiques et communicationnelles de notre recherche. La partie théorique comprend d'autre part une présentation de la conceptualisation et de la modélisation par des graphes. C'est le chapitre 2 qui explique les structures formelles adoptées.

La seconde partie est une partie pratique. Elle aborde en premier lieu le matériel d'étude ainsi que son pré-traitement (chapitre 3), puis pose la méthode d'analyse mise en place dans cette recherche (chapitre 4). Viennent ensuite les résultats obtenus (chapitre 5) et les éléments de discussion que nous apportons (chapitre 6).

Première partie

Partie théorique

1 Nature et propriétés d'un corpus d'écrits de travail

1.1 Introduction

Le mot Nature en tant que simple substantif provient du latin « natura », terme lui-même dérivé du participe passé du verbe naître. Les travaux de recherches présentés ici portent sur un recueil de productions textuelles et sur la nature de ce recueil. La description de cette nature utilisera des qualificatifs qui pointent la réalité discursive de ces écrits.

Il s'agit de porter notre attention sur de multiples discours, homogènes dans leur forme, regroupés au sein d'un ensemble constitué. L'assemblage de tous ces textes reflète les activités de travail d'un groupe professionnel, et s'établit sur une longue durée. Ainsi l'objet de ces recherches est un discours, qui s'inscrit au sein d'un ensemble d'échanges socio-discursifs intégrés dans une globalité et une période sociétale. Mis bout à bout, ces discours écrits forment un corpus cohérent.

Quant à lui, le terme corpus a été utilisé très tôt dans le vocabulaire religieux chrétien pour désigner l'hostie, puis changeant de cadre, a été repris dans le domaine juridique où il a été utilisé dans le sens d'un recueil de pièces et de documents concernant une même discipline [51].

De nos jours et dans le vocabulaire général des sciences, un corpus désigne un recueil large de documents ou de données [11]. Aussi, dans l'étude des langues, un corpus peut se définir comme un ensemble d'énoncés effectivement émis par des utilisateurs d'une langue commune à une époque donnée [17], ce qui permet de rappeler la dimension commune (même langue et même territoire) aux éléments qui le composent. Ici, le corpus servant d'appui à notre recherche a une représentativité uniquement réduite à son

lieu et à son époque. Il est simplement exhaustif d'un exercice d'écriture spécifique propre à une profession et temporalisé sur une chronologie étendue sur quatre décennies.

C'est une collection de données langagières qui a été sélectionnée et organisée pour servir d'échantillon à un type de langage professionnel, voire à un jargon. En tant qu'objet construit [33], un corpus est donc critique et problématisé. Il informe ses propres composants et ainsi les textes qui le composent peuvent s'éclairer mutuellement [33]. Dans notre démarche, sa construction permet d'apprécier l'évolution du lexique choisi par un ensemble d'auteurs successifs, pour exprimer un bilan. En observant les transformations ou continuités des récits produits année après année, il est possible d'apprécier la dynamique des choix lexicaux.

Les écrits étudiés rendent compte d'activités professionnelles. Or de tout temps, travail et langage ont toujours été très connectés. Aujourd'hui, ils sont même inséparables. L'être humain dans un travail organisé [26] se regroupe avec d'autres humains pour former une équipe. Ensemble, ils visent à la réalisation d'objectifs de production communs, au moyen d'une ou plusieurs techniques, tout en étant soumis à un ensemble de contraintes matérielles et de contraintes sociales. C'est par le langage que la coordination et l'organisation du travail va pouvoir s'opérer.

L'agir et le travail sont inséparables de l'échange langagier. Cet échange langagier s'exerce dans un contexte composé de multiples modalités d'organisation du symbolique, en particulier dans la tension entre l'oral et l'écrit. Il n'y a pas d'écrit sans oral quelle que soit la profession concernée. Il n'y a pas de production écrite sans une forme de production orale préalable. Les modalités de production se chevauchent et sont présentes dans de nombreuses professions qui alternent échanges oraux et productions écrites, ce qui est le cas dans le contexte de la recherche présentée ici. Ce phénomène a été régulièrement mentionné par des linguistes ayant étudié le rapport entre l'oral et l'écrit. L'écrit et l'oral restent malgré tout d'une nature inégale du point de vue de la codification et de la validation sociale.

Il est alors intéressant de reprendre une distinction ancienne [28] et produite par Grant Johnson et Kaplan en 1979 pour exprimer certaines des

subtilités du rapport entre le langage et le travail : talk-about-the-work / talk-in-the-work / talk-as-the-work. Le langage peut porter sur le travail, en le décrivant, le rapportant, le commentant. Mais le langage peut aussi être une composante du travail, ce qui s'avère le cas dans la quasi-totalité de l'industrie tertiaire et de la production de services. Le langage peut enfin être le travail lui-même, comme dans l'enseignement ou les métiers à forte composante sociale. Le verbe "talk" a le mérite de rappeler ici la permanente présence de l'oral au proche voisinage de l'écrit.

Il est alors nécessaire de s'interroger sur les obligations d'écriture dans le cadre de la situation professionnelle afin de savoir si celle-ci relève uniquement de la prescription.

1.2 Les signes et les mots

1.2.1 Définitions

Dans son sens courant, le signe est la « chose perçue qui permet de conclure à l'existence ou la vérité (d'une autre chose absente à laquelle elle est liée) » [51]. Il est aussi possible de ramener ce vocable à sa dimension philosophique et « sémiotique » avec l'aide d'un linguiste. Benveniste définit le signe comme tout objet perceptible pouvant être « le représentant d'autre chose qu'il évoque à titre de substitut » [63].

En épistémologie des sciences sociales, le signe n'existe pas en lui-même, il ne se pose que dans les termes de l'absence d'autres signes en lieu et place de la sienne [5]. C'est également de cette façon, que dans le cadre du cours de linguistique générale, Ferdinand de Saussure, aujourd'hui reconnu comme le fondateur du structuralisme en linguistique, produit sa définition du signe.

Charles Sanders Peirce, selon une manière tout à fait complexe et dans une perspective globale, définit le signe comme « quelque chose qui tient lieu pour quelqu'un de quelque chose sous quelque rapport ou à quelque titre » [15].

Des difficultés fondamentales se situent autour de l'interprétation du signe. C'est une problématique connue depuis longtemps par les linguistes car la problématique se situe dans le rapport du signe avec ce qu'il désigne, ce qu'il signifie, ce qu'il veut dire, ce qu'il peut ou ce qu'il pourrait vouloir

dire. Il y a là une zone particulièrement floue.

Roland Barthes, dans un article de 1964 écrit que le signifié et le signifiant sont les composants du *signe*. Un terme, ajoute-t-il, qui a lui-même une histoire très riche mais qui n'élimine pas son ambiguïté. Celle-ci peut en effet se percevoir à travers « une série de termes affinitaires et dissemblables : signal, indice, icône, symbole, allégorie » [4].

Peirce propose dans ses écrits l'idée d'une séparation du monde des signes et du monde des choses : « tout peut servir de signe, il n'y a rien qui ne puisse être un signe, et notre pensée n'a accès qu'à des signes » [5]. Dans l'analyse de Peirce, être vivant, c'est interpréter le monde, c'est à dire interposer un espace de traitement entre le sujet et les signes qu'il reçoit. La sémiologie de Peirce place le signe dans un rapport avec l'objet et l'interprétant. La notion d'interprétant est considérée comme plutôt obscure. Cependant, elle peut être expliquée par le sens dans lequel le signe se trouve momentanément traduit, voire l'idée à laquelle il donne naissance [18].

1.2.2 Une organisation des signes

La sémiologie est la science qui étudie les systèmes de signes au sein de la vie sociale [17]. Le terme de sémiotique est issu de l'anglais et il recouvre quasiment le même objet d'étude que la sémiologie. En effet, le signe est considéré dans les deux dénominations comme la notion de base pour l'étude des systèmes symboliques [17]. Cependant, le signe linguistique, bien que privilégié au sein des activités sémiotiques humaines, n'est simplement qu'une des catégories de la sémiotique.

Le rapport de sémiologie est globalement problématique pour les signes linguistiques.

A. Polguère propose la définition suivante pour le signe linguistique. Selon sa conception, « un signe au sens large est une association entre une « idée » (le « contenu » du signe) et une « forme ». Les trois termes *association*, *idée* et *forme* étant pris dans leurs acceptions les plus générales possibles [38].

Les signes peuvent être classés en fonction du type de rapport entre le contenu et la forme. Cette classification se ramène à un ensemble de trois types de signes : les indices, les icônes et les symboles. A partir d'un point

de vue synthétique et dans la perspective de Polguère :

- un indice est un signe pour lequel il existe un « lien de proximité » entre son contenu et sa forme,
- « un » icône (*mot masculin*) est un signe qui pointe une marque de ressemblance entre l'idée véhiculée et la forme associée,
- un symbole est un signe pour lequel il n'existe pas de lien logique visible entre l'idée et la forme.

Ce partitionnement est propre à la « lexicologie » et aux « sciences de l'information et de la communication ».

D. Bounoux reprend ce partitionnement en rappelant la distinction entre communication analogique et digitale proposée par Watzlawick [44, p 57] dans le cadre de l'école de Palo Alto, et la nécessité de subdiviser la communication analogique en deux parties. Ces deux parties recouvriront l'une l'indice et l'autre l'icône, alors que la communication digitale recouvrira seulement le symbole.

Si A. Polguère classe le rapport de sémiologie en proximité, ressemblance et séparation, D. Bounoux précise que cette classification ne partage que des catégories mentales et non des choses. Il les classe en variation passant progressivement du continu au discontinu.

- Les indices sont les traces sensibles d'un phénomène, son expression directe, un fragment de celui-ci. Ils restent cependant définitivement ambigus hors de leurs contextes.
- Les icônes sont des signes liés par ressemblance de la forme avec le contenu mais sans la contiguïté des indices. Ils s'ajoutent au monde.
- Les symboles « rompent autant avec la continuité qu'avec la contiguïté » par l'arbitraire du signe.

1.2.3 Les signes linguistiques

Le signe linguistique n'est finalement qu'une des catégories de la sémiologie. Il « se distingue des autres signes par au moins cinq propriétés » [38].

C'est une association entre un signifié et un signifiant, c'est-à-dire une image acoustique. Il a un caractère arbitraire tout en pouvant présenter une certaine flexibilité. Il est figé par son immutabilité et sa stabilité. Il présente un caractère évolutif, pouvant adopter des variations individuelles et/ou temporelles qui traduisent sa dimension dynamique. Il est linéaire car traduisible par une suite de sons.

Cet ensemble de propriétés des signes linguistiques peut être décomposé selon au moins deux grandes dimensions. La première permet d'opposer le caractère lexical au caractère grammatical du signe. La seconde permet d'opposer le caractère élémentaire du signe à son caractère complexe. Le rapport de lexicalité pose problème. La lexicologie appelle lexie l'unité lexicale de la langue, elle appuie cette notion sur les concepts de mot-forme et de lexème. Une lexie regroupe en fait des mots-formes qui finalement ne se distinguent entre eux que par la flexion.

Pour la définition des termes lexique et vocabulaire, nous nous rallierons ici à celle de J. Picoche qui propose « d'appeler lexique l'ensemble des mots qu'une langue met à la disposition des locuteurs, et vocabulaire l'ensemble des mots utilisés par un locuteur donné dans des circonstances données » [11].

Une difficulté réside dans la définition de ce qu'est un mot ainsi que dans celle des unités significatives. Les unités significatives doivent répondre à l'obligation d'appartenir au stock fourni par la langue [17]. De fait, jusqu'à la fin du 18^{ème} siècle, la plupart des linguistes occidentaux se sont accordés à attribuer ce statut d'unité significative au *mot*, car si sa décomposition est possible, elle ne le sera qu'en unités non significatives (syllabes ou lettres) [17].

L'apparition progressive de la linguistique comparative issue de l'étude du sanscrit (langue dont les éléments significatifs dans les mots sont très juxtaposés) a apporté une possibilité théorique de décomposition du mot en plusieurs sous-ensembles spécifiques : les « sémantèmes » et les « morphèmes ».

Ces sous-ensembles sont des traces de marque d'étapes formalisées pendant les processus de la théorisation des sciences du langage.

Selon les auteurs et en fonction de leur orientation théorique, ces termes peuvent être utilisés pour contribuer au découpage notionnel du terme « mot » via le terme de lexème. Le « lexème », cité dans le « dictionnaire d'analyse du discours », est assimilé au morphème. Il a été introduit dans la terminologie linguistique par le biais d'un autre terme, celui de « vocable » qui en désigne l'occurrence dans le discours. C'est une bipartition notionnelle qui fait penser à celle entre « type » et « token », comme dans certains langages informatiques de programmation.

Selon J. Lyons [11], le terme de lexème est utilisé « pour dénoter les unités plus abstraites qui se présentent sous différentes formes flexionnelles selon les règles syntaxiques mises en jeu dans la génération de phrases ». Le concept de dénotation est utilisé pour mentionner la relation entre le signe linguistique et ce à quoi il réfère. En d'autres termes, les vocables sont considérés comme des unités rapportées au texte, c'est à dire des « unités lexicales réalisées dans un discours ».

1.3 Une archive, un corpus, un discours

Pour Dominique Maingueneau [30], « le segment textuel auquel on s'attache le plus spontanément quand on veut pénétrer une archive, c'est bien évidemment l'unité lexicale, qui joue un rôle privilégié dans la conscience que les sujets ont des positionnements discursifs : on reconnaît une archive aux mots qu'elle emploie, on la résume à l'aide de quelques mots-clé ».

Notre objet d'étude constitue bien une archive, en ce sens qu'il est composé d'une collection de comptes rendus d'activité professionnelle. Ces comptes rendus adoptent une forme particulière de discours, avec des contraintes propres à cet exercice discursif. C'est un genre de discours particulier avec des dates et des obligations de production dans l'écriture. Ce genre de discours « s'attache à rapporter des paroles à des lieux », où le lieu et la période sont clairement définis.

Dans notre cas d'étude, c'est une association-employeur qui est l'instance locutaire et prend en charge la validité de l'écriture. Elle en est responsable.

Les textes considérés se rapportent à un secteur professionnel qui relève du « travail social ». Sur la forme, ce dispositif de communication est défini socio-historiquement, s’agissant de comptes rendus professionnels réalisés de longue date. Cette production est archivée chez le locuteur, peut-être également chez l’allocutaire.

Dans le cadre d’un échange communicationnel, l’allocutaire désigne le sujet auquel s’adresse le locuteur. C’est généralement au destinataire que l’on fait allusion en parlant d’allocutaire, mais plus largement ce terme peut désigner le récepteur indépendamment de l’énonciation. Il s’agit donc de l’individu aussi bien que l’entité construite par le locuteur à travers son énoncé. L’allocutaire est une représentation pour le locuteur. Le problème est que la communication est un processus qui met en scène deux locuteurs actifs. L’allocutaire est un coénonciateur actif, il interprète en se mettant à la place du locuteur, et influence par ses réactions les propos du locuteur. L’usage du singulier pour le terme « archive », réfère à un sens tout à fait spécifique et s’apparente à la même racine que le terme « archéologie ». D’une archive, il est possible d’approcher un « savoir » en considérant le choix des mots employés par les scripteurs. Mais ce savoir ne fournit des indications que sur les rapports des signes avec leurs utilisateurs, dans une approche pragmatique. Le lexème « archives » (au pluriel) est utilisé pour désigner un « ensemble de documents hors d’usage courant, rassemblés, répertoriés et conservés pour servir à l’histoire d’une collectivité ou d’un individu » avec un emploi exceptionnel au singulier [50]. Le grand Robert n’utilise que ce sens habituel puisqu’il ne possède pas d’entrée « archive » au singulier.

1.3.1 Langage et travail

Notre travail s’intéresse à un ensemble de textes de métier. De par leur fréquence de production, ces comptes rendus se retrouvent décalés dans le temps, entre les moments de la réalisation de la pratique professionnelle et le moment de l’énonciation.

S’agissant d’écrits produits par des travailleurs sociaux, il est fort probable que l’acte d’écriture ne soit pas vécu comme un élément primordial de la mission professionnelle [13]. Ce n’est pas une tâche première, mais plutôt

secondaire eu égard à une vision davantage orientée vers une mission d'accompagnement éducatif vers la jeunesse des territoires d'affectation, avec un projet de prévention de la marginalité, de facilitation de l'insertion ou de la promotion sociale des jeunes et des familles. Les écrits considérés sont pourtant bel et bien des "écrits de travail", puisqu'il s'agit là du motif de l'énonciation, et que leur production s'effectue "dans des situations de travail". Il y a un intérêt à faire cette différence dans la mesure où l'on veut prendre en compte les propriétés générales de ces écrits.

Il est raisonnable de s'interroger sur la force d'obligation de l'exercice de l'écrit parmi la liste de tâches attendues du métier. On peut donc se demander si ces écrits relèvent du travail prescrit ou du travail réel [8]. Ici, l'énonciation écrite n'est pas monolithique. Plusieurs personnes participent à la rédaction. Celle-ci retrace une activité collective et prend ainsi une forme impersonnelle. Chaque document du corpus a en effet été écrit par une équipe de plusieurs travailleurs, et on peut se demander si il est légitime de les confondre derrière une seule dénomination, derrière un seul nom. Est-ce un énoncé pluriel qui existe sous la dénomination d'équipe ?

Des différences jouent sur les motivations de chacun à l'écriture, qu'elles aillent de l'écriture pour soi à l'écriture prescrite. Il faut mentionner l'existence de carnets de bord personnels, non réellement obligatoires mais suggérés. Ce complément d'écriture est pratique et libre. Il est destiné à faire mémoire et à noter des indications en prévision de difficultés liées à l'oubli, voire de faciliter les étapes ultérieures liées à la rédaction des comptes rendus. L'écriture prescrite est une forme de communication externe destinée à satisfaire aux obligations légales. Elle est destinée aux autorités de tutelle et vise à rendre compte du travail effectué. Cette communication prescrite est une forme de garantie, une anté-réponse au renouvellement du financement de la structure. Elle se retrouve donc en aval du travail réalisé et en amont de l'attribution de moyens pour renouveler ce travail.

1.3.2 Travail prescrit et travail réel

La différence entre travail prescrit et travail réel a été mise en valeur par

les recherches ergonomiques [62].

Le travail prescrit c'est le travail décrit, prévu, appris ou organisé. Il présente de singulières différences entre le monde de sa prescription et celui de son exécution, car dans le monde réel vient se placer ce qui échappe à l'abstraction conceptuelle, à la généralité ou à la prévision anticipatrice.

En effet dans quelle mesure, la multiplicité des situations rencontrées au cours d'une action éducative relève-t-elle d'une simple répétition qui ne comporterait pas d'interprétation du sens des événements en cours ?

On peut sans doute se demander en quoi un travail éducatif est-il différent de l'exécution d'une tâche à la chaîne, ou d'une tâche parfaitement répétitive. La nécessité de rentabiliser toutes les actions humaines qui relèvent du monde du travail a peut-être contaminé toutes les approches du travail professionnel, jusqu'aux tâches liées aux relations humaines. Pourtant, il serait raisonnable de penser que l'exercice du travail éducatif est difficilement réductible à l'application pure et simple de prescriptions et de consignes.

Une part de la souffrance au travail trouve son origine entre travail prescrit et travail réel. C'est une souffrance qui peut émerger d'une déception liée à la difficulté de l'exécution, et va éventuellement se compenser par une recherche (sauvegarde) de solutions personnelles. La création est l'alternative à la composante létale de cette crise. Création qui soutient la nécessité d'un investissement personnel comme élément vital de l'économie psychique dont aucun travailleur (y compris social) ne peut se dispenser en totalité.

Comme dans beaucoup de structures sociales, la difficulté des éducateurs à la réalisation de la production écrite du compte rendu s'avère tout à fait régulièrement. Les voies silencieuses et les chemins de traverses d'accès à la production écrite sont nombreuses. Le compte rendu est un énoncé, c'est le produit d'un acte d'énonciation [31, p 55] lui même défini par « la mise en fonctionnement de la langue par un acte individuel d'utilisation » [31, p 56]. Chaque travailleur social énonce par son écrit ce qu'il tente de transmettre.

1.4 Des calculs automatiques sur le langage

Le traitement automatique des langues dont on connaît depuis de nombreuses années l'acronyme TAL désigne également une revue du même nom

[59]. Son domaine relève d'une recherche pluridisciplinaire qui tend à faire collaborer des scientifiques issus de champs théoriques différents.

La liste des champs concernés peut être concentrée autour de quatre principaux pôles d'attraction. Ces pôles relèvent, dans un ordre décroissant, de la linguistique pour le premier, puis de l'informatique, enfin des mathématiques au sens large, et pour finir d'un ensemble mixte composé à parts inégales des sciences cognitives et de l'intelligence artificielle [12].

La construction historique du traitement automatique des langues montre qu'elle a été influencée par des pôles dont la force d'attraction a pu varier en fonction de diverses contraintes qui ont été soit technologiques, soit pratiques soit sociales. Les tensions existantes à travers les rapports de force en jeu entre ces champs théoriques ont orienté son champ d'investigation. Ce qui a fait jouer des alternatives sur sa dénomination que celle-ci soit française, européenne ou américaine. Le TAL français, dès le début des années soixante, a pu prendre au cours des années qui ont suivi des dénominations très différentes : linguistique mathématique, linguistique computationnelle, linguistique algorithmique, linguistique calculatoire, linguistique quantitative, linguistique appliquée, informatique linguistique, informatique quantitative, analyse des données textuelles ou statistique textuelle.

L'automatisation du traitement des langues a évolué avec la généralisation de l'usage des ordinateurs, et a profité des progrès d'une puissance de calcul toujours plus grande, corrélée à un volume quasi constant du processeur (conjecture de Moore). Cependant, malgré cette amélioration des capacités de calcul des processeurs, l'automatisation du traitement informatique ou automatique du langage reste confrontée de façon permanente aux ambiguïtés et à l'implicite de celui-ci. L'ambiguïté du langage est un trait caractéristique incontournable et historiquement avéré.

Ce trait va venir se problématiser dans l'interprétation automatique de « l'entité linguistique pertinente pour un niveau de traitement » [45].

Par exemple, lors de l'interprétation et de la compréhension automatique de la phrase suivante, « *il poursuit l'homme à vélo* », le lecteur ou bien l'automate peut se demander si l'énonciateur poursuit l'homme qui est sur un vélo ou bien si il le poursuit en étant lui même sur un vélo.

L'implicite du langage dans une production langagière verbale ou écrite prend place dans presque toute interaction humaine. La connaissance du monde et de son fonctionnement par les personnes présentes en contexte permet d'éviter de mentionner l'ensemble des éléments qui pourraient empêcher la compréhension d'une situation. L'absence de perception contextuelle est handicapante.

1.4.1 Les fréquences lexicales

Le langage naturel ne combine pas de façon aléatoire les mots au sein des phrases car il est contraint par la syntaxe. Pour toute mesure automatique, il est nécessaire de définir une norme pour isoler dans la chaîne textuelle des unités sur lesquelles portent un certain type de dénombrement [27].

J.R. Firth [19] propose de pouvoir « connaître » un mot par son voisinage, et de comprendre ainsi davantage de la langue et du monde dans lequel elle existe à travers cette manière d'être ensemble pour des mots.

Le terme de « collocation » créé en 1957 par Firth au sens d'habituelle combinaison de mots, a généré de nombreuses confusions terminologiques relevées ultérieurement par les linguistes. Cette notion, au sens de « combinaison récurrente de mots dans un texte spécifique », est une opération dont la détection plus ou moins automatisable a été validée par les successeurs de Firth.

Wikipédia, définit la collocation comme une cooccurrence privilégiée, *une association habituelle* (c'est nous qui soulignons) « d'un mot à un autre au sein d'une phrase qui sans être fixe, n'est pas pour autant fortuit ». Cette encyclopédie collaborative propose pour illustration les exemples suivants, « voix suave », « courir vite », « entraîner des conséquences ».

Ailleurs [50], la collocation est l'emploi d'un terme relativement à d'autres, toutes variantes morphologiques confondues, et sans égard à la classe grammaticale.

Un autre type de définition a été formalisé via l'usage d'un « mot base » et d'un « mot collocalisé » ou « collocatif » comme par exemple dans l'expression « gros fumeur » ou « pluie diluvienne ». Cette formalisation a été utilisée en TAL à l'aide de fonctions de détections basées sur des lexiques de

compatibilité [19].

La dimension collocative peut se perdre si on joue sur la composition sémantique de la paire, sur la syntaxe (l'ordre est donc essentiel), et on ne peut en aucun cas la retrouver en utilisant des synonymes. Il est préférable de se référer à la dimension distributionnelle, c'est à dire à une « combinaison récurrente » quand on utilise le terme de cooccurrence [19].

Nous conserverons au terme de « cooccurrence » sa dimension statistique. Le problème à résoudre concerne la nature ou la définition du « mot » comme unité significative.

Pour y répondre, l'influence de la typographie se manifeste, par sa définition du mot comme une production graphémique encadrée d'espaces et éventuellement clôturée par un élément de ponctuation. C'est la définition à laquelle cette thèse souscrit. Cette définition est complétée dans notre recherche par la mention du contexte de l'unité linguistique de la phrase. La cooccurrence est ramenée ici à la structure d'une paire de mots repérée par sa position au sein d'une séquence dans la phrase. Il existe malgré tout des mots associés, à des degrés divers de figement et propres à des corpus situés [11].

Une des premières définitions de la cooccurrence stipule que le critère essentiel est celui de la distance entre les mots cooccurrents, distance appréciée en fonction du nombre de mots intercalés [19, 21]. Définition progressivement affinée avec la taille du contexte de référence ainsi que le fait de pouvoir être aisément observable dans les données. La notion de cooccurrence pointe pour un mot donné, la coexistence possible dans un même contexte de plusieurs mots distincts de pouvoir se trouver à proximité d'un autre. Z.S. Harris [11] a introduit dans les années cinquante cette notion de cooccurrence textuelle, dans laquelle il précise la nécessité de référer au texte particulier dont est issue l'occurrence. La définition suivante est proposée par L. Lebart et A. Salem dans leur ouvrage sur la statistique textuelle : « attirances particulières entre couples de formes au sein d'une unité de contexte donnée » [27]. Ils rappellent que la recherche de cooccurrences s'est faite à partir de deux méthodes différentes. La première consiste en une recherche organisée à partir d'une forme pôle, pôle qui contrôle le couple des co-occurents (couple orienté

ou non, avec ou sans mesure de la force d'association), alors que la seconde s'oriente vers une recherche de co-occurents multiples.

Des cooccurrences observées dans des données peuvent être visualisées dans un graphe. A. Mehler a collationné de nombreux et différents types de graphes linguistiques [34, p 349]. Il cite en particulier des graphes d'associations, de citations, de concepts, de cooccurrences, de journaux, de phrases, de thésaurus, de réseaux web ou wiki (web collaboratif).

1.4.2 La lemmatisation

L'appréciation du dénombrement des cooccurrences peut se faire à l'aide d'un dépouillement textuel par lemmatisation ce qui a le mérite d'unifier sous une même dénomination les différentes flexions d'un même lemme, définit ici comme une forme graphique choisie conventionnellement comme adresse dans un lexique.

Dans un article de la revue *Lexicométrica* écrit en 2000 [10], E. Brunet pointe l'absence de ce terme technique de lemmatisation dans les dictionnaires et le présente comme « l'opération de regroupement qui rassemble les formes différentes appartenant au même vocable ».

Dans un autre article [29], B. Lemaire dans un comparatif de méthode d'analyse de corpus, définit la lemmatisation comme le remplacement de « chaque mot par sa forme canonique ». L'avantage principal réside dans la réduction du nombre de formes et l'augmentation des occurrences [29]. Pour des « petits corpus spécialisés » ce peut être un but poursuivi, mais cette méthode présente des inconvénients qui peuvent être préjudiciables à l'appréciation des contextes des mots considérés. Le caractère préjudiciable de cette opération a tendance à être davantage marqué sur les formes non verbales [29].

La lemmatisation consiste pour un texte écrit en français à ramener les verbes sous la forme de leurs infinitifs, les substantifs sous leur forme au singulier, les adjectifs au masculin singulier, et les formes élidées à leur forme sans élision. La lemmatisation génère malgré tout des difficultés liées à la résolution de multiples ambiguïtés qui se construisent autour de la détermination du lemme de rattachement. La résolution de cette difficulté oblige

dans de nombreuses occasions à renoncer à certaines systématisations ou automatisations pour mettre en place des analyses manuelles ou linguistiques.

L'usage de la lemmatisation renvoie directement à l'usage et au choix d'un dictionnaire particulier.

Une lemmatisation complète, sur un corpus assez important, reste une opération coûteuse mais indispensable dans un travail de recherche [10]. Le processus de lemmatisation est donc une opération de regroupement qui rassemble sous un même vocable des formes différentes issues d'un mot spécifique. La notion de lemme a été définie comme une forme arbitraire mais conventionnelle pour l'ensemble des formes d'un paradigme flexionnel.

Par définiton, c'est l'identification du mot sous sa forme canonique c'est à dire sous la forme générique d'où sont dérivées les formes fléchies. La désambiguïsation manuelle complète des homographes représente toujours un effort certain en raison de la taille du corpus textuel à traiter.

1.4.3 Le lexique et les représentations sociales

Dans un article intitulé « Grammaires et lexiques, vecteurs des représentations sociales » inséré dans le livre « Les Représentations Sociales », Horace Rom Harré, philosophe et psychologue avance l'idée que les représentations sociales utilisent un répertoire linguistique comme support [24]. L'idée de cet auteur est que les représentations sociales sont transmises par un lexique et il pose une analyse des représentations sociales des émotions à travers l'usage de leurs répertoires et de leurs lexiques.

Les énoncés linguistiques sont le produit d'une activité de communication. Le « langage, en tant qu'activité symbolique de relation à la réalité et de construction de représentations, ajuste la relation des sujets par rapport à eux-mêmes, transcende leur interaction » [25].

2 La conceptualisation par graphe

2.1 Introduction

La modélisation de données par l'intermédiaire d'un graphe est une idée assez ancienne en sciences et couvre aujourd'hui l'ensemble du champ scientifique. Dans le dictionnaire historique de la langue française [51], l'entrée « graphe » se trouve dans le groupement lexicographique de « graphique ».

Ce terme y est défini par son emprunt à l'anglais, avec une étymologie équitablement partagée entre les deux langues et surtout par une orientation vers la notion au sens large d'écriture. Il a d'abord été appliqué dans le domaine de la chimie, puis vers 1880 son usage s'est étendu aux mathématiques à travers un double sens. D'une part celui de représentation graphique d'une fonction mathématique, et d'autre part celui d'un ensemble de couples d'éléments qui vérifient une relation donnée.

Le terme de graphe est un mot d'usage courant dans les productions scientifiques. Il se confond aisément, par cooccurrence, avec celui de réseau [42]. L'étymologie de réseau l'accroche aux filets (rets) qui servaient à attraper des oiseaux, des poissons ou du gibier. Peu à peu, avec du temps, « réseau » a évolué vers d'avantage d'abstraction, se rapprochant ainsi de celui de graphe. L'encyclopédie Universalis propose d'ailleurs dans son dictionnaire les deux définitions suivantes de réseau. Un réseau y est alors défini comme un « ensemble de lignes entrecroisées, dessin que forment ces lignes » ou un « maillage d'éléments unis par des liens de nature variée ».

Un graphe, quel qu'il soit, n'est finalement rien d'autre qu'un système de couplage entre un ensemble de sommets et un ensemble de liens. C'est le fait que chaque lien relie ou non deux sommets (dimension binaire) qui construit l'existence du graphe.

Les exemples de la figure 1 montrent différentes formes de liaisons entre des entités, qu'il s'agisse de routes entre villes, de messages entre personnes ou d'influences entre mouvements artistiques.

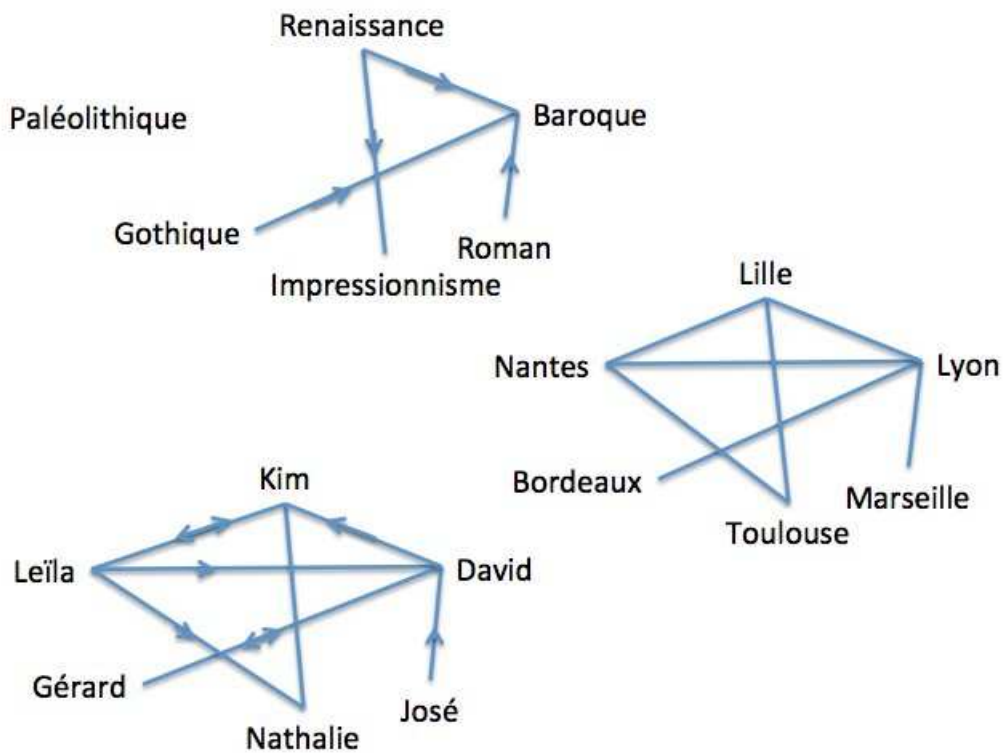


FIG. 1 – Graphes connexes ou non, orientés ou non

Le graphe des mouvements artistiques est orienté et n'est pas connexe car le sommet Paléolithique est isolé. Le graphe des relations entre personnes est orienté et connexe. Le graphe des villes est à la fois non orienté et connexe.

Dans la suite de ce chapitre, nous utiliserons indifféremment les termes de réseau ou de graphe.

2.2 La formalisation par un graphe

2.2.1 Arêtes et sommets

Un graphe permet de représenter la structure et les connexions d'un ensemble complexe comportant un grand nombre d'objets. Il permet d'exprimer

à la fois les relations et les dépendances précises entre les éléments, même si les éléments sont très nombreux. Ainsi, un réseau est une collection de sommets qui interagissent à travers des liens [32] ou des relations. Cette collection, insérée dans la réalité change au cours du temps, des éléments nouveaux peuvent apparaître ou disparaître. L'analyse des liens, des arêtes ou des relations, a révélé des propriétés structurelles.

Les éléments qui composent un graphe sont habituellement appelés noeuds ou sommets (vertices) et les relations appelées arêtes ou arcs (edges). Les termes de noeuds ou de sommets sont équivalents et s'échangent sans modification de signification. En revanche, le choix de la dénomination des relations n'est pas pas totalement neutre. En raison de la question de l'orientation des liens entre les éléments, certains graphes sont orientés alors que d'autres ne le sont pas. Dans l'étude des propriétés structurelles des graphes, l'orientation des liens peut jouer un rôle. C'est par exemple le cas dans un graphe modélisant un réseau routier sans sens interdit où les voitures peuvent aller dans les deux sens. Dans d'autres situations, elle joue un rôle fondamental, comme dans le cas d'un graphe modélisant des échanges intracellulaires ou encore des échanges de messages entre individus.

Dans une modélisation, le terme d'arête est réservé pour la désignation d'un lien non orienté et celui d'arc pour une relation orientée. Un graphe est dit non orienté si la relation binaire qui existe entre ses sommets est symétrique, ce qui revient à dire qu'un graphe non orienté est un graphe dont tous les arcs sont symétriques.

Dans le cadre d'une approche théorique, un graphe est défini par deux ensembles, V (vertices) et E (edges), où chaque élément de E est une paire d'éléments de V . Il vient alors assez naturellement la question de la représentation visuelle de toutes les liaisons entre sommets qui composent le graphe. Si cette représentation visuelle peut se faire sans qu'aucun arc (ou arête) n'en recoupe un autre, le graphe est dit planaire. Toutefois il est le plus souvent nécessaire de disposer les différents sommets dans un ordre plutôt qu'un autre pour obtenir une représentation planaire, encore que ce type de représentation s'avère impossible à obtenir en toute généralité.

Un graphe est une modélisation topologique sans géométrie particulière. il

peut avoir plusieurs représentations ou diagrammes. Les graphes de la figure 2 sont isomorphes. Si visuellement ils évoquent des formes différentes, il n'en demeure pas moins que ces graphes sont bien les mêmes : même nombre de sommets et mêmes liaisons entre les sommets

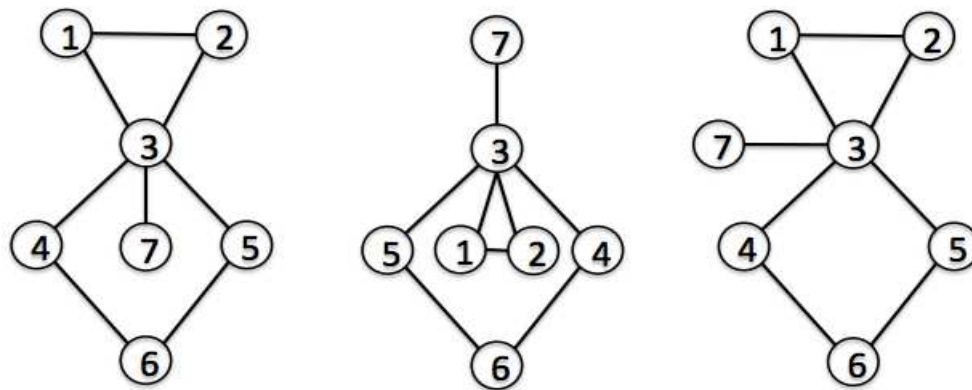


FIG. 2 – Graphes isomorphes

2.2.2 Connexité et parcours

Un graphe sans orientation est donc un ensemble composé de sommets et d'arêtes.

Il est appelé graphe simple si il y a au plus une arête entre deux sommets et s'il n'y a pas de boucle réflexive d'un sommet sur lui même. Dans le cas où une paire de sommets peut être reliée par de multiples arêtes, ce type de graphe est dénommé multigraphe. La notion de multigraphe autorise également la boucle.

Un graphe est qualifiable par le nombre de ses sommets et par le nombre de ses arêtes. Le nombre de ses sommets est l'ordre du graphe et le nombre de ses arêtes en est la taille. Si on augmente le nombre d'arêtes, le graphe va se

densifier. Cependant, la densification n'apportera pas forcément sa connexité. En effet, si pour chacune des paires de sommets distincts u et v , il n'est pas forcé qu'il existe un chemin de u à v . Seule une caractéristique de connexité peut apporter la satisfaction de cette condition. Dans le cas où l'ensemble du graphe n'est pas connexe en totalité, il est possible de ramener la connexité du graphe au sous-graphe connexe maximal qu'il contient. Un graphe est connexe si pour tous sommets x et y , on peut atteindre y par un déplacement à partir de x et via une suite de sommets consécutifs.

Un sommet est isolé si il n'a aucun voisin, si aucune arête ne le prend pour point de départ, si il n'a aucun sommet adjacent. Si un sommet n'est pas isolé, son degré représente le nombre d'arêtes qui l'ont pour extrémité. Le terme d'arité peut être utilisé pour exprimer cette connectivité d'un sommet. Le degré d'un sommet correspond donc au nombre d'arcs ou d'arêtes qui lui sont incidents et qui peuvent être empruntées pour se déplacer à partir de là.

La visualisation d'un graphe incite en effet à évaluer la possibilité d'un déplacement à l'intérieur, d'un cheminement. Un chemin ou une chaîne dans un graphe correspond à une séquence alternée de sommets et d'arêtes commençant et se terminant par un sommet. Une séquence de cheminement démarre à partir d'un noeud donné et se clôture après avoir parcouru l'ensemble des noeuds incidents traversés au long de la chaîne. Le terme de chaîne est utilisé pour les graphes non orientés. Le nombre d'arêtes parcourues est la longueur de la chaîne.

Chaque sommet rencontré sur le chemin propose ainsi une ou plusieurs opportunités de déplacement, opportunités dépendantes de son degré. La distance entre deux sommets u et v si elle existe est la longueur du plus court chemin qui relie u à v (décompté en nombre d'arêtes).

Les métriques opérées sur les graphes relèvent de mesures locales ou globales. Le nombre de liaisons par sommet est utilisé dans une mesure locale alors que la distance est utilisée dans une mesure globale du graphe.

La mesure de la distance est utilisée dans une des métriques de référence des propriétés d'un graphe sous le nom de distance caractéristique (moyenne des distances pour tous les sommets). Elle est également utilisée pour la

mesure du diamètre d'un graphe : le diamètre désignant le meilleur raccourci parmi les plus grandes distances existantes entre deux sommets.

Dans cette terminologie des graphes, le terme de cluster souligne le caractère d'amas ou d'agrégat dans une vision partitionnée des sommets d'un graphe. Plus précisément, la tendance à l'agrégation ou au regroupement désigne pour chaque sommet le ratio arêtes présentes / arêtes possibles des voisins immédiats du sommet de référence. La moyenne de ce ratio est appelée coefficient d'agrégation. Si un sommet v a k_v voisins, il y a $\binom{k_v(k_v-1)}{2}$ connexions possibles entre les voisins, alors le coefficient d'agrégation du sommet v est la fraction suivante [40][42]

$$C_v = \frac{2 \cdot \text{le nombre de liens entre les voisins de } v}{k_v(k_v-1)}.$$

Le coefficient d'agrégation de l'ensemble du graphe est la moyenne des C_v avec $C = \frac{1}{N} \sum_v C_v$

Cependant cette définition, que nous avons effectivement utilisée ici, présente un biais en faveur des sommets n'ayant qu'un petit degré ou une petite arité, et peut être remplacée de façon non biaisée par la suivante [37]

$$C_v = \frac{3 \cdot \text{le nombre de triangles du graphe}}{\text{le nombre de triples connectés à } v}$$

Définition dans laquelle les triangles sont les ensembles de trois sommets où chaque sommet est relié aux deux autres et où les triples représentent les deux sommets liés à v qu'ils soient eux-mêmes ou non reliés entre eux.

2.2.3 Les graphes petit-monde

Les célèbres travaux de D.J. Watts et S.H. Strogatz en 1998 ont montré que les propriétés des réseaux ou des graphes construits de façon déterministe n'expliquent pas directement les données issues de la réalité. Construire un graphe de façon déterministe, c'est utiliser un même mode de construction. Ces auteurs ont montré que les graphes utilisant comme méthode de construction la reproduction d'un même motif (graphe régulier) présentaient une distance caractéristique supérieure à celle issue de la réalité. De même, les graphes construits avec des méthodes reposant sur des liaisons aléatoires régulièrement distribuées ont tendance à présenter une distance caractéristique inférieure à celle issue de graphes traduisant des activités de l'homme.

Les graphes de terrain qui sont des graphes issus de la modélisation d'observation d'activités humaines ou naturelles ont tendance à se ressembler en terme de domaine de valeurs de leurs distances caractéristiques.

Par cette modification de l'écriture du mode de connexion des arêtes, ces auteurs ont modélisé le concept d'un réseau petit monde. Ils mentionnent eux-mêmes le rôle clé du raccourci [42]. Le caractère petit-monde d'un graphe incarne une caractéristique tout à fait particulière. C'est une structure qui optimise à la fois la préservation d'une capacité de régulation locale sachant que celle-ci tend normalement à agrandir la distance entre les sommets tout en assurant le maintien permanent d'une possibilité de raccourci [32].

A partir des travaux de Erdos et Rényi [23], un graphe aléatoire se définit comme un graphe à N sommets connectés par n arêtes choisies aléatoirement avec la probabilité p parmi les $N.(N-1)/2$ arêtes possibles. Ce graphe a donc en moyenne $p.N.(N-1)/2$ arêtes distribuées aléatoirement. Ce type de graphe possède dans son propre espace de probabilité, un faible diamètre, une faible distance caractéristique ainsi qu'un faible coefficient d'agrégation.

Pour les graphes réguliers, il est possible de ramener leur conception à partir d'un anneau ou d'une grille. Les sommets sont répartis N fois en voisinage régulièrement le long de l'anneau et tous les sommets ont le même degré de connectivité. Ce type de graphe possède un grand diamètre puisque pour se déplacer d'un sommet quelconque au sommet opposé (le plus éloigné) il faudra parcourir $N/2$ arêtes. Il possède également une forte distance caractéristique et un fort coefficient d'agrégation.

Les grands graphes de terrain sont peu denses, ils ont peu d'arêtes différentes relativement au nombre de leurs sommets [22], et cette densité se traduit par sa distance caractéristique. La limite théorique de celle-ci est marquée par le rapport du logarithme du nombre de ses sommets sur le logarithme de son degré [2]. Cette faible densité est compensée par une plutôt forte connectivité et est exprimée par un taux d'agrégation élevé.

Les graphes de terrain présentent en outre une distribution particulière du degré d'incidence de chaque sommet [2, 3] sous la forme d'une loi puissance. La probabilité k qu'un sommet donné dans le graphe ait k voisins décroît comme une loi puissance de la forme $k^{-\lambda}$ avec λ comme constante. Gaume

[22] synthétise cet ensemble de caractères avec le tableau 1.

	Distance caractéristique	Taux d'agrégation
Graphes aléatoires	faible	faible
Graphes de terrain	faible	fort
Graphe type grille	forte	fort

TAB. 1 – Caractéristiques structurelles des graphes de terrain

2.3 Les graphes lexicaux

Il existe différentes sortes de réseaux ou de graphes lexicaux. La nature et la catégorie des arêtes présentes dans le graphe déterminent l'appartenance à plusieurs grands types de relations.

Un premier type de relation fait appel à une structure syntagmatique. Un syntagme au sens large peut se limiter à une suite de mots qui se suivent et produisent un sens acceptable. Il peut également être défini par la co-présence de deux ou plusieurs unités dans un énoncé [17]. Le terme d'énoncé est lui-même utilisé de manière polysémique dans les sciences du langage [17] selon qu'on le conçoit comme résultat d'une production spécifique (l'énonciation) ou qu'on le conçoit simplement comme « une séquence verbale de taille variable » [11]. Un graphe de ce type aura construit des relations entre les sommets (les mots) à partir d'une co-présence au sein d'une même phrase, d'un même texte ou d'un même corpus, co-présence déterminée à l'aide d'une distance, ou d'une fenêtre d'observation.

Un second type de relation fait appel à une structure paradigmaticque. Ce type de structure peut prendre des formes plus ou moins variables, puisqu'au sens large [17] un paradigme est une réunion d'unités linguistiques quelque en soit le principe de réunion. La synonymie, mais aussi la relation

entre l'adjectif et le substantif, voire la métaphore sont des relations de type paradigmatic.

Un troisième type de relation proposé par B. Gaume [22] fait appel à une notion de proximité sémantique. La proximité sémantique entre deux verbes peut être appréciée si l'un est utilisé dans la définition de l'autre à l'intérieur d'un dictionnaire général. Un graphe lexico-verbal de ce type peut être construit à partir de cette démarche.

Cependant, tout graphe lexical peut présenter des caractéristiques spécifiques. De fait, de tels graphes peuvent se focaliser sur la question des cooccurrences au sein de la phrase. Ainsi, Ferrer i Cancho [21] décrit les réseaux de cooccurrences comme une construction issue d'un corpus à partir de l'idée suivante : des mots proches entre eux au sein d'une phrase sont liés à un niveau syntaxique, sachant qu'un corpus est vu dans cette proposition comme un *grand* ensemble de phrases.

Chaque mot d'un réseau de cooccurrence est un sommet de graphe. Deux mots « x » et « y », « x » étant placé avant « y » dans la même phrase, sont reliés par une arête si leur distance de séparation est inférieure à une distance D. Cette distance, qui est arbitraire, est la taille de la fenêtre qui permet de mesurer le nombre de mots présents entre « x » et « y » nombre auquel on ajoute 1. Nous avons utilisé une fenêtre de taille deux.

Par phrase, l'ensemble des cooccurrences saisies avec cette méthode représente un lien de nature syntaxique dans 70% des cas [21]. De tels réseaux de cooccurrences sont considérés comme une approximation grossière en comparaison avec celle d'un réseau de dépendance syntaxique.

Avec une chronologie temporelle, un graphe de cooccurrences lexicales offre une possibilité d'apprécier sa dynamique. Au cours du temps, les entrées lexicales évoluent, les sommets qui en sont les reflets apparaissent ou disparaissent. L'évolution du vocabulaire peut se mesurer à l'aide de la perspective chronologique.

Deuxième partie

Partie pratique

3 Le matériel d'étude

3.1 L'association de prévention spécialisée de Bègles

Notre recherche explore un matériel textuel dont la rédaction s'est échelonnée avec une fréquence annuelle entre 1971 et 2010. C'est un ensemble de textes destinés à rendre compte d'une action de prévention spécialisée sur l'une des vingt-sept communes de la communauté urbaine de Bordeaux (CUB).

L'action de prévention spécialisée relève depuis les lois sur la décentralisation de 1986 d'une compétence légale des départements au titre de l'aide sociale à l'enfance (art L 121-2 et L 221-1 du code de l'action sociale et des familles). Elle vise à prévenir la marginalisation et à faciliter l'insertion ou la promotion sociale des jeunes et des familles. Elle s'adresse aux jeunes et aux familles en difficulté ou en rupture avec leur milieu.

Le département de la Gironde a choisi jusqu'à présent de confier cette mission à des associations loi 1901. Pour ce type d'intervention, les associations emploient des éducateurs spécialisés qui interviennent dans la rue pour entrer en contact avec les jeunes et les familles. Les éducateurs spécialisés concernés relèvent de la politique départementale de l'aide sociale à l'enfance (ASE). Les actions de prévention spécialisée mettent donc en relation les associations qui exécutent la mission et le conseil général qui contrôle cette prestation à l'aide de rencontres régulières avec les associations, par l'échange de documents et avec l'aide d'une convention [48]. Cette intervention est formalisée en Gironde à l'aide d'une "charte départementale de la prévention spécialisée" aujourd'hui actualisée et toujours valide pour la période 2008-2012. Cette charte précise la place de la prévention spécialisée dans un schéma départemental conjoint, celui de la prévention et celui de la protection de l'enfance, de la jeunesse et de la famille [48]. Le conseil général finance les

neuf associations qui sont présentes sur le département. Selon les départements, il peut exister d'autres formes de coopérations institutionnelles qui ne sollicitent pas les associations loi 1901, comme par exemple dans le cas de services intégrés dans le fonctionnement du département ou dans le cas de délégation partielle de la prévention spécialisée aux communes concernées.

Dans le cas de l'Association de Prévention Spécialisée de Bègles (APSB), et pour la période allant de 1971 à 2010, le lien avec le conseil général peut être analysé à travers le lexique utilisé dans les textes des bilans. Il s'agit là de la principale forme d'échange par des écrits. La taille de l'équipe éducative pendant la période considérée a varié de trois à huit personnes. Cette taille s'est stabilisée depuis près de vingt ans autour de huit personnes.

Dans le cadre de l'exercice de son métier, chaque éducateur se rend au travail à l'heure prévue par l'emploi du temps. Celui-ci est avalisé par la direction. La prise de poste professionnel débute dans les locaux de l'association. Puis, en fonction des régularités internes et institutionnelles propres à l'APSB (réunions de régulation, rencontre avec des partenaires, préparation ou mise en place de projets d'activité), les éducateurs se rendent ensemble ou séparément sur leurs lieux d'affectation.

Les intitulés des écrits bilans ont largement évolués au cours des années. Ils ont été alternés sous le nom générique de "rapports" de natures diverses (rapports administratifs, rapports d'activité) ou sous un autre nom générique, celui de bilans de natures également diverses (bilans d'activité, bilans associatifs). Ils donnent ainsi un peu à voir d'une certaine variabilité terminologique. Un tableau récapitulatif des dénominations utilisées au cours du temps est fourni en annexe.

Cependant, les lois relatives à tous les types d'interventions sociales ont nettement évolué ces dernières années.

La loi du 2 janvier 2002 (art L 312-8 du code de l'action sociale et des familles) stipule que les établissements et services sociaux et médico-sociaux procèdent à l'évaluation de leurs activités et de la qualité des prestations qu'ils délivrent, au regard notamment de procédures, de références et de recommandations de bonnes pratiques professionnelles validées ou élaborées, selon les catégories d'établissements ou de services. Un premier conseil de

l'évaluation sociale et médico-sociale a été créé en 2003, puis a été suivi en 2007 par celle de l'ANESM (Agence nationale de l'évaluation et de la qualité des établissements et services sociaux et médico-sociaux). En Gironde, le conseil général a formalisé un nouvel écrit pour rendre compte des actions de prévention spécialisée. Il s'agit du « compte rendu de mission », élaboré en 2006. Le « compte rendu de mission » fait partie intégrante des documents écrits échangés. Chaque association conventionnée fournit le sien depuis 2006. Il est structuré sur des dimensions quantitatives et qualitatives, et hors les tableaux qui composent, celui de 2010 est consultable en annexe. La demande de compte rendu est centrée sur le territoire d'intervention, la population concernée, le partenariat, les problématiques des jeunes, les problématiques prioritaires pour l'année et les réponses développées par l'association. Le compte rendu de mission utilise les facilités informatiques pour la gestion de sa mise en forme narrative, mais il reste orienté dans une forme générique de comptabilité. En 2006, l'APSB s'est contenté de produire son compte rendu de mission et n'a pas fourni de bilan d'activité. A partir de 2007, l'APSB a fourni tant le compte rendu de mission que le bilan de l'année considérée. C'est ce qui explique l'absence de bilan pour 2006.

L'APSB intervient sur la commune de Bègles. Il s'agit d'une commune d'environ 10 km² et peuplée d'environ 25000 habitants (chiffres 2009). Pour tout ce qui concerne la politique en faveur des quartiers défavorisés, les pouvoirs publics ont mis en place une organisation basée d'abord sur des institutions et agences nationales dont nous ne citerons que les plus connues (ministère de la ville, agence nationale pour la rénovation urbaine (Anru), l'agence nationale pour la cohésion sociale et l'égalité des chances (Acsé)), puis sur une organisation territoriale qui coordonnent services de l'Etat et acteurs locaux. C'est donc dans des zones urbaines spécifiques que vont plutôt se localiser les interventions de prévention spécialisée. Ces zones spécifiques sont toujours urbaines, souvent sensibles (ZUS), redynamisées (ZRU) ou bénéficiant d'une aide de cohésion sociale (CUCS). Ces zones sont définies par le comité interministériel des villes et du développement social urbain (CIV). Cette définition prend en compte des références communales, départementales, régionales et nationales qui peuvent parfois être en tension. Il y a

treize ZUS et huit ZRU dans la CUB.

Sur la commune de Bègles, il y a trois ZUS. L'une d'entre elles bénéficie également d'un statut de ZRU. En ce qui concerne les zones CUCS, elles sont habitées par environ 36% des habitants.



FIG. 3 – Zone d'intervention de l'équipe éducative

3.2 Les documents étudiés

La période de production analysée s'étend sur une quarantaine d'années. Le matériel rassemblé est composé de 38 écrits différents qui sont produits dans un contexte spécifique par son articulation entre énoncé et énonciation.



FIG. 4 – Présentation de l'ensemble des documents

Le contexte d'énonciation inclut le caractère obligatoire du *compte rendu* (au sens large) des actions de prévention spécialisée réalisées au cours de l'année précédente. L'énoncé est ensuite adressé au destinataire qui dispose de la compétence légale pour le maintien de l'habilitation de cet exercice associatif.

Chaque texte présente plusieurs caractéristiques dans son rapport à la question de l'évaluation. L'association valide l'écrit comme bilan annuel. Elle est l'auteur, l'énonciateur théorique, global.

Cependant, il existe des auteurs plus incarnés, choisissant leurs mots pour exprimer les actions réalisées, les difficultés rencontrées, les imprévus. Les auteurs ont investi au niveau de l'éducation spécialisée un espace désigné du territoire, ils y sont connus et perçus à travers cet exercice de leur métier. Ils l'ont fait plus ou moins individuellement, plus ou moins en investissement binomial, en investissement d'équipe, en rapport avec leurs implications dans cet exercice. Il faut bien devoir choisir ses mots pour exprimer les difficultés des populations rencontrées en accord avec les représentations de l'auteur. De plus, le passage à l'écriture est décalé dans le temps par rapport à la réalisation de cette série d'exercices professionnels. Ce décalage bien qu'il puisse se compenser par des notes prises en cours d'année peut se réguler dans l'échange oral [28], et peut-être aussi par une forme de répétition scripturale. Cette répétition est une possibilité entre renouvellement et redite, recommencement et reprise.

Les auteurs exercent leur liberté dans le choix de leurs mots, illustrant alors les « relations des signes avec leurs utilisateurs, à leur emploi et à leurs effets » [31].

L'écriture du bilan contraint l'écrit comme un préalable à un retour de financement pour l'année qui suivra.

Les bilans d'activité sont donc adressés une fois par an au conseil général de la Gironde, accompagnés des comptes rendus de mission.

3.3 Le prétraitement des documents

L'ensemble des bilans représente 1895 pages dactylographiées. Trente deux pages blanches destinées à aérer la présentation ont été insérées par une ou par deux dans quelques bilans de la période.

Pendant la première partie de la période 1972-2010, les documents ont été tapés à la machine à écrire. La machine à écrire a d'abord été manuelle, puis elle est devenue électrique. Chaque machine à écrire a utilisé les polices propres à son époque. Les documents du début de cette période n'ont été conservés que sous format papier. Tous les textes concernés ont dû être photocopiés pour être numérisés.

La numérisation nécessaire des photocopies de chaque année a dû être faite page par page et enregistré au format pdf. Chaque fichier pdf a ensuite été passé à travers une série logicielle OCR (Optical Character Recognition) pour fournir par page photocopiée un fichier de type RTF (Rich Text Format). Pour chaque année de cette première période, chaque fichier de page de type RTF a été transféré dans un traitement de texte disposant d'un correcteur orthographique afin de s'assurer de la conformité orthographique des productions OCR. Les nombreuses erreurs orthographiques de la reconnaissance de caractère ont été manuellement corrigées, ce qui implique un temps de lecture conséquent.

Ce n'est qu'après 1986 que les productions des textes ont été réalisées à l'aide de traitements de textes installés sur des ordinateurs, avec les formats standards des systèmes d'exploitation en place. L'ensemble des textes obtenus

après ces processus, de la première à la dernière période, présentaient tous suivant l'origine de leur création, le défaut de ne pas être normalisés sur les deux plans d'encodage qui suivent. En effet, l'encodage informatique et l'encodage typographique utilisés par le secrétariat de l'association ont évolué pendant ces trois décennies.

En ce qui concerne l'encodage informatique, les textes n'utilisaient pas tous les mêmes ensembles de caractères. Les charsets suivants ont été régulièrement rencontrés : ASCII, ISO-8859-1, ISO-8859-15, Windows-1252, UTF-8. Ils ont ensuite été ramenés par choix sous le même charset en UTF-8. Les traitements de textes qui ont été utilisés à l'APSB n'avaient jamais le même ensemble d'encodage pour les délimitations et les dénominations de champ. Les premiers bilans de la période informatique ont eu chacun leur gestion des espaces (insécables ou non), des tabulations ou des retours à la ligne. Les systèmes d'exploitation ont été différents sur l'ensemble de la période partant de Dos 6.22, passant par Windows 3.1 et 3.11 pour aboutir à Windows XP.

Pour l'encodage typographique utilisé par le secrétariat, le point le plus caractéristique a été la question de la gestion des majuscules. La présence ou l'absence des majuscules a surtout été problématique pour la gestion des noms propres, des prénoms et des titres. *Monsieur* est il la même chose que *monsieur* ou *Mr* ou *M* ou *mr* ? Il en a été de même pour les acronymes et les sigles. L'APSB désigne-t-il la même chose que l'A.P.S.B ? La typographie des chiffres a également évolué incluant ou pas un espace entre les centaines et les centaines de milliers. 10500 € = 10 500 € ? De même, les dénominations des activités ont évoluées en fonction des modes, est-ce que le bodyboard est la même chose que le morey ? Il a fallu gérer et prendre en compte la normalisation des éléments mentionnés ci-dessus afin de pouvoir stocker chaque forme lexicale sous un identificateur unique. On obtient ainsi un texte dans lequel on peut mesurer des présences lexicales et organiser des comparaisons entre années.

Le nombre de phrases par ligne a été fixé à un.

3.4 La construction des graphes lexicaux

3.4.1 L'étiquetage morpho-syntaxique et la lemmatisation

Nous avons effectué l'étiquetage morphosyntaxique de chaque texte. Cet étiquetage consiste à identifier pour chaque mot sa classe morphosyntaxique à partir de son contexte dans la phrase et avec l'aide d'un ensemble de paramètres lexicaux extérieurs qui sont spécifiques et propres à une langue. En accord avec le dictionnaire de l'Encyclopédie Universalis, nous avons pris pour définition du sens large d'une phrase, le concept d'*un énoncé de mots ayant un sens complet*. Cet énoncé doit se clôturer par un point à la ligne, ou un signe spécifique de fin de phrase.

Nous avons choisi Tree-Tagger comme étiqueteur morphosyntaxique. Cet étiqueteur a été développé par Helmut Schmid dans le cadre du projet TC (« Textual Corpora and tools for their exploration ») à l'Institut de Linguistique Computationnelle de l'Université de Stuttgart [55]. Ce logiciel n'est pas en open source mais, selon les termes de sa licence, il est librement utilisable dans le cadre de la recherche.

Helmut Schmid a élaboré par la suite une fonction complémentaire pour intégrer un lexique externe. En effet, le lexique utilisé par Tree-Tagger repose au départ sur un fichier de paramètres linguistiques français intégré par compilation. L'utilisation de cette fonction supplémentaire a amené à la construction progressive d'un lexique français complémentaire composé au fur et à mesure de la rencontre de mots nouveaux. Un tel dictionnaire a été construit avec les mots issus de nos captures, mais inconnus des compilations originelles de l'étiqueteur.

Outre l'annotation morpho-syntaxique et automatique du texte, Tree-Tagger (part-of-speech tagging) fournit également en parallèle les lemmes des mots annotés ou « taggés », bien qu'il ne les désambiguïse pas, laissant cette appréciation à son utilisateur. Nous avons utilisé ces lemmes pour transformer les textes dans une version lemmatisée.

Un traitement complémentaire a été réalisé sur tous les documents rassemblés. Nous avons supprimé les sommaires, les tableaux, les graphiques et les listings nominatifs en raison de notre volonté de prendre en compte la

dimension narrative de l'écrit.

L'utilisation de l'étiqueteur en lemmatiseur a généré des indéci- sions de lemmatisation propres au langage français dont voici quelques exemples. SUIS a-t-il pour lemme SUIVRE ou ÊTRE, FILS a-t-il pour lemme FILS ou FIL, CONVIENT a-t-il pour lemme CONVIENIR ou CONVENIR, DROITE a-t-il pour lemme DROIT ou DROITE ? Dans les cas d'ambiguïtés décisionnelles de l'éti- queteur, un seul lemme a été conservé par forme.

Les « tags » sont les étiquettes morpho-syntaxiques de l'étiqueteur. Leur liste est consultable sur le site de Tree-Tagger [58] ainsi que dans l'an- nexe 10.

Les lemmes associés à des tags relevant de catégories grammaticales ou « morpho-syntaxiques » spécifiques ont été éliminés. Ont alors été supprimés des textes, et n'ont pas été intégrés dans les graphes, les prépositions avec ou sans article, les pronoms possessifs, les conjonctions, les articles. De même, ont été enlevés en tant que tags catégoriels les signes spécifiques de ponctua- tion, du point d'exclamation au point d'interrogation jusqu'aux parenthèses ou aux guillemets.

Le fichier texte final représentatif d'une année donnée sert de base à la création du graphe, et est composé de phrases dont les mots sont uniquement des lemmes. L'ensemble ressemble à une sorte de télégramme souffrant d'une perte de sens déjà perceptible à sa lecture.

Voici un exemple de deux phrases dans leur forme originale et à leur suite les mêmes phrases dans une forme lemmatisée AVEC ou SANS tags :

La Prévention Spécialisée sur Bègles a vu le jour en 1967.
--

Née d'initiatives privées, elle bénéficie de la collaboration étroite de la Municipalité, du Club de Prévention de Talence et du soutien de la DRASS.

Ces deux phrases sont directement extraites des premiers rapports d'activité.

En voici, une forme intermédiaire qui a éliminé les tags catégoriels mentionnés plus haut.

le prévention spécialisé Bègles avoir voir le jour en 1967.

naître de initiative privé, elle
bénéficiaire de la collaboration
étroite de la municipalité, du
club de prévention de Talence
et du soutien de la DRASS.

Pour terminer, voici la forme lemmatisée finale supportant la suppression des tags grammaticaux, forme intégrée dans le graphe.

prévention spécialisé Bègles
avoir voir jour 1967.

naître initiative privé elle
bénéficiaire collaboration étroite
municipalité club prévention
Talence soutien DRASS.

3.4.2 Les cooccurrences lexicales séquentielles

Les cooccurrences lexicales séquentielles ont été obtenues de chaque phrase dans chaque texte selon la méthode suivante.

L'extraction a été réalisée à l'aide d'une fenêtre mobile de récupération et sa capacité de déplacement est délimitée par la taille de la phrase. Chaque phrase a été parcourue par cette fenêtre mobile, appelée fenêtre de cooccurrence. Cette fenêtre a une taille qui lui permet d'intégrer simultanément 3 mots. Le choix d'une taille trois est inspiré de travaux de Ferrer et Solé [20]. Au cours de leurs travaux, les auteurs, renonçant à une interprétation triviale de la nature des cooccurrences ont fait le choix de se limiter à « des mesures quantitatives de corrélation (entre mots) indépendamment de leur nature » [20]. Il ont choisi pour chaque paire cooccurrence un mot et son voisin phrastique immédiat ou bien un mot et le voisin de son voisin [20]. C'est cette méthode simple et robuste que nous avons élue pour la récolte de nos cooccurrences.

Dans chaque phrase, les mots sont comptabilisés par un processus de coulissement de la fenêtre en accord avec le sens de lecture. Pour une phrase, la fenêtre commence par le début de la phrase, et respectant l'ordre suivant, le premier mot est connecté avec le second, puis le premier avec le troisième, et enfin le second avec le troisième. Puis la fenêtre se déplace vers le mot suivant, elle démarre donc à partir du second. Elle ne mémorise pas le lien entre le second mot et le troisième, puisqu'il a déjà été mentionné pendant la première position. Sont alors mémorisés les liens entre le second et le quatrième mot ainsi que les liens entre le troisième et le quatrième mot. Le déplacement se poursuit jusqu'à la fin de la phrase et enregistre donc pour une phrase de dix mots dix-sept connexions. Le nombre de connexions α pour un phrase composée de X mots se calcule par la formule

$$\alpha = 3 + 2 * (X - 3)$$

Avec ce mode de construction, il n'existe pas de lien entre le dernier mot d'une phrase et le premier de la suivante. Ce processus se poursuit jusqu'à la dernière phrase et s'arrête à la fin du document.

3.4.3 Le mode de construction du graphe

Un ensemble d'arêtes a été constitué pour chaque document annuel. Cette extraction d'arêtes au sein des phrases lemmatisées a été réalisée à l'aide d'un ensemble de traitements informatiques composés en langage Java. Ces arêtes présentent les caractéristiques particulières suivantes :

- Aucune arête n'est assurée d'être la seule représentante de ce type de lien entre deux mots lemmes. Elle peut être présente plusieurs fois en raison de la probabilité de répétition d'une phrase-lemme par document mais elle n'est comptabilisée qu'en mode tout ou rien.
- Une arête ne représente que le lien d'un mot (lemme) à son voisin ou à son « après-voisin » dans le sens de lecture de la phrase. Elle ne représente en aucune façon un lien plus lointain. Cette petite proximité (par séquence) recouvre des liens de nature syntaxique dans environ 70% des cas [21] en raison du mode de captation des liens au sein des phrases. Nous rappelons qu'il n'existe pas de lien entre le dernier mot d'une phrase donnée et le premier de la phrase suivante
- Tous les calculs sur le graphe ont été réalisés sur sa plus grande partie connexe. Cependant, bien que le graphe n'ait pas toujours été totalement connexe, la partie laissée de côté reste parfaitement marginale.
- La dynamique du graphe est perceptible par l'évolution des liens entre les mots qui composent ses cooccurrences, c'est à dire par l'évolution des arêtes.

Le principe de construction du graphe est basé sur les arêtes récupérées depuis les fichiers de phrases-lemmes. Les variations entre le nombre de sommets et le nombre d'arêtes sont présentées dans la figure 5.

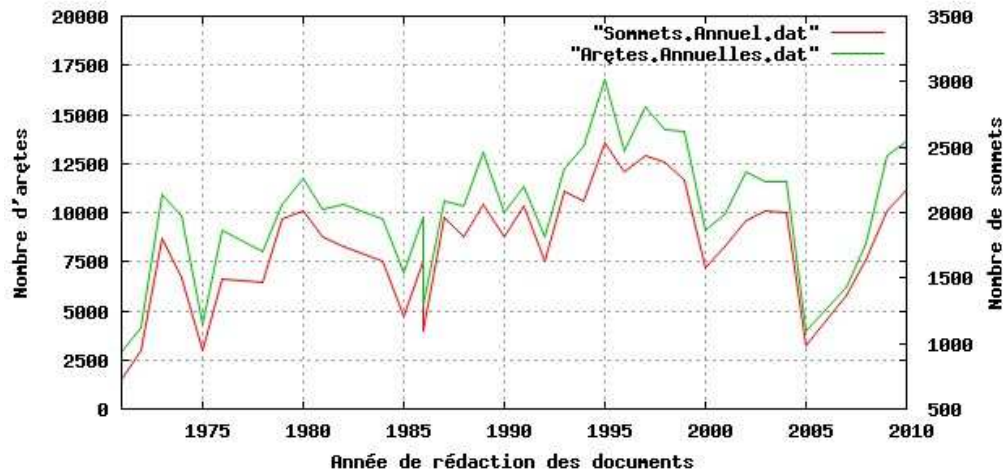


FIG. 5 – Relation entre le nombre d'arêtes, le nombre de sommets, et le temps dans les ensembles annuels

Dans la figure 5, on retrouve superposées les deux courbes exprimant le nombre de sommets et le nombre d'arêtes au cours du temps. Chaque document annuel a conduit à un ensemble de 1000 à 2500 sommets (échelle de droite) pour un ensemble de 3000 à 17000 arêtes (échelle de gauche). Il faut noter la proportionnalité entre les deux courbes, malgré une hétérogénéité des valeurs au cours du temps. Le rapport multiplicatif annuel entre le nombre d'arêtes et le nombre de sommets présente une moyenne de 5.65 avec un écart type d'échantillon de l'ordre de 0.66. Ce rapport annuel est également encadré par un minimum de 4.01 et un maximum de 6.62.

4 La méthodologie du traitement

4.1 La triennalisation

Les graphes annuels construits par la méthode décrite précédemment mettent en valeur des mots, des lemmes, par leurs occurrences dans les textes, par leur degré de connexion aux autres sommets. Pour apprécier ces types d'évolutions statistiques, nous avons souhaité étendre ces mesures à des ensembles pluri-annuels.

Notre méthode a consisté en la production de graphes de trois ans que nous avons nommés graphes triennaux. Le choix de la triennalisation permet en effet une appréciation plus fine de la variation du choix des mots des scripteurs, ainsi que celle de l'évolution de cette dynamique au cours des différentes périodes de rédaction.

La triennalisation est le regroupement de chaque document de l'année X avec les deux documents qui le suivent aux années $X+1$ et $X+2$. Sous une forme plus triviale, c'est la concaténation des fichiers de phrases-lemmes des années respectivement concernées. Cette concaténation n'invalide pas la normalisation des textes mentionnée plus haut. Cette concaténation a généré trente-six regroupements triennaux en appliquant une fenêtre glissante de 3 ans sur les 38 années d'observation. Les graphes triennaux rendent compte des sommets et des arêtes selon la méthodologie de production des graphes annuels.

Les sommets montrent un aspect de la richesse du vocabulaire et les arêtes montrent un aspect de l'intensité des connexions utilisées dans les bilans d'activité. L'orientation des arêtes est probablement ambiguë, car pour reprendre Molière et le maître de philosophie dans le bourgeois gentilhomme, en prose « il faut bien étendre un peu la chose ». Molière fera dire à Monsieur Jourdain qu'il pourra dire belle Marquise, vos beaux yeux me font mourir d'amour. Ou bien : d'amour mourir me font, belle Marquise, vos beaux yeux. Ou bien : vos yeux beaux d'amour me font, belle Marquise, mourir. Ou bien :

mourir vos beaux yeux, belle Marquise, d'amour me font. Ou bien : me font vos yeux beaux mourir, belle Marquise, d'amour.

Molière dans son texte a joué avec l'ordre d'apparition des mots autorisé par la prose (Molière, *Le Bourgeois Gentilhomme*, Acte II, Scène IV).

De même, l'antéposition ou la postposition d'adjectifs qualificatifs, de participes passés, de participes présents au voisinage d'un nom, montrent des usages inhabituels. Ces usages laissent une place à la possibilité d'une expression subjective voulant privilégier le propre et/ou le figuré. L'ordre d'apparition des mots dans la phrase détient, d'un point de vue pragmatique, des vertus illocutoires pour informer et convaincre, voire croire transformer la situation.

Les arêtes des graphes triennaux ont donc été symétrisées dans cette perspective pour construire un graphe non orienté. Dans chaque triennal, la concaténation des phrases-lemmes issues de chaque année maintient un risque de répétition multiple des arêtes et donc des sommets par la coprésence probable de phrases similaires. Les arêtes ont donc été rendues uniques dans chaque triennal pour mettre en valeur la seule cooccurrence des mots. Les arêtes peuvent être réflexives en particulier dans le cadre d'un texte lemmatisé. Par exemple, « j'ai eu » lemmatisé deviendra « je avoir avoir » et générera entre autres une arête « avoir avoir ».

Dans la figure 6, on retrouve superposées les deux courbes exprimant le nombre de sommets et le nombre d'arêtes au cours du temps. Ces valeurs concernent les 36 ensembles triennaux composés pour lisser dans le temps l'observation du choix et de l'association des termes contenus dans les documents étudiés. Chaque ensemble triennal comprend entre 2000 et 4000 sommets (échelle de droite) pour un volume de 15000 à 40000 arêtes (échelle de gauche). Sur la figure 6, les échelles des ordonnées ont été représentées aux valeurs doubles de celles de la figure 5. En effet, il est possible d'observer qu'il y a peu de variations entre le nombre de sommets et le nombre d'arêtes, les courbes sont visuellement corrélées et l'écart entre les deux est relativement petit et il n'y a d'ailleurs pas vraiment de variation d'un triennal à un autre.

Le nombre de sommets de chaque triennal est relativement proche de son nombre d'arêtes, modulo le changement d'échelle. Si on compare la figure 6 avec la figure 5, il est possible d'observer un apaisement dans le rapport entre le nombre de sommets et le nombre d'arêtes. Il est envisageable de pouvoir apprécier un phénomène de lissage propre à la méthodologie triennale.

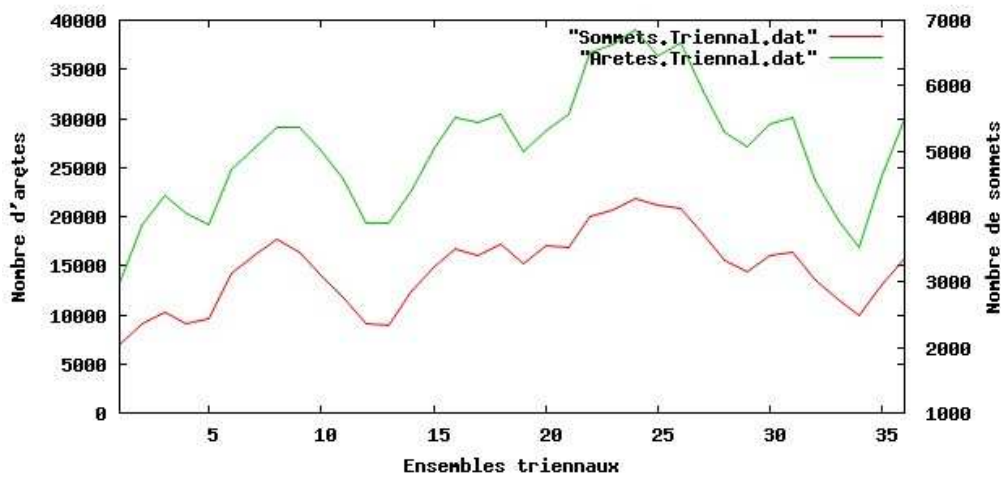


FIG. 6 – Relation entre le nombre d'arêtes, de sommets, dans les triennaux et le temps

Les métriques des graphes ont été calculées et rapportées uniquement à la partie connexe la plus grande du graphe triennal.

Le mode de calcul du coefficient d'agrégation provient d'un article qui fait référence [42]. Ce coefficient est également issu de l'étude des graphes sociaux et il tente de mesurer la tendance du voisinage immédiat d'un sommet

à posséder une tendance à l'interconnexion. Le calcul est réalisé à l'aide de la moyenne des coefficients d'agrégation c'est à dire la moyenne des coefficients calculables. Quand un sommet n'a qu'un seul voisin, son coefficient d'agrégation est incalculable si on utilise le mode de calcul de la partie théorique. Le nombre de feuilles correspond au nombre de sommets qui n'ont qu'un seul voisin. La taille du nombre de coefficients non calculables est immédiatement en lien avec le nombre de sommets à un seul voisin. Si un sommet a deux ou plusieurs voisins non connectés, on obtiendra un coefficient égal à zéro. Ces mesures sont accessibles dans les résultats.

La distance caractéristique a été calculée à l'aide d'un algorithme en recherche de largeur (Breadth First Search). Cet algorithme cherche à épuiser la liste des sommets voisins du sommet de départ avant de poursuivre l'exploration du graphe. Il permet de trouver le plus court chemin entre le sommet de départ et tous les autres sommets accessibles. Le diamètre est partiellement dépendant de la distance caractéristique, puisqu'il est la plus grande distance caractéristique extractible de l'ensemble des plus courts chemins entre les sommets. La valeur de la distance caractéristique d'un graphe aléatoire, plutôt faible, est limitée par le logarithme népérien du nombre de ses sommets divisé par le logarithme népérien du produit entre la probabilité de connection et le nombre de sommet.

Ces mesures, distance caractéristique et coefficient d'agrégation, sont accessibles dans les résultats et ont été fournis par les programmes spécifiques que nous avons développés pour traiter les données observées.

4.2 Le corpus de comparaison et le test statistique

Nous nous sommes intéressés aux fréquences relatives des mots utilisés dans les documents professionnels pour tester un lien entre la fréquence d'un usage professionnel et la fréquence d'un usage non spécifique. Ce ratio a été apprécié sur tous les sommets des graphes triennaux à l'aide d'une comparaison.

Cette comparaison a consisté à comparer les fréquences relatives des lemmes de chaque document avec celles fournies par un corpus présent sur le

site de Lexique 3. Lexique 3 est une base de données lexicales libre [36]. Notre comparaison a pris en compte la release 3.71 de Lexique 3 parue en 2011. La page d'accueil de Lexique 3 déclare que « cette version de Lexique 3 concerne 135 000 mots du français [61] ». On peut supposer qu'il s'agit spécifiquement de formes puisque avec les « mots » sont fournis les représentations orthographiques et phonémiques, la syllabation, la catégorie grammaticale, le genre et le nombre, les fréquences fléchies et les fréquences des lemmes associés. Parmi les bases de données libres, celle de Lexique 3 offre donc surtout la fourniture de fréquences de lemmes.

Cependant, les fréquences de lemmes de Lexique 3 sont basées sur deux corpus construits à partir de deux méthodes qui sont différentes. La première utilise les documents écrits alors que la seconde utilise des sous-titres de films [64], ce qui fournit une intéressante approximation des fréquences de mots davantage rapportée aux interactions verbales humaines.

La première méthode s'appuie sur l'utilisation de documents écrits fournis par la base de l'Atilf, et elle est composée d'un corpus de textes littéraires. Les textes concernés sont récents, ils proviennent de Frantext. Il s'agit de 218 romans publiés entre 1950 et 2000. Ces livres représentent un corpus de 14.7 millions d'items filtrés par une méthodologie qui est déclarée de bonne qualité sur le site Lexique.Org : logiciel Aspell, dictionnaires Français-Gutenberg 1.0 et Le Grand Robert 1996.

La seconde méthode repose sur l'utilisation d'un grand nombre de sous-titres de films pour approximer les fréquences de mots dans les interactions humaines. Lexique 3 a compilé 52 millions de mots français provenant d'une grande variété de films français, anglais ou américains, européens quoique non-anglais, ainsi que de séries anglo-américaines [64]. Cette compilation concerne 9474 films et séries généralement de langue anglaise en rapport avec leur situation dominante dans cette industrie.

Nous avons choisi de privilégier la seconde méthode de construction du corpus des lemmes de Lexique 3 pour comparer des fréquences de mots issues de professionnels pratiquant essentiellement l'oral au travail. Les textes numérisés à partir de sous-titres de films ont fourni un accès complémentaire aux mesures de fréquences des textes écrits. C'est une possibilité hautement

corrélée à ce qui peut être obtenu à partir de sources sérieuses et également à ce qui peut être obtenu à partir des fréquences écrites produites avec les textes sélectionnés par Lexique 3 dans Frantext [64].

Ainsi, dans le tableau récapitulatif présenté par le manuel de Lexique 3, nous avons retenu les troisièmes et septièmes colonnes. Celles-ci concernent respectivement les lemmes (l’infinitif pour un verbe, le singulier pour un nom et le masculin singulier pour un adjectif) et leurs fréquences selon le corpus de sous-titres par million d’occurrences.

Nous avons extrait par calcul de nos propres documents les occurrences de lemmes dans chaque triennal. Lexique 3 présente de son côté la fréquence d’un lemme donné par million en référence au corpus nommé « `freqlemfilm2` ». La comparaison des occurrences de lemmes, par triennal ou par million en corpus « `freqlemfilm2` », a été réalisée à l’aide d’un test statistique.

Pour la réalisation de cette mesure, un « test de Fisher exact » a été retenu.

Son exécution sur notre matériel a été réalisée par le logiciel R qui propose en freeware un langage et un environnement de calculs statistiques classiques [57]. Ce test statistique permet de tester si des fréquences entières observées sur deux échantillons sont identiques ou non, même si les effectifs sont faibles, ce qui est notre cas pour les ensembles triennaux.

Par défaut, le test de Fisher est sans a priori sur quel échantillon a une plus grande fréquence que l’autre, mais on peut tester seulement un sens ou l’autre.

Ici, nous avons choisi de définir H_0 comme l’hypothèse soutenant que la fréquence des lemmes de nos textes est inférieure à la fréquence présentée dans Lexique 3. Chaque fois que cette possibilité est avérée par calcul via le test de Fisher, on rejette cette hypothèse ce qui nous permet de ne retenir que son inverse, c’est à dire seulement les cas où les fréquences professionnelles sont significativement supérieures à l’usage commun au risque de 0.05.

L’explication suivante concerne le choix de l’application du test exact de Fisher. Le choix de ce test relève d’une problématisation spécifique qui est traduisible par un tableau de contingence à quatre cases. Ce qui fait

automatiquement penser au test du χ^2 et à sa capacité à traduire dans une réponse en tout ou rien statistique, c'est à dire en oui ou en non, la question de la plus grande fréquence d'un usage lexical professionnel.

Dans un tableau de contingence à quatre cases, désirant répondre à la question de la fréquence relative, chaque lemme est représenté en colonne par sa fréquence entière, c'est à dire son occurrence et par sa non-occurrence dans le cadre de la référence choisie, c'est-à-dire soit par ensemble triennal soit par corpus des sous-titres [64]. La fréquence dans le corpus des sous-titres est fournie en occurrences par million. Chaque lemme triennal a donc été évalué en comparaison avec la fréquence d'effectif par million de Lexique 3. Il est ainsi possible que la fréquence de Lexique 3 soit fournie dans son corpus avec des nombres non entiers mais R utilise ici un script qui arrondi automatiquement cette fréquence. De plus, outre cet arrondi de fréquence, le script R effectue à travers une matrice générale et pour chaque lemme du triennal, une somme globale pour déterminer le nombre de ses présents et de ses absents.

	Triennal 1	Lexique 3
autorité	13	32
tout autre mot candidat	12 116	999968

TAB. 2 – Tableau de contingence

Le tableau 2 présente dans sa deuxième ligne le lemme « autorité » et son occurrence dans le triennal concerné ou son occurrence par million dans le corpus de Lexique 3. Dans sa troisième et dernière ligne, sont affichées les occurrences des autres lemmes dans le triennal ou dans le corpus de comparaison, ce qui limite obligatoirement la portée de la sélection aux lemmes inclus dans Lexique 3. En l'occurrence, alors que le premier triennal est composé de 2031 sommets différents, seuls 1829 sommets étaient présents en tant que lemmes dans le corpus de Lexique 3. En effet, les noms propres ou les sigles sont absents du corpus généraliste.

Les tableaux 3 et 4 récapitulent en exemple pour deux triennaux, le premier et le vingt-troisième, des lemmes retenus ou rejetés, suivant leurs occurrences triennales et la réponse au test exact de Fisher sur les présences respectives telles que l'on peut les trouver dans Lexique 3. Ces tableaux fournissent dans leurs deux dernières colonnes la p-value et la réponse à la question du rejet de H_0 .

Les lemmes disposant de réponses « TRUE » ont donc été retenus. Les occurrences par million de Lexique 3 sont présentées ici telles qu'elles sont fournies dans le corpus dans la colonne « freqlemfilm2 », mais nous rappelons qu'elles sont arrondies automatiquement par le script R.

Lemme	Présence	Présence Lexique 3	p-value	rejet H0
ville	4	295.14	43.49	FALSE
autorité	13	32.95	0	TRUE
étude	7	44.22	0	TRUE
autonomie	2	2.76	0.12	TRUE

TAB. 3 – Réduction par sur-présence dans le Triennal 1

Lemme	Présence	Présence Lexique 3	p-value	rejet H0
ville	22	295.14	0.03	TRUE
autorité	1	32.95	64.2	FALSE
étude	7	44.22	0.08	TRUE
autonomie	15	2.76	0	TRUE

TAB. 4 – Réduction par sur-présence dans le Triennal 23

Sélectionner sur un ensemble triennal de fichiers lemmes une liste de mots par une comparaison significative de fréquence est une méthode qui tend à réduire l'ensemble des mots disponibles pour la présentation de la nature professionnelle de l'écriture. Outre une telle tendance de réduction, cette méthode limite les diverses formes de saillances lexicales, que celles-ci appartiennent aux critères formels (prosodie au sens large, syntaxe dans la structure de l'énoncé) ou cognitifs (récence, mémoire, familiarité) ou qu'elles soient inhérentes à la mise en corpus de compte rendu.

Si l'on prend par exemple le premier ensemble triennal composé des années 1972-1973-1974. Il correspond à un ensemble de 86 pages dactylographiées (cf annexe 1) dont la concaténation conduit à 2031 lemmes différents. La réduction effectuée par le test de Fischer ramène les 2031 lemmes différents à 1175 lemmes retenus, soit une baisse de 58%. Il est possible d'avoir une appréciation de cette réduction pour les 36 ensembles triennaux de l'ensemble de la période à travers les tableaux 6 et 7 d'une part, et 9 et 10 d'autre part, présentés plus loin.

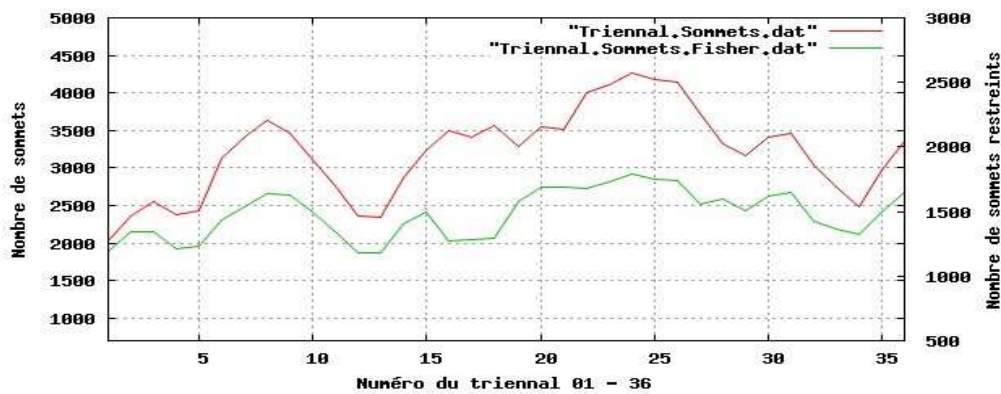


FIG. 7 – Relation dans le temps entre le nombre de sommets des ensembles triennaux et le nombre de sommets des ensembles triennaux restreints

La figure 7 présente deux courbes superposées exprimant au cours du temps d'une part le nombre de sommets composant les ensembles triennaux d'observation (échelle de gauche) et d'autre part le nombre de sommets composant les ensembles triennaux restreints par le test de Fisher (échelle de droite). Une différence d'échelle entre ces deux courbes permet de ne pas aplatir la courbe du bas, et de lui faire présenter une certaine variation dans le temps.

Même si l'allure générale de ces deux courbes est assez similaire, certaines périodes (15-20 par exemple) montrent entre elles des différences d'amplitude et d'inflexion. En effet, si le ratio de ces deux quantités présente en réduction avec le test de Fisher une moyenne de 2.15 et un écart type d'échantillon de 0.24, ce ratio change nettement pour les périodes triennales 16, 17 et 18, localisant à ce moment précis la présence de ses trois maximums [2.75, 2.67, 2.76]. Ce ratio est encadré par son maximum 2.76 et son minimum 1.70. Il est également possible de retrouver un autre point d'amplitude avec des variations moindres sur les périodes triennales 22 à 27, venant localiser un ratio compris de 2.37 à 2.39. Les périodes triennales 16 à 18 recouvrent les années 1987 à 1991 et les périodes triennales 22 à 27 concernent celles entre 1993 et 2000.

4.3 Les deux enquêtes auprès des professionnels

Notre recherche s'appuie sur un ensemble de graphes lexicaux triennaux élaborés à partir d'écrits. Ces écrits sont spécifiques dans la mesure où ils témoignent d'une activité professionnelle. Les auteurs de ces documents exercent leur travail dans une pratique essentiellement orale. L'activité annuelle est toutefois rythmée par ces écrits qu'il s'agit de renouveler régulièrement. Nous nous sommes donc penchés sur la perception qu'ont les auteurs sur leurs bilans d'activité. Notre intention est de trouver dans une interaction orale avec les professionnels du secteur (des auteurs potentiels des documents étudiés) des marqueurs ou des points focaux à observer dans les graphes lexicaux constitués.

Nous avons mené une enquête en deux phases, espacées de plusieurs mois, avec l'aide de deux questionnaires d'entretiens.

La question de la première enquête cherche à percevoir les représentations singulières des professionnels sur les bilans d'activité. En d'autres termes, de quoi parlent les bilans d'activité ? Est donc posée aux professionnels la question suivante :

« A partir de votre expérience, quels sont selon vous les thèmes abordés de façon récurrente dans les bilans d'activité (au plus douze thèmes) ? Classez les par ordre d'importance décroissante ».

Une synthèse des réponses collectées à l'aide cette première question a été ensuite élaborée. Cette synthèse permet de déterminer une liste de thèmes potentiels en relation avec l'exercice professionnel.

C'est cette liste qui est ensuite proposée aux professionnels dans une nouvelle question, posée à plusieurs mois d'intervalle de la première. Cette seconde phase d'enquête a pour but d'aborder la présence potentielle de constellations de mots cooccurrents d'une thématique spécifique. La question suivante est posée :

« A partir de votre expérience professionnelle et dans le cadre de l'écriture de votre bilan annuel d'activité, pouvez-vous fournir un ensemble d'une dizaine de mots (noms, verbes ou adjectifs) que vous croyez pouvoir associer dans ce bilan à chacun des thèmes suivants » :

travail de rue	actions collectives	filles	précarité	scolarité
errance	activités	chantiers	familles	partenariat

Comme pour la question de la première phase de l'enquête, nous avons utilisé un questionnaire sur papier qui a été rempli sous une quinzaine de jours.

5 Les résultats

5.1 Les métriques sur les ensembles triennaux

Cette partie présente les résultats obtenus à partir des calculs menés sur les graphes triennaux. Chaque graphe triennal modélise le vocabulaire employé dans les comptes rendus d'activité de trois années glissantes. Les sommets du graphe sont composés des mots utilisés et les arêtes symbolisent une proximité lexicale des mots dans les textes.

Pour rapporter ici ces résultats, nous avons adopté les désignations suivantes. La lettre « T », suivie d'un numéro, désigne un graphe triennal. Il y en a 36 au total, allant de T01 (pour les trois premières années 1972, 1973, 1974) à T36 (pour les années 2008, 2009, 2010). L'année 1986 ayant produit deux documents et l'année 2006 n'ayant pas donné lieu à la rédaction d'un compte rendu, c'est bien 36 triennaux que nous avons à traiter pour trente-huit comptes rendus. La lettre « S » désigne les sommets du graphe ; la lettre « A » désigne ses arêtes. La composante principale d'un graphe (sa plus grande composante connexe) étant celle sur laquelle les calculs peuvent être menés, les sommets connexes de cette composante principale (SC) sont distingués des sommets non connexes (SNC) ; et parmi les premiers, les sommets connexes feuille (SCF) sont distingués des sommets connexes non feuille (SCNF), pour autoriser les calculs des coefficients de connectivité. AC désigne les arêtes de la principale composante connexe. Toute cette terminologie associée aux variables et à la présentation de leurs valeurs est rappelée dans le tableau 5.

T	Le triennal avec son numéro
S	Le nombre de sommets
A	Le nombre d'arêtes capturées
AC	Le nombre d'arêtes de la composante connexe majeure
SC	Le nombre de sommets de la composante connexe majeure
SNC	Le nombre de sommets hors de la composante connexe majeure
SCF	Le nombre de sommets connexes n'ayant qu'un seul voisin (sommets feuilles)
SCNF	Le nombre de sommets connexes non feuilles

TAB. 5 – Signification des noms de variables employés dans les tableaux de résultats 6 et 7.

La concaténation des 36 graphes triennaux constitue un graphe général composé de 10893 sommets distincts. Il s'agit là de l'ensemble du vocabulaire employé sur la période étudiée, une fois retirés les prépositions, articles et autres lemmes non retenus (cf 3.4). Faire porter cette concaténation sur les graphes triennaux procure le même résultat que si elle est menée sur les graphes annuels ; le vocabulaire est bien le même, et les doublons générés par le recouvrement des triennaux successifs sont retirés. Pour chacun de ces 10893 sommets, nous avons attribué un identificateur unique, de manière à obtenir une référence univoque pour chaque lemme manipulé.

Les tableaux 6 et 7 dénombrent les arêtes et les sommets des différents graphes considérés. Le tableau 6 porte sur les 18 premiers graphes triennaux, et le tableau 7 sur les 18 suivants. On y retrouve les éléments globaux de chaque graphes (S et A), ainsi que les éléments de la composante connexe principale de chacun (SC, AC) en mentionnant le nombre de sommets écartés (SNC), puis les sommets impliqués dans le « centre de gravité » du graphe (SCNF), et ceux relatifs à la « périphérie » du graphe (SCF).

Les graphes triennaux regroupent chacun entre 2000 et 4000 sommets (moyenne de 3202 ; écart-type de 585), pour un total de 13000 à 39000 arêtes

(moyenne de 26900 ; écart-type de 6246). Ceux ayant le plus de sommets sont bien évidemment ceux qui présentent le plus d'arêtes (cf figure 8, où chaque graphe triennal est symbolisé par un point). Ce rapport (nombre d'arêtes par sommet) calculé pour chacun des graphes triennaux varie autour d'une moyenne de 8.33.

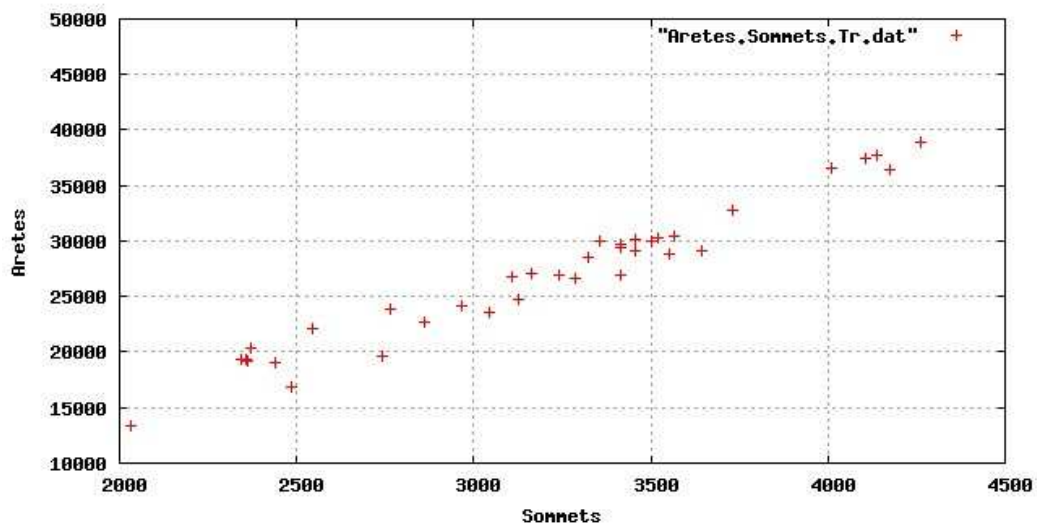


FIG. 8 – Nombre d'arêtes par sommets pour les graphes triennaux

La figure 8 est un nuage de points présentant, pour chaque ensemble triennal, le rapport existant entre le nombre de sommets et le nombre d'arêtes de cet ensemble. Ce rapport multiplicatif présente une moyenne de 8.33 avec un petit écart type de 0.6, et un déplacement du minimum au maximum qui va de 6 à 9.

A chaque étape de calcul, un certain nombre de sommets sont laissés de côté (SNC puis SCF), mais il s'agit à chaque fois d'un effectif très faible, comparé au nombre de sommets total de chaque graphe. Le nombre de sommets feuilles varie entre un minimum de 2 et un maximum de 20, avec une moyenne autour de 9 et une médiane autour de 8. De même, la taille de la composante non connexe est extrêmement petite, avec une grande variabilité entre les graphes (d'un minimum de 0 à un maximum de 26 sommets concernés). Le nombre moyen de sommets non connectés est proche de la valeur médiane (moyenne : 10.6 ; médiane 9.5).

Ce faible nombre de sommets laissés de côté, n'affecte pas beaucoup le nombre d'arêtes mises à l'écart. Ainsi les colonnes A et AC ont sensiblement les mêmes valeurs, et le fait de retirer un SCF ne va à chaque fois retirer qu'une seule arête (cf tableaux 6 et 7).

T	S	A	SNC	SC	AC	SCF	SCNF
T_01	2031	13297	0	2031	13297	4	2027
T_02	2363	19236	0	2363	19236	3	2360
T_03	2546	22128	2	2544	22127	4	2540
T_04	2370	20370	2	2368	20369	4	2364
T_05	2438	19104	8	2430	19097	8	2422
T_06	3128	24774	2	3126	24773	9	3117
T_07	3415	26967	9	3406	26961	8	3398
T_08	3644	29074	11	3633	29064	6	3627
T_09	3456	29079	13	3443	29068	9	3434
T_10	3107	26849	11	3096	26837	10	3086
T_11	2763	23863	11	2752	23854	7	2745
T_12	2359	19327	13	2346	19316	7	2339
T_13	2345	19311	8	2337	19304	6	2331
T_14	2862	22655	15	2847	22643	18	2829
T_15	3241	26956	16	3225	26942	18	3207
T_16	3503	30003	11	3492	29993	11	3481
T_17	3417	29662	26	3391	29643	4	3387
T_18	3567	30491	22	3545	30477	12	3533

Table 6: Caractéristiques générales des graphes triennaux T01 à T18

Les en-têtes de colonnes sont les suivantes :

T_xx : Triennal_xx

S : Sommets

A : Arêtes

SNC : Sommets non connexes

SC : Sommets connexes

AC : Arêtes connexes

SCF : Sommets connexes feuilles

SCNF : Sommets connexes non feuilles

T	S	A	SNC	SC	AC	SCF	SCNF
T_19	3285	26664	20	3265	26654	13	3252
T_20	3552	28782	16	3536	28774	20	3516
T_21	3521	30359	18	3503	30347	19	3484
T_22	4012	36617	8	4004	36611	11	3993
T_23	4106	37496	9	4097	37490	11	4086
T_24	4264	38964	23	4241	38945	14	4227
T_25	4175	36355	19	4156	36341	13	4143
T_26	4136	37699	15	4121	37687	19	4102
T_27	3730	32754	7	3723	32749	10	3713
T_28	3323	28584	10	3313	28576	9	3304
T_29	3161	27110	8	3153	27106	7	3146
T_30	3415	29411	6	3409	29408	8	3401
T_31	3458	30160	2	3456	30159	6	3450
T_32	3045	23628	7	3038	23624	4	3034
T_33	2740	19639	14	2726	19627	3	2723
T_34	2486	16838	7	2479	16830	2	2477
T_35	2967	24202	7	2960	24197	2	2958
T_36	3355	29998	7	3348	29993	2	3346

TAB. 7 – Caractéristiques générales des graphes triennaux T19 à T36

Les en-têtes de colonnes sont les suivantes :

T_xx : Triennal_xx

S : Sommets

A : Arêtes

SNC : Sommets non connexes

SC : Sommets connexes

AC : Arêtes connexes

SCF : Sommets connexes feuilles

SCNF : Sommets connexes non feuilles

T	DC	D	CA	T	DC	D	CA
T_01	2.88	6	0.46	T_19	2.87	9	0.50
T_02	2.77	6	0.46	T_20	2.86	9	0.50
T_03	2.75	6	0.47	T_21	2.78	7	0.51
T_04	2.75	6	0.48	T_22	2.76	7	0.52
T_05	2.79	7	0.49	T_23	2.76	7	0.52
T_06	2.80	8	0.51	T_24	2.76	7	0.52
T_07	2.79	8	0.52	T_25	2.79	7	0.51
T_08	2.80	8	0.51	T_26	2.78	8	0.51
T_09	2.80	6	0.50	T_27	2.79	9	0.50
T_10	2.81	7	0.49	T_28	2.79	9	0.50
T_11	2.81	9	0.50	T_29	2.79	9	0.51
T_12	2.81	9	0.49	T_30	2.81	8	0.51
T_13	2.81	9	0.49	T_31	2.81	7	0.50
T_14	2.90	9	0.49	T_32	2.85	8	0.49
T_15	2.89	8	0.49	T_33	2.87	8	0.48
T_16	2.87	9	0.48	T_34	2.90	7	0.47
T_17	2.83	10	0.49	T_35	2.86	7	0.48
T_18	2.87	10	0.49	T_36	2.83	7	0.48

TAB. 8 – Métriques petit-monde de la composante connexe principale des triennaux 1 à 36

Le tableau 8 reprend les valeurs associées aux métriques petit-monde calculées sur chaque graphe triennal. En colonne, DC désigne la distance caractéristique, D désigne le diamètre et CA désigne le coefficient d'agrégation moyen. La distance caractéristique et le diamètre se mesurent sur la plus grande composante connexe du graphe ; le coefficient de connexité (ou coefficient d'agrégation) nécessite de ne considérer que les sommets non feuilles

(SCNF) puisque le coefficient d'agrégation est incalculable sur les sommets feuilles.

Les distances caractéristiques de tous les triennaux sont très homogènes entre tous ces graphes, sans dépendre de la taille des graphes (nombres de sommets, nombre d'arêtes). Ces distances caractéristiques valent en moyenne 2.81, avec un écart type de l'ordre de 0.04. Il en va de même pour les coefficients d'agrégation qui évoluent autour d'une moyenne de 0.49 et pour un écart type de 0.01. Le diamètre est finalement la métrique qui présente le plus de fluctuation, avec un maximum de 10 et un minimum de 6. Mais là aussi les valeurs sont relativement homogènes avec une moyenne de 7.8 et une médiane de 8.

Toutes ces métriques sont donc homogènes dans leurs valeurs et cohérentes avec les caractéristiques attendues pour les graphes petits-mondes.

5.2 Les graphes triennaux restreints

Il est possible d'apprécier la réduction produite par le test exact de Fisher sur le nombre de sommets retenus pendant la construction du graphe à l'aide du graphique de la figure 9. Cette figure présente la courbe du rapport calculé entre le nombre de sommets d'un ensemble triennal restreint et le nombre de sommets de l'ensemble triennal correspondant. Il s'agit donc du rapport entre les deux courbes de la figure 7, fluctuant entre les valeurs de 58% (pour TR01/T01) et de 36% (pour TR16/T16 ou TR18/T18). La valeur moyenne de 46% est reportée sur un axe horizontal.

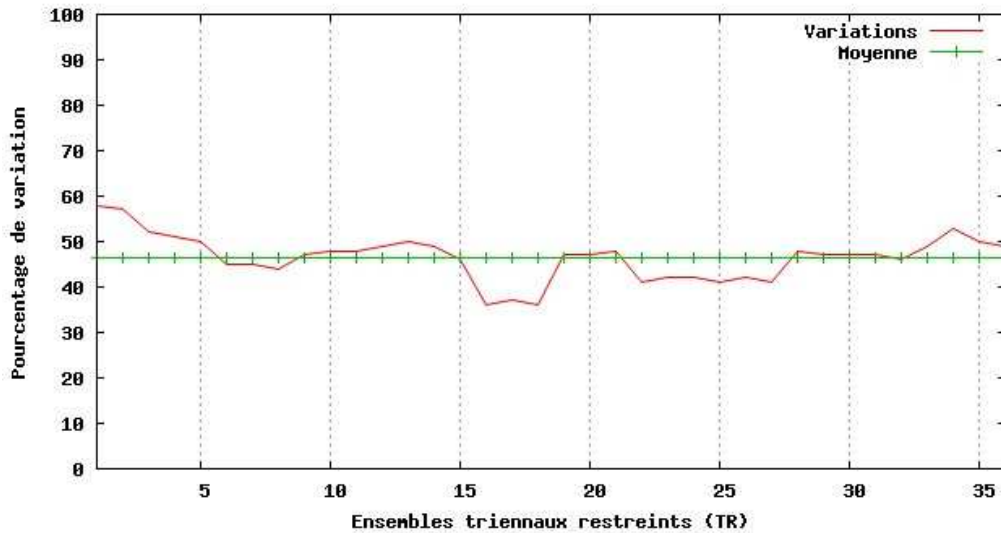


FIG. 9 – Pourcentage temporel de réduction du nombre de sommets des graphes triennaux par application du test exact de Fisher

Les tableaux 9 et 10 ci-dessous permettent les observations suivantes.

La dénomination « TR » signifie la présence de la restriction due à l'application du test exact de Fisher. Les tableaux 9 et 10 peuvent se lire de la même façon que les tableaux 6 et 7 sachant que les intitulés ont intégré la restriction mentionnée. Ainsi, hormis la première, toutes les colonnes des tableaux neuf et dix concernent le dénombrement des sommets et des arêtes. Les colonnes qui dénombrent les sommets commencent par la lettre S et celles qui dénombrent les arêtes commencent par la lettre A.

Chaque type de sommet est différent. Les sommets de type S désignent des sommets qui ne présentent pas d'autre caractéristique que celle d'appartenir au graphe. Les deux autres types de sommets issus de la catégorie S présentent des spécifications en rapport avec le fait d'appartenir ou pas

à la composante connexe majeure : les sommets SC y appartiennent et les sommets SNC sont en dehors de cette composante. Enfin, les deux dernières colonnes précisent la position du sommet par rapport au centre de gravité du graphe. La colonne SF (sommets feuilles) désigne le nombre de sommets qui sont à la périphérie de la composante connexe majeure. Ces sommets non feuilles, comme les sommets feuilles appartiennent à la composante connexe.

Les sommets feuilles présentent un minimum de 7 pour un maximum de 93, avec une moyenne autour de 56.7, une médiane à 60 et un écart type à 19.6. La composante non connexe est petite, avec une moyenne assez proche de la médiane (moyenne : 2.36 ; médiane 2). Elle s'étale de 0 au minimum pour atteindre un maximum de 9, avec un écart type à 2.48. Le nombre de sommets est homogène allant d'un minimum de 1167 à un maximum de 1779 avec un écart type de 183.

Les arêtes présentent un minimum de 6364 et un maximum de 18 442 avec une moyenne à 12 760 et une médiane à 12 960 ainsi qu'un écart type dans cet échantillon de 2928. Le rapport entre nombre de sommets et nombre d'arêtes est plutôt constant avec un rapport multiplicatif de 8.63 en moyenne et une médiane à 8.53. Il est possible de rapprocher ce rapport avec le rapport multiplicatif obtenu à partir du graphe non restreint dans les tableaux 6 et 7 [moyenne 8.33, médiane 8.55].

TR	S	A	SNC	SC	AC	SCF	SCNF
TR_01	1175	6364	9	1166	6358	58	1108
TR_02	1336	9374	4	1332	9372	48	1284
TR_03	1334	11119	2	1332	11118	34	1298
TR_04	1202	10157	0	1202	10157	37	1165
TR_05	1218	8616	7	1211	8612	56	1155
TR_06	1406	10626	4	1402	10624	67	1335
TR_07	1528	11913	4	1524	11911	70	1454
TR_08	1620	12710	4	1616	12708	76	1540
TR_09	1610	13279	0	1610	13279	93	1517
TR_10	1481	13107	0	1481	13107	87	1394
TR_11	1330	11995	4	1326	11993	61	1265
TR_12	1167	9040	8	1159	9036	57	1102
TR_13	1173	9156	4	1169	9154	49	1120
TR_14	1393	10945	4	1389	10943	72	1317
TR_15	1477	12776	0	1477	12776	59	1418
TR_16	1268	14143	0	1268	14143	8	1260
TR_17	1279	14246	0	1279	14246	10	1269
TR_18	1288	14642	0	1288	14642	7	1281

TAB. 9 – Tableau récapitulatif des graphes triennaux restreints 1-18

Les en-têtes de colonnes sont les suivantes :

TR : Triennal restreint xx

S : Sommets

A : Arêtes

SNC : Sommets non connexes

SC : Sommets connexes

AC : Arêtes connexes

SCF : Sommets connexes feuilles

SCNF : Sommets connexes non feuilles

TR	S	A	SNC	SC	AC	SCF	SCNF
TR_19	1557	13052	4	1553	13050	60	1493
TR_20	1672	13813	4	1668	13811	66	1602
TR_21	1680	15073	0	1680	15073	70	1610
TR_22	1671	17249	1	1670	17248	37	1633
TR_23	1721	17800	1	1720	17799	46	1674
TR_24	1779	18442	1	1778	18441	56	1722
TR_25	1732	16947	0	1732	16947	60	1672
TR_26	1718	17217	0	1718	17217	55	1663
TR_27	1541	14881	2	1539	14880	42	1497
TR_28	1575	13580	0	1575	13580	60	1515
TR_29	1487	12869	0	1487	12869	67	1420
TR_30	1608	14823	0	1608	14823	74	1534
TR_31	1629	14762	6	1623	14758	68	1555
TR_32	1412	11082	4	1408	11079	68	1340
TR_33	1338	8965	2	1336	8963	65	1271
TR_34	1305	8096	2	1303	8095	72	1231
TR_35	1473	12153	2	1471	12152	53	1418
TR_36	1622	14368	2	1620	14367	76	1544

TAB. 10 – Tableau récapitulatif des graphes triennaux restreints 19-36

Les en-têtes de colonnes sont les suivantes :

TR : Triennal restreint xx

S : Sommets

A : Arêtes

SNC : Sommets non connexes

SC : Sommets connexes

AC : Arêtes connexes

SCF : Sommets connexes feuilles

SCNF : Sommets connexes non feuilles

TR	DC	D	CA	TR	DC	D	CA
TR_01	2.98	7	0.38	TR_19	2.72	7	0.39
TR_02	2.85	8	0.38	TR_20	2.75	7	0.39
TR_03	2.64	7	0.41	TR_21	2.63	6	0.44
TR_04	2.61	7	0.43	TR_22	2.52	6	0.43
TR_05	2.82	7	0.39	TR_23	2.52	6	0.42
TR_06	2.79	8	0.40	TR_24	2.54	6	0.44
TR_07	2.69	10	0.44	TR_25	2.57	6	0.41
TR_08	2.70	7	0.43	TR_26	2.65	7	0.38
TR_09	2.78	7	0.39	TR_27	2.63	7	0.37
TR_10	2.72	6	0.38	TR_28	2.72	7	0.39
TR_11	2.68	6	0.40	TR_29	2.72	7	0.40
TR_12	2.75	8	0.39	TR_30	2.62	7	0.43
TR_13	2.76	8	0.39	TR_31	2.71	7	0.39
TR_14	2.79	7	0.40	TR_32	2.76	7	0.38
TR_15	2.73	7	0.38	TR_33	2.84	7	0.38
TR_16	2.50	6	0.34	TR_34	2.90	7	0.38
TR_17	2.50	7	0.32	TR_35	2.78	6	0.38
TR_18	2.47	5	0.32	TR_36	2.76	7	0.39

TAB. 11 – Métriques des graphes triennaux restreints 1-36

Le tableau 11 reprend les valeurs associées aux métriques petit monde calculées sur chaque graphe triennal restreint. En colonne, DC désigne la distance caractéristique, D désigne le diamètre et CA le coefficient d'agrégation. La distance caractéristique et le diamètre ont été appréciés sur la plus grande composante connexe du graphe. Le coefficient de connexité (ou coefficient d'agrégation) nécessite de ne considérer que les sommets non feuilles (SCNF) puisque le coefficient d'agrégation est incalculable sur les sommets feuilles.

Les distances caractéristiques de tous les triennaux restreints sont cohérentes par leurs tailles avec un graphe de type petit monde. Elles présentent une moyenne de 2.69, une médiane de 2.72, toutes deux associées à un écart type de l'ordre de 0.11.

Les coefficients d'agrégation ont une moyenne et une médiane de 0.39, ainsi qu'un écart type de 0.02. Il est intéressant de noter la diminution de ces coefficients si on les compare avec ceux du tableau 8 car on obtient une baisse de 8%.

Les diamètres présentent un maximum de 10 pour un minimum de 5, avec une moyenne de 6.88 et une médiane de 7 associées toutes deux à un écart type de 0.85.

Ces mesures sont toutes cohérentes avec une caractéristique petit monde pour chaque graphe restreint concerné.

Avec un logiciel d'exploration interactive de grandes masses de données comme Tulip [60], il est possible d'obtenir une représentation visuelle de ces graphes triennaux. Les figures 10 et 11 proposent ainsi une sortie graphique pour les TR_01 (composé à partir des textes des années 1971/72, 1972/73 et 1973/74) et TR23 (composé à partir des textes des années 1994, 1995 et 1996). Ces exemples de graphes triennaux restreints ont été choisis car ils correspondent aux périodes retenues pour les extraits de textes proposés dans les annexes.

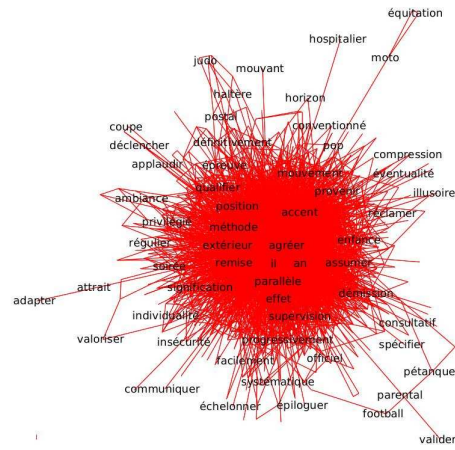


FIG. 10 – Visualisation du graphe triennal restreint TR_01

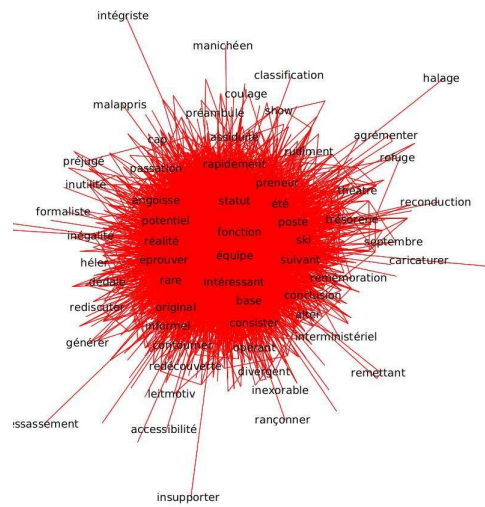


FIG. 11 – Visualisation du graphe triennal restreint TR_23

La physionomie générale d'un tel graphe ressemble à une grosse pelote où le détail des liens entre nœuds ne peut pas être distingué. Le logiciel Tulip propose de mettre en avant certains labels de nœuds et affiche ainsi certains mots choisis en fonction de leur répartition dans l'affichage visuel. Toutefois, en fonction de l'interaction opérée dans l'exploration du graphe (zoom, rotation, mode d'agencement, déplacement du graphe), ce sont des mots différents qui sont susceptibles d'être affichés. Le nombre de pixels nécessaires à l'affichage des intitulés et des liens entre nœuds est trop important pour la résolution de l'écran et la taille de l'image.

Il est possible d'agrandir et de zoomer sur une zone plus précise. Apparaissent alors les termes voisins d'un terme choisi. Sur la figure 12, le graphe TR_01 a été scruté dans le voisinage du mot « équipe ». On y lit des termes voisins comme « utilisation », « fonctionnement » ou encore « activité ». Pour rappel, il s'agit du graphe triennal restreint, dans lequel n'ont été conservés du texte initial que les termes apparaissant dans une occurrence significativement plus fréquente que dans le français « ordinaire ».

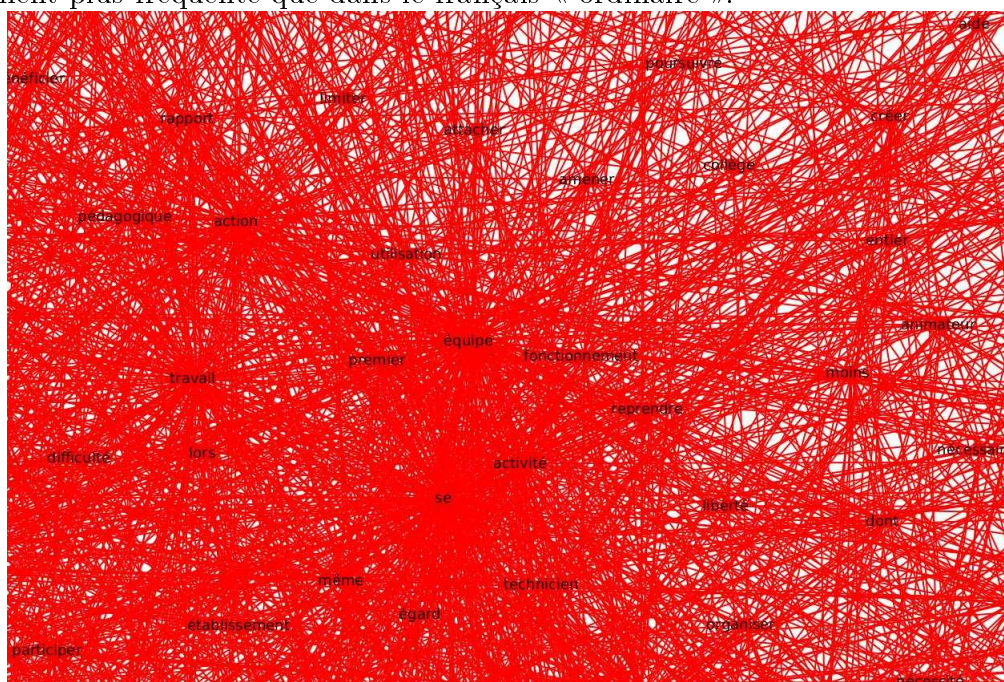


FIG. 12 – Détail du graphe TR_01, centré sur le mot « équipe »

5.3 La dynamique des graphes triennaux

Sur la période d'observation, la production écrite des bilans d'activité perçue à travers notre prisme de composition des ensembles triennaux présente 10893 sommets différents (lemmes). En appliquant à chaque lemme le test exact de Fisher au sein de chaque triennal où il apparaît, le nombre total de sommets a été ramené à 5867 lemmes. Un lemme présent dans deux ensembles triennaux peut très bien n'être retenu par le test de Fisher dans aucun de ces deux ensembles triennaux, ou bien n'être retenu que dans l'un des deux triennaux restreints, ou encore satisfaire au test de Fisher pour ces deux ensembles triennaux. Les 10893 lemmes de notre ensemble d'observation ont ainsi autant de "chances" de figurer parmi les 5867 lemmes retenus qu'ils figurent dans un ensemble triennal de la période. Chacun des 5867 lemmes retenus peut ainsi être affecté d'un score correspondant au nombre de triennaux restreints dans lesquels il a été conservé.

5.3.1 Les constantes

Au final, 146 lemmes ressortent particulièrement pour avoir été non seulement toujours présents sur les 36 regroupements triennaux mais également statistiquement sélectionnés à chaque fois. Ces 146 lemmes relèvent de différentes catégories de mots que l'on peut répartir au sens large en trois groupes.

La majeure partie est composée de 102 mots que l'on peut classer comme des lemmes substantifs ou adjectifs. Ces mots sont présentés dans le tableau 12, dans l'ordre alphabétique (de haut en bas et de gauche à droite).

Le second groupe est constitué de 20 verbes. Ce sont des verbes d'articulation pour l'ensemble du vocabulaire global. Ils sont présentés dans le tableau 13, dans l'ordre alphabétique (de haut en bas et de gauche à droite).

Vingt-quatre autres lemmes composent le troisième groupe. Ce sont des lemmes de type adverbial ou propres à l'usage de la langue française dans le cadre d'une extraction réalisée à partir de phrases lemmatisées. Non spécifiques à la profession, ils témoignent peut-être d'une manière d'expression propre aux documents observés et constituent finalement un bruit perma-

ment dans l'observation. Le tableau 14 présente ces 24 lemmes dans l'ordre alphabétique (de haut en bas et de gauche à droite).

absence	cit�	effet	intervention	p�dagogique	r�alisation
accueil	collaboration	�l�ment	jeune	permanent	recherche
action	collectif	emploi	lieu	personnel	r�flexion
activit�	commune	�quipe	local	population	relation
adolescent	communication	�volution	loisir	positif	rencontre
adulte	compte	exp�rience	mesure	possibilit�	r�union
aide	condition	fa�on	mise	premier	r�le
an	connaissance	familial	moyen	pr�paration	secteur
animation	conseil	famille	n�cessaire	pr�sence	situation
ann�e	cours	fonction	niveau	pr�vention	social
apprentissage	cr�ation	formation	nombre	prise	sortie
association	demande	g�n�ral	nouveau	professionnel	sp�cialis�
autre	d�marche	groupe	objectif	projet	stage
bilan	diff�rent	habitant	part	propre	structure
cadre	difficult�	important	participation	quartier	tel
camp	�ducateur	individuel	particulier	quotidien	travail
charge	�ducatif	int�r�t	partie	rapport	vacance

TAB. 12 – Substantifs et adjectifs permanents significatifs

aborder	d�finir	organiser	situer
agir	engager	participer	souhaiter
amener	�tablir	permettre	
concerner	exprimer	proposer	
constituer	faciliter	rencontrer	
cr�er	intervenir	sembler	

TAB. 13 – Verbes permanents significatifs

ainsi	certain	etc	nous
auprès	chacun	eux	particulièrement
auquel	ci	lequel	plusieurs
ce	divers	lors	régulièrement
celui	dont	moins	se
cependant	également	nombreux	souvent

TAB. 14 – Bruit permanent significatif

Un regard posé sur la dynamique des graphes concerne ces 146 lemmes que l'on retrouve de façon permanente sur les 36 triennaux restreints. Nous avons observé la présence d'éventuelles arêtes au sein de chacune des paires de ces lemmes tout au long de la période étudiée, soit sur presque quarante années.

Les 146 lemmes considérés pourraient être à la base d'un nombre maximal de $(146*146)/2$ arêtes différentes (soit 10 658), y compris les arêtes réflexives d'un mot ou lemme sur lui-même. En considérant l'arité moyenne, mesurée à 8,63 pour les noeuds au sein du corpus étudié, on pourrait même s'attendre à une valeur estimée de 1260 arêtes pour ces 146 sommets. En fait, les arêtes toujours présentes entre ces 146 mots sur toute la période des 36 triennaux ne sont qu'au nombre de 48. Ces arêtes sont des éléments de stabilité de la dynamique du lexique professionnel. Parmi elles, 14 arêtes méritent d'être mentionnées et sont présentées dans le tableau 15. Ce sont des liaisons établies entre les termes professionnels et forment d'une certaine façon des bigrammes du langage professionnel (jeune adulte, groupe de jeunes, travail d'équipe, action éducative, prise en charge, etc.). Les autres arêtes relevées et non présentes dans le tableau 15 mettent en oeuvre au moins un lemme du tableau 14 (comme ce-travail, nous-sembler, nous-important, ...) voire pour 6 de ces associations, deux lemmes de ce tableau (comme celui-ci, se-ce, ce-nous, ce-ce, nous-nous, et nous-lequel).

travail-prévention	prise-charge
jeune-adulte	travail-équipe
groupe-jeune	activité-loisir
social-familial	prévention-spécialisé
emploi-formation	éducateur-prévention
action-éducatif	jeune-an
équipe-éducatif	autre-part

TAB. 15 – Arêtes permanentes significatives

La dynamique des graphes triennaux restreints se place également autour de l'évolution du vocabulaire. Nous rapportons ici quelques résultats concernant cette évolution.

Le vocabulaire de chaque ensemble triennal restreint s'appuie pour partie sur des lemmes qui lui sont propres et pour partie sur des termes partagés avec les autres ensembles. De façon stricte, aucun ensemble triennal ne comporte de termes qui ne sont pas partagés avec les autres. En effet, deux triennaux successifs ont par construction un recouvrement de deux-tiers. Concernant les triennaux restreints le recouvrement est du même ordre. Il est donc plus judicieux de se focaliser sur le vocabulaire commun entre ces ensembles, et de regarder les recouvrements qui existent entre ces ensembles.

Comme vu plus haut, le recouvrement exact entre tous les ensembles ne porte que sur 146 lemmes. Nous allons donc considérer certains recouvrements partiels, et pour apprécier cette évolution du vocabulaire au cours du temps, nous choisissons de prendre comme point fixe le dernier ensemble (TR36). A partir de là nous mesurons le taux de partage de chaque ensemble TR avec le TR36. C'est ce que représente la figure 13, en superposant deux courbes sur une échelle commune : l'évolution au cours du temps du nombre de lemmes des ensembles TR (déjà présente sur la figure 7), et l'évolution au cours du temps du nombre de lemmes partagés avec TR36.

La partie droite de ces courbes (à partir de TR34) et leur jonction finale (TR36 partage tout son vocabulaire avec TR36), s'explique par le mode de

construction des ensembles triennaux avec un recouvrement glissant.

Sur le reste de la période temporelle, la figure 13 présente une tendance à l'harmonisation avec le vocabulaire de TR36. Le taux de recouvrement de chaque TR avec TR36 reste de l'ordre de 50%, mais avec une tendance haussière. Le lexique professionnel évolue au cours du temps et les TR récents partagent plus de lemmes avec TR36 que les TR plus anciens.

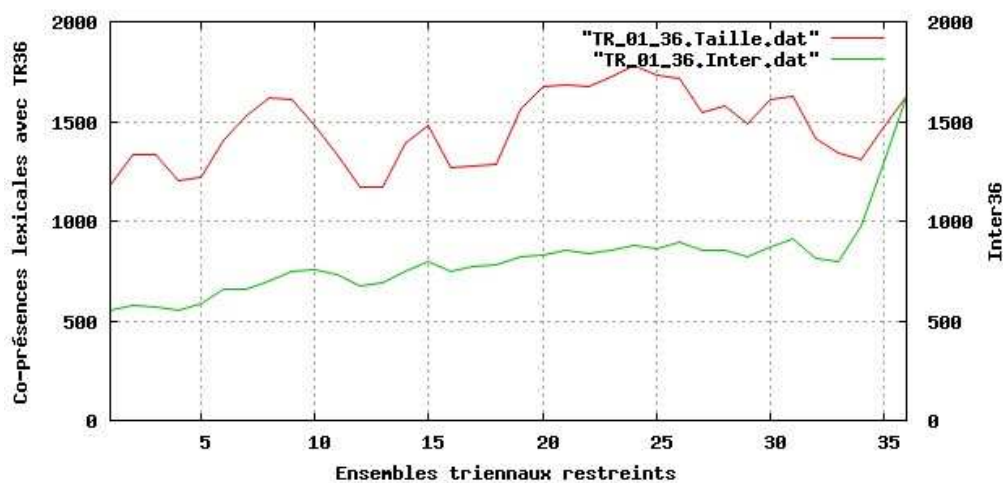


FIG. 13 – Partage lexical entre chaque triennal restreint et le 36ème triennal

Une autre façon de considérer la situation est de répondre à la question "Quand est-ce que les lemmes composant TR36 ont-ils fait leur apparition?". C'est ce que propose de faire la figure 14. Les deux extrémités de la courbe proviennent des artefacts de construction. D'une part le mode de construction de TR36 amène ici aussi une "bosse" à partir de TR34; d'autre part l'apport

des 1175 lemmes de TR01 parmi les 5867 lemmes au total produit un pic en début de courbe. L'aire sous la courbe représente les 1622 lemmes de TR36. Cette figure représente bien deux choses : l'apport régulier de chaque TR au vocabulaire global (la courbe décroît mais ne s'annule jamais), et la part de renouvellement de chaque TR (symbolisé ici par la "bosse" artificielle finale). Ce dernier point pourrait témoigner de problématiques professionnelles nouvelles au sein d'un métier dont les bases restent stables.

La figure 15 reprend les courbes temporelles précédentes mais les présente sous la forme de rapports. La courbe rouge correspond au pourcentage de lemmes partagés entre chaque TR et TR36. On voit que cette valeur oscille autour de 50%, avec une tendance haussière. La courbe verte correspond au pourcentage de lemmes de chaque TR qui est nouveau par rapport au vocabulaire déjà rencontré. Il est normal que cette seconde courbe décroisse, il faut noter qu'elle ne s'annule jamais.

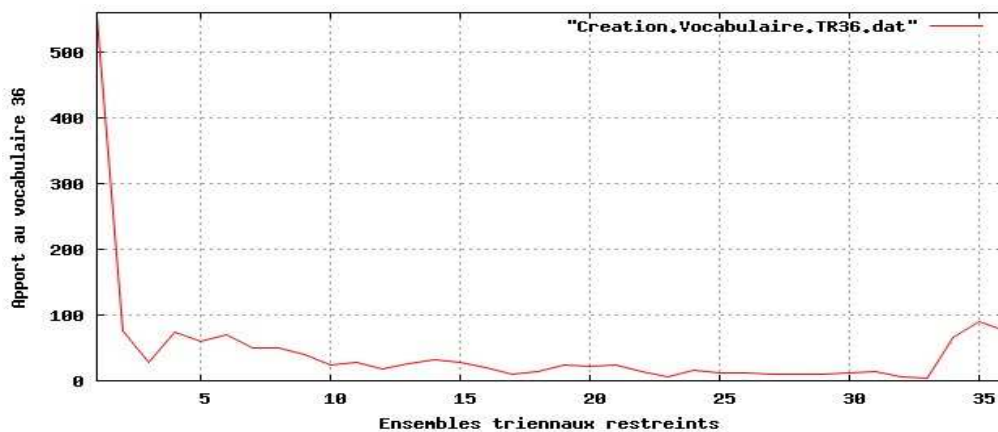


FIG. 14 – Moments d'apparition des mots du TR36

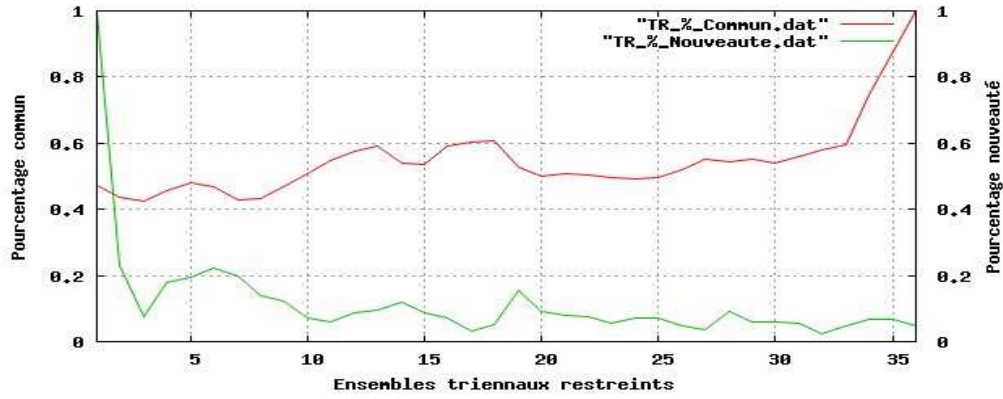


FIG. 15 – Distribution du vocabulaire en pourcentage

La figure 16 présente le moment d'apparition de chacun des 5867 lemmes composant la globalité des ensembles TR. Malgré l'étendue de la période d'observation (environ 40 ans), on ne parvient pas à épuiser le vocabulaire "significatif", c'est-à-dire employé dans une proportion significativement différente de l'usage du français courant. Cette courbe pourrait être sous-titrée : "au niveau professionnel, il y a toujours quelque chose de nouveau dont on doit parler"...

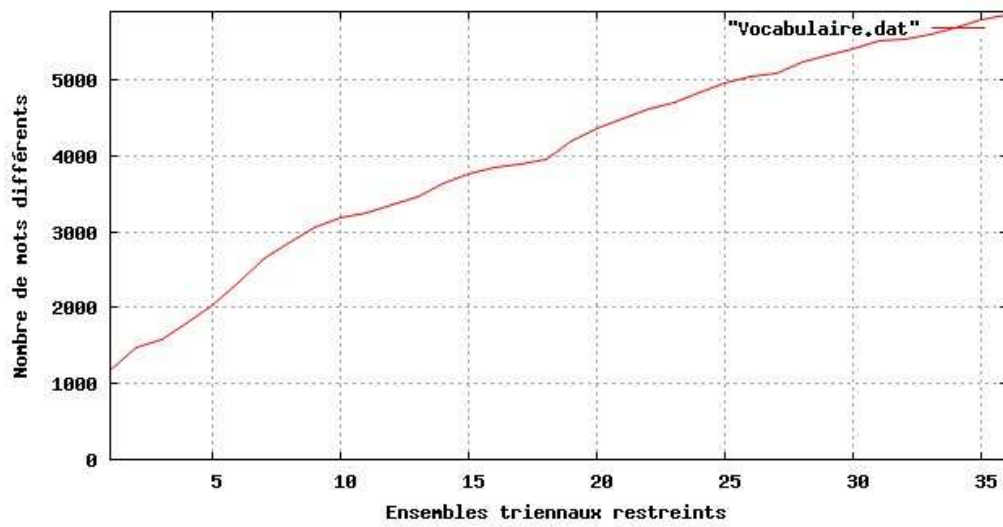


FIG. 16 – Nombre de mots différents par triennal restreint

5.4 La frise chronologique de la période

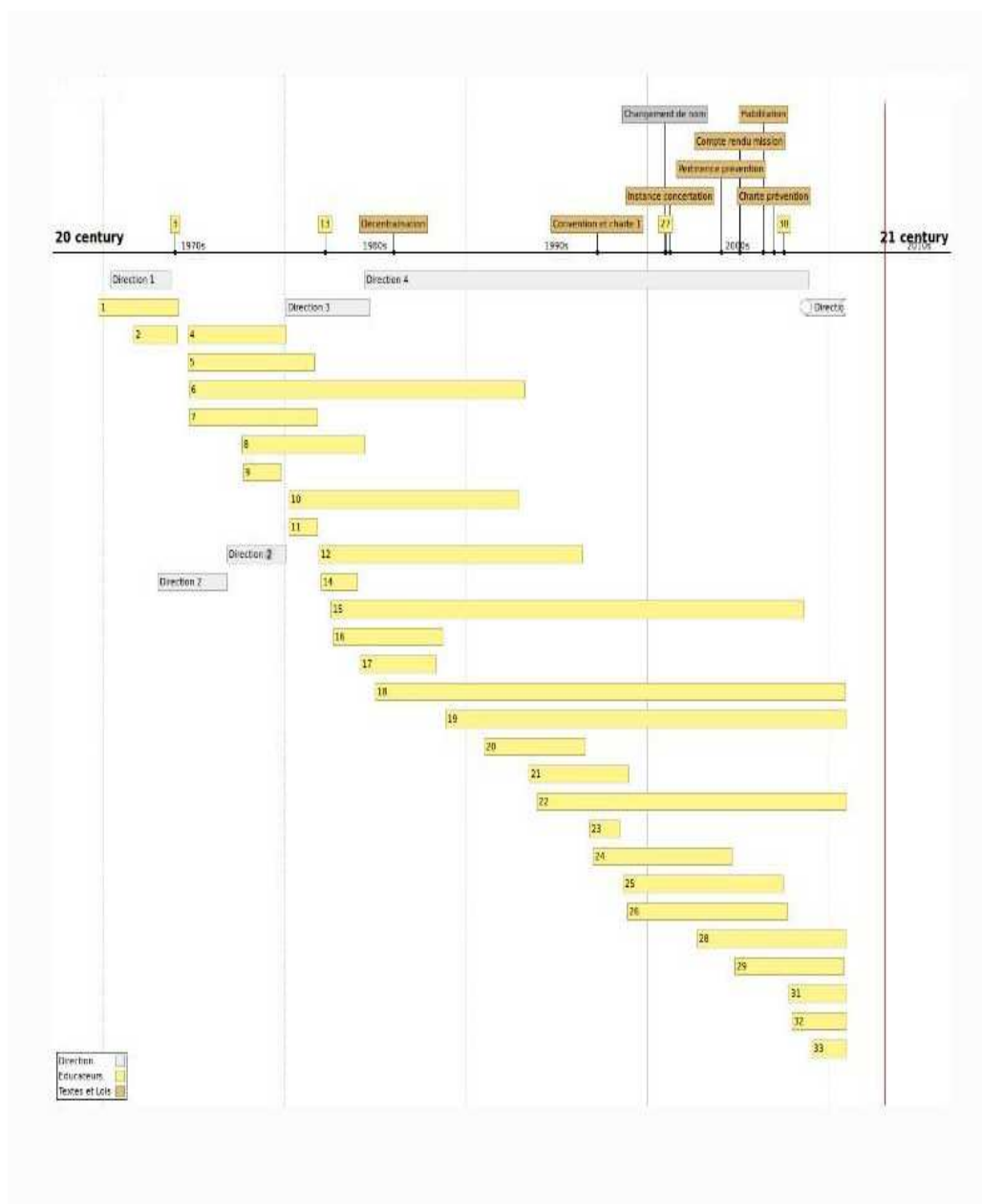


FIG. 17 – Frise chronologique

Entre 1970 et 2010, trente-trois nouveaux éducateurs de rue ont été embauchés.

Ils ont tous exercé leur compétence professionnelle et ont donc contribué à l'écriture du bilan annuel d'activité. Certains ont été présents de longues années alors que d'autres n'ont exercé que sur des périodes plus courtes. Parmi les intervenants de longue durée, sept d'entre eux ont travaillé au moins douze ans d'affilée. Dans le groupe de ces sept personnes de longue présence, il est possible de considérer une partition temporelle avec deux périodes successives .

La première époque débute en 1974 et fini en 1996 avec essentiellement trois personnes. Les durées concernées sont de 12 à 18 ans et ont abouti à deux exercices de fonction de directions en prévention spécialisée en Gironde. La seconde période présente des moments partagés avec la première, globalement elle débute alors que l'autre n'est pas finie, de 1985 jusqu'à 2010 et est composée de quatre personnes personnes. L'une de ces quatre personnes embauchée en 1982 a investi le poste de direction à Bègles en 1988.

Outre l'arrêté de 1972 qui fonde légalement et globalement un mode d'intervention dénommé « prévention spécialisée », le premier texte légal à intervenir en prévention concerne la décentralisation et l'adaptation de la législation sanitaire et sociale au niveau du département sous la responsabilité du président du conseil général. Le rapprochement avec le conseil général a été d'abord validé par la signature de la première charte départementale et d'une convention en 1997. Peu après, le service de prévention spécialisée de Bègles a été renommé sous le nom d'association de prévention spécialisée de Bègles. Une étude départementale sur la pertinence de la prévention spécialisée a été réalisée en 2004 par un organisme choisi par le département, puis a été suivie par la mise en place officielle du compte rendu de mission en 2005. L'habilitation des associations déléguées à l'exercice de la compétence de prévention spécialisée a été suivie un an plus tard d'une seconde charte, respectivement en 2006 et 2007.

La période de légifération la plus fournie a démarré en 2002, via la loi du 2 janvier 2002 qui rénove les actions sociales et médico-sociales, puis a été suivie par la loi de 2007 sur la protection de l'enfance, que d'aucuns ont

qualifié de loi de la prévention de la délinquance.

Si l'on remonte à la figure 9, il est possible de corréler les variations à la baisse du corpus par le triennal 15 (composé des trois années qui suivent le rapport 1986-87) avec la mise en place et l'installation de la décentralisation, ainsi que la stabilisation de phénomènes induits. De même, toujours sur cette figure, le creux des triennaux 22 à 27 peut être mis en lien avec la variation des embauches entre 1993 et le tout début des années 2000. Pour finir et toujours avec la figure 9, il paraît intéressant de rapprocher l'augmentation de la production de nouvelles formes (lemmes) spécifiques avec l'entrée de six nouveaux éducateurs dès le triennal 33 (à partir de 2007) automatiquement impliqués dans la rédaction des bilans d'activité.

5.5 Les enquêtes auprès des professionnels

La question de la première enquête a été présentée à huit personnes. Ces huit professionnels présents ont fourni en réponse à cette question, soit des syntagmes, soit des mots. En fonction de chaque professionnel, les réponses vont de l'unité extrêmement simple (lemme, mot unique) à des ensembles syntagmatiques pouvant atteindre jusqu'à neuf éléments différents. Les réponses fournies sont consultables en annexe 11. Cette diversité et cette variété de réponses ont été synthétisées dans un ensemble de dix éléments, tous fortement liés à l'exercice professionnel. Cet ensemble synthétisé a été présenté ensuite dans la deuxième enquête pour solliciter les éducateurs sur les thématiques suivantes :

actions collectives, activités, chantiers, errance, familles, filles, partenariat. précarité, scolarité, travail de rue.

Cette seconde enquête a été présentée en 2011, soit à plusieurs mois de distance de la première. Elle n'a pu être proposée qu'à six personnes en raison des absences afférentes à l'évolution de la taille de l'équipe de l'APSB. Les six éducateurs présents sur cette période ont chacun fourni des mots d'extraction nominale, verbale ou adjectivale. Cependant, quelques syntagmes sont présents dans les listes réceptionnées. Les réponses fournies à cette seconde enquête sont consultables en annexe 12.

Les résultats de la première enquête ont donc permis d'élaborer la seconde et ouvrir ainsi un questionnement à postériori sur la présence d'un fond lexical partagé par les professionnels pour faire le bilan écrit de leur travail. Avec les réponses de la seconde enquête, il est possible de comparer les mots choisis par association à une thématique avec les mots appartenants au langage professionnel tel que nous l'avons défini.

Les mots d'association thématique, consultables en annexe 12, fournis par les éducateurs ont tous été lemmatisés parfois au prix d'une division syntagmatique. Ainsi, individualités est devenu par lemmatisation « individualité », « projet commun » a été comptabilisé comme « projet » et comme « commun », « coordination de proximité » a été comptabilisé comme « coordination » et comme « proximité » avec suppression de « de ». En d'autres termes, cette méthodologie d'analyse de l'enquête a traité les formes par lemmatisation individuelle. Seuls les noms propres suivants n'ont pas pu l'être (Hostens, MDSI, Remue_Méninges, SAEMCIB). Tous les mots proposés par thématique n'ont été pris en compte que si ils appartenaient au « langage professionnel » tel que nous l'avons défini.

Thématique	Diff Liste	Présence Texte	Fisher
Travail de rue	47	44	39
Actions collectives	44	43	41
Activités	39	39	36
Chantiers	41	41	39
Errance	43	42	37
Familles	43	43	39
Filles	45	42	37
Partenariat	48	46	44
Précarité	37	34	30
Scolarité	35	34	34

TAB. 16 – Connexions aux premiers voisins de l’aréopage des réponses par thème

Dans le tableau 16, la première colonne liste les thématiques interrogées. La seconde, le nombre de lemmes différents dans l’aréopage des réponses. La troisième s’assure de la présence de ces lemmes dans le texte global lemmatisé. La quatrième mesure l’appartenance au lexique professionnel.

Les variations entre les colonnes deux, trois et quatre permettent d’anticiper des interrogations autour de leurs valeurs. La première et la dernière ligne s’opposent au niveau de leurs variations entre ces trois colonnes. Nous avons voulu construire une représentation de la tendance à faire discours selon le thème considéré, ici le travail de rue et la scolarité.

Les courbes des figures 18 et 19 ont été construites à partir des graphes triennaux (représentation matricielle) et fournissent pour un « mot professionnel », selon le triennal considéré, sa tendance à être connecté à ses « premiers voisins ». Elles représentent en pourcentage, par triennal, en fonction de la thématique donnée, le rapport entre la somme brute des connexions aux premiers voisins pour chacun des lemmes de réponse d’enquête et la somme brute des connexions aux premiers voisins pour tous les mots du langage professionnel.

C'est dans la période étudiée la tendance à faire discours sur un sujet donné. Cette tendance est mesurée par les connexions de premier voisinage réalisées autour des mots des experts dans une thématique donnée. Il ne faut pas oublier que les lemmes d'associations des experts ont été fournis début 2011. Les courbes considérées sont en quelque sorte une rétroprojection dans le temps d'une appréciation associative récente.

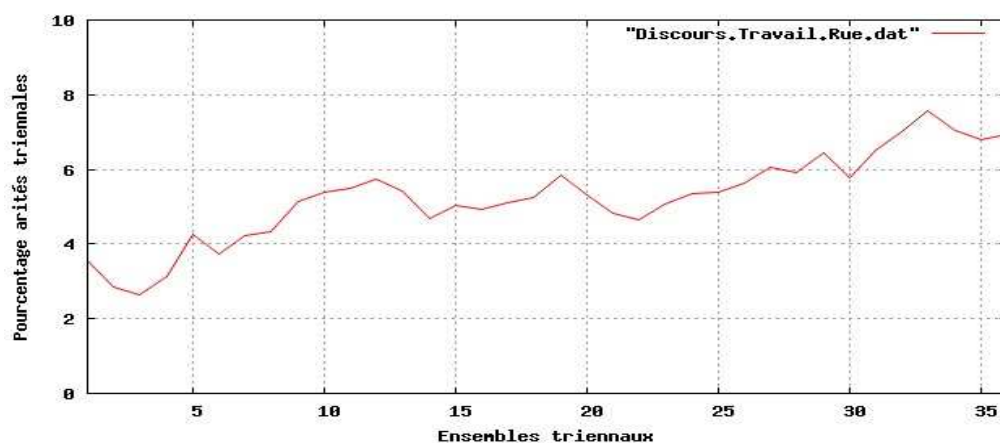


FIG. 18 – Relation en pourcentage entre le temps et les arités de lemmes de la thématique travail de rue

Il est possible de commenter la figure 18 avec l'idée qu'il y a toujours eu beaucoup à dire, pour faire une description des actes de communication réalisés dans un espace urbain ou social. Dire ce qu'est le travail de rue est une synthèse difficile à réaliser, sa condensation lexicale, voire sémantique, tend à échapper à des modalités figées. Il y a de si nombreuses manières de le dire que la mise en application déborde, au moins en termes lexicaux. Ce qui vient très probablement pointer ici la dimension pragmatique de cet exercice, c'est à dire la relation des utilisateurs avec leurs signes [31] et sa richesse.



FIG. 19 – Relation en pourcentage entre le temps et les arités de lemmes de la thématique scolarité

La figure 19 montre dans la période étudiée, sauf l’exception du triennal 10 au triennal 15, une stabilisation avérée des associations verbales récentes des éducateurs autour des questions de l’école ou de la scolarité. Un pic se caractérise par l’augmentation brutale, en termes de premiers voisins, des lemmes fournis par les éducateurs autour de cette thématique. Bien que ceci ne soit pas mentionné dans les éléments de la chronologie, la structure étudiée a participé à la mise en place d’une structure de dimension locale, en l’occurrence une association de soutien et d’aide pour des jeunes du territoire souffrant de difficultés scolaires. Il est très probable que ce pic appartienne à cette période puisque le triennal 10 relève d’années centrées à minima autour de l’année 1983. Le compte rendu d’activité de 1985 mentionne l’existence et le nom de l’association concernée (Remue_Méninges). Nous avons tendance à supposer que de telles actions restent perceptibles au regard de la durée de la période étudiée.

5.6 Investigations complémentaires

L'étude du langage professionnel a montré les constantes du vocabulaire professionnel. Ces constantes peuvent être perçues comme des éléments d'une langue de bois. Cette expression est une métaphore pour parler d'un langage stéréotypé, dénotant une certaine rigidité [11]. Dans le cadre d'investigations complémentaires, nous nous sommes donc davantage penchés sur la dynamique réelle de cette langue d'expression de compte rendu, dynamique qui puisse être perceptible à travers l'usage de notre méthodologie.

La figure 20 est un histogramme qui vient illustrer à partir des graphes la répartition relative et dans le temps des deux termes « étranger » et « religion ». Ces deux termes appartiennent au lexique professionnel et n'apparaissent qu'à des moments tout à fait limités dans la période étudiée. Le temps qui sépare ces deux mots marque la réelle différence des temps de leurs usages dans les comptes rendus.

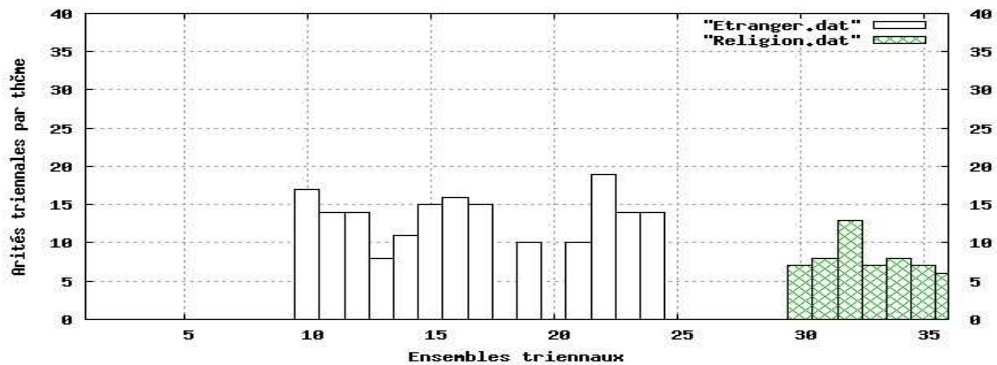


FIG. 20 – Distribution temporelle du degré de connexion au voisinage des termes affichés

Dans la mission des éducateurs, la relation s'exerce dans un espace urbain confronté à des besoins sociaux émergents, ce qui facilite une intimité entre la prescription du travail et un certain vacarme des repères sociaux [65, p 42]. L'histogramme de la figure 21 montre la distribution dans le temps de la présence des lemmes « usager » et « client ». Différencier l'usage de ces termes est une problématique propre à des travailleurs sociaux qui se servent de la communication orale et lui reconnaissent un réel pouvoir performatif [65].

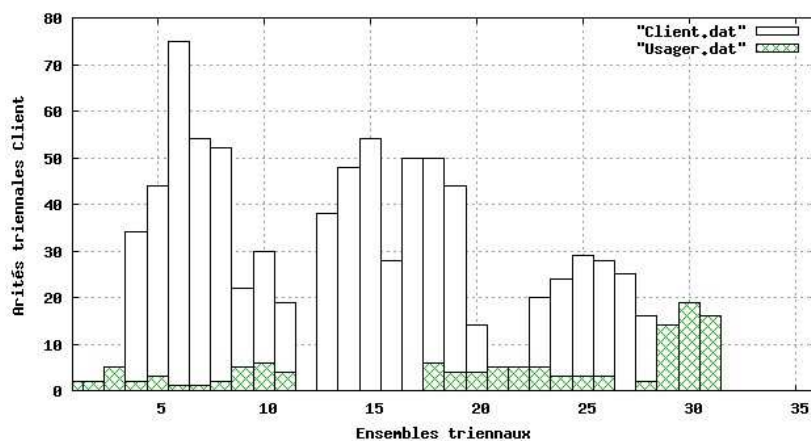


FIG. 21 – Histogramme des lemmes « Client » / « Usager » dans la période étudiée

5.7 Un graphe global construit par concaténation d'arêtes issues de triennaux restreints

Un graphe complémentaire a été produit à partir de la concaténation de toutes les arêtes qui ont pu être sélectionnées à un moment ou à un autre dans un des triennaux restreints.

A l'instar des graphes restreints, ce graphe présente des caractéristiques petit monde bien qu'il ne soit constitué que d'arêtes composées de sommets sélectionnés au sein d'un triennal dont on ignore la période d'origine.

Les caractéristiques générales de ce graphe sont présentées dans le tableau 17 et les caractéristiques petit monde dans le tableau 18.

S	A	SNC	SC	AC	SCF	SCNF
5867	118 831	12	5855	118 825	248	5607

TAB. 17 – Graphe global concaténé

Les en-têtes de colonnes des tableaux 17 et 18 sont les suivantes :

S : Sommets

A : Arêtes

SNC : Sommets non connexes

SC : Sommets connexes

AC : Arêtes connexes

SCF : Sommets connexes feuilles

SCNF : Sommets non connexes feuilles

DC : Distance caractéristique (moyenne)

DC	Diamètre	Coefficient d'agrégation
2.61	7	0.53

TAB. 18 – Métriques du graphe global concaténé

La figure 22 donne une représentation visuelle de ce graphe global composé à partir de la concaténation des graphes triennaux restreints exposés précédemment. Comme pour les figure 10 et 11, la physionomie générale de ce graphe est celle d'une grosse pelote, ici composée de près de 6000 nœuds inter-reliés à l'aide de plus de 118 000 arêtes.

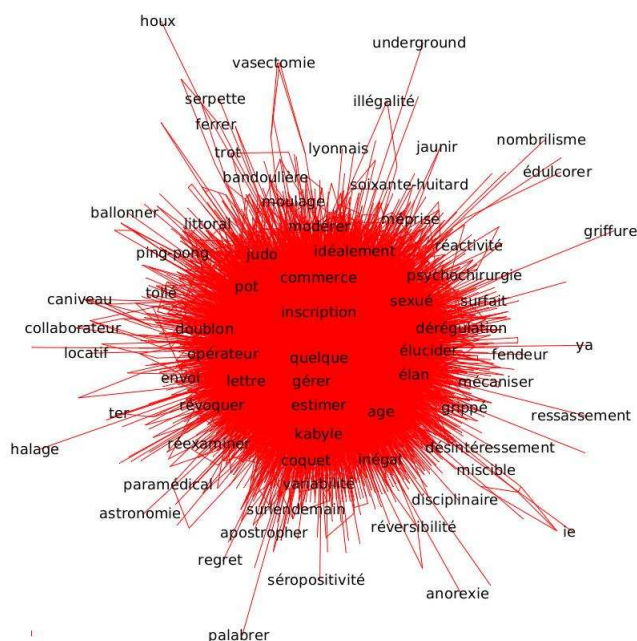


FIG. 22 – Représentation globale du graphe des TR concaténés

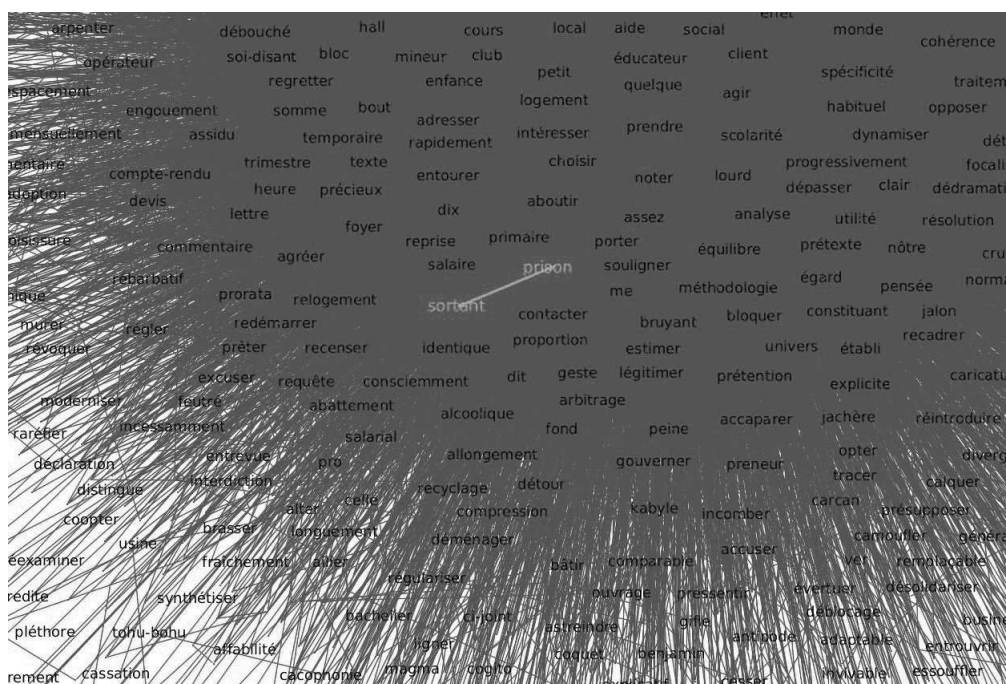


FIG. 23 – Détail du voisinage du mot « prison » dans un extrait du graphe des TR concaténés

La figure 23 cible ce même graphe sur le mot « prison ». La comparaison de ces deux figures (22 et 23) montre bien qu'à ce niveau de globalité, ce sont les algorithmes d'affichage du logiciel Tulip qui déterminent quels sont les termes affichés. Un simple ciblage sur un terme donné, modifie l'ensemble des labels affichés. Pour pouvoir distinguer et suivre les liens entre nœuds, il faut procéder à un niveau de zoom plus élevé (comme précédemment sur la figure 12) ou bien solliciter de façon spécifique les cooccurrences du mot « prison » dans le graphe des cooccurrences durables comme dans la figure 24.

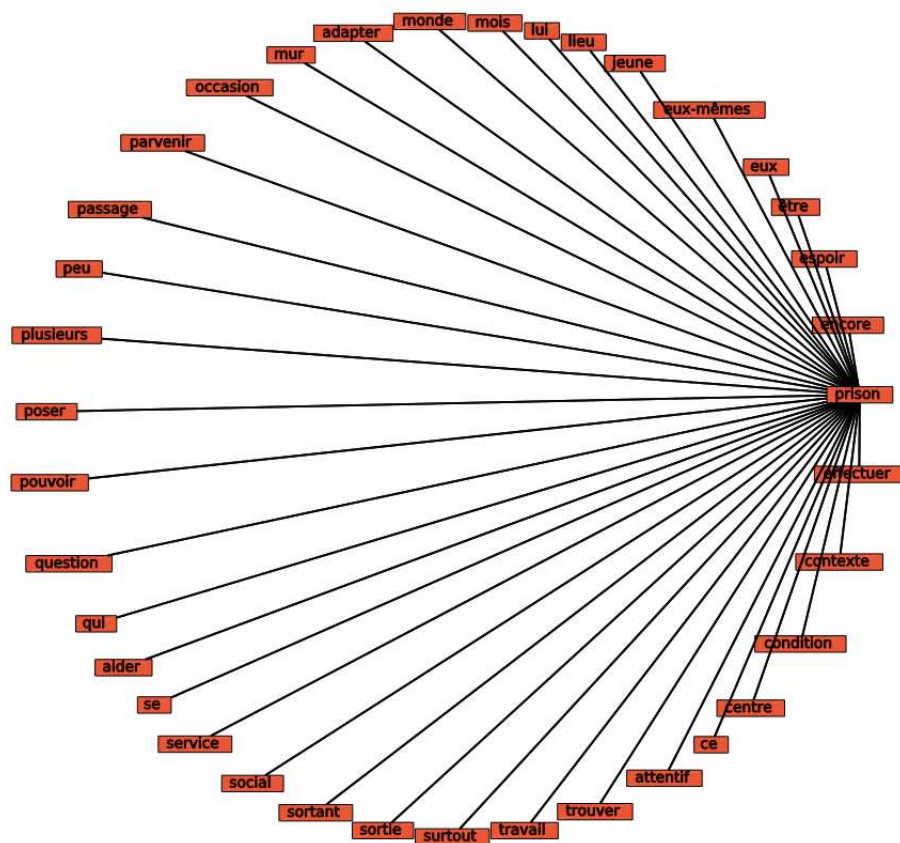


FIG. 24 – Détail du voisinage du mot « prison » dans le graphe des TR concaténés

La figure 24 présente un détail précis du voisinage du mot « prison », détail composé de 36 cooccurents labelisés. Ces cooccurences appartiennent toutes au graphe des TR concaténés. Cependant, elles peuvent être situées dans le triennal restreint dans lequel elles sont apparues la première fois. En l’occurrence, « prison » ne manifeste sa présence significative que dans quatre triennaux spécifiques : TR_22, TR_23, TR_24, TR_25. Ce qui renvoie son caractère significatif aux années comprises entre 1993 et 1998 quand ces années sont regroupées par trois pour valider une telle sur-représentation fréquentielle.

6 Discussion

6.1 Rappel synthétique de la démarche

Tout au long de notre recherche, nous avons voulu approcher une perspective d'évolution pour des graphes de cooccurrences lexicales, cooccurrences extraites de comptes rendus écrits. Les documents utilisés pour atteindre cet objectif se rapportent à des usages professionnels et contextualisés dans cet usage de la langue. Nous avons choisi de capter des sens ou des idées manifestés dans ces archives, dans ces rapports d'activités de travailleurs sociaux, via les mots en usage dans les textes concernés.

Les mots utilisés traduisent des habitudes d'expression ou de vocabulaire qui sont, ou bien ont été, présentes au cours de ces années pour ces éducateurs de prévention spécialisée. L'ensemble de la recherche repose sur un matériel écrit annuel composé de trente-huit bilans, dits d'activités. Ces textes de bilans sont étalés sur presque quarante ans et ils composent un total de plus de 1800 pages imprimées.

Pour produire l'écriture de toutes ces pages, 10 893 entrées lexicales, lemmatisées, différentes, uniques ont été employées.

Toutes les variations des fréquences lexicales annuelles, qu'elles soient excessives ou fortes, ont été en partie lissées par un processus de triennalisation. L'ensemble de toutes les formes graphiques fléchies a été lemmatisé, puis trié significativement par le biais de sa sur-représentation relative de fréquences pour approcher ce coeur de langage « professionnel » dans ce type de communication formelle.

Ainsi, les 10893 entrées lexicales ont été ramenées par tri à 5867 termes lemmatisés et concentrés dans 36 triennaux restreints construits par cette sélection de fréquence. Chaque triennal restreint se retrouve ainsi composé d'arêtes dont les sommets ont prouvés qu'au moins un fois leur usage se faisait dans une fréquence significative.

Les graphes triennaux, comme les graphes annuels possèdent des propriétés de réseaux petit monde (pour rappel faible diamètre et fort taux d'agrégation). Ces caractéristiques de type petit-monde propres à nos ré-

seaux construits selon notre méthodologie s'avèrent régulières tout au long de ce temps, pluri-décennal.

Ces graphes de cooccurrences lexicales ont été puisés dans des écrits d'une même profession sociale, avec pour enjeu d'être destinés à décrire un bilan annuel des actions réalisées l'année précédente dans l'exercice professionnel.

Pour l'essentiel, la mission de la profession sociale concernée n'a fondamentalement pas changé au cours de la période d'étude. La motivation à l'écriture des textes, pas plus que la nature de l'expression de son objet, n'ont subi des transformations notables. La formalisation du cadre d'astreinte étant prospectivement encore stable, on peut se demander si une dynamique réelle existe dans le choix des mots pour exprimer l'acte professionnel.

6.2 Le langage professionnel

Le discours professionnel a été de façon statistique ramené à 5867 entrées significatives sur l'ensemble de la période étudiée. Cependant, cet ensemble de mots est distribué différemment suivant la période triennale considérée. Chaque triennal restreint n'est composé en moyenne que de 1467 sommets, ce qui est finalement peu. Cette restriction triennale conduit malgré tout, et à chaque fois, à un ensemble de lemmes représentant environ un quart du vocabulaire considéré comme significatif.

Si l'on regarde la totalité des triennaux restreints et que l'on cherche à savoir comment les auteurs des bilans d'activité ont choisi les paramètres de leurs expressions, on voit bien qu'ils exercent un choix dans un vocabulaire assez large. Un vocabulaire toujours accessible par la langue mais qui a progressivement stabilisé son extension sur l'ensemble de la période que nous avons étudiée. Ce vocabulaire se confronte à celui des comptes rendus de missions institués annuellement par le conseil général de la Gironde dès 2005. Il montre dans cette confrontation une certaine forme de résistance ou de poésie face à un comptage comptable de type analytique. Modalité de comptage qui peut se comprendre pour une institution qui gère l'intervention et le financement de plusieurs équipes de prévention spécialisée. Hors les tableaux qui le composent, la part narrative du compte rendu de mission de

l'année 2010 est fournie en annexe.

Pour une perception de la dynamique lexicale globale de la période étudiée, en considérant uniquement le moment de début et le moment de fin, c'est à dire en prenant en compte le premier triennal restreint et le trente-sixième, on peut se rendre compte avec la figure 10, du rapport de recouvrement des deux triennaux. C'est-à-dire de l'apport de chacun de ces deux triennaux particuliers pour l'ensemble du lexique professionnel. Ce partage, ce recouvrement de l'ensemble du vocabulaire par les apports spécifiques du premier et du dernier triennal est de l'ordre de 38% sur l'ensemble du vocabulaire.

Sur toute la période étudiée, triennal après triennal, des lemmes de fréquences différentes et significatives sont apparus au cours de chaque période, de façon régulière, valorisant par étapes successives, des lemmes inconnus du triennal précédent. Ces nouveautés régulières du vocabulaire sont ainsi progressivement parvenues à devenir un fond permanent de lexique professionnel. Fond lexical à l'intérieur duquel chaque auteur a pu venir puiser une entrée de son choix au moment de l'écriture. Cette possibilité représente une vitalité de l'expression lexicale, mais elle la limite aussi par un lent processus d'érosion créative. Notre recherche n'ayant pas pu être poursuivie sur un trente-septième triennal, on peut supposer que cette évolution lexicale n'est pas encore complètement terminée. Ceci peut se percevoir à l'aide des figures 12 et 13.

Dans le vocabulaire professionnel décrit, les cinq premiers milliers de mots et les huit cents soixante-sept qui les suivent n'ont finalement qu'une présence discontinue tout au long des différentes étapes de l'étude. Occasionnellement, cette discontinuité se fige, mais on peut considérer qu'elle se stabilise parfaitement à travers les 146 lemmes spécifiques présentés dans les tableaux 12, 13 et 14. En fait, cette stabilisation doit être réduite aux 123 entrées lexicographiques présentées dans les tableaux 12 et 13, soit aux 102 substantifs ou adjectifs ainsi qu'aux 21 verbes. Ces 123 termes ont *toujours* été présents dans chaque triennal restreint. Nous considérons que le tableau 14 relève d'un bruitage propre à notre méthodologie de capture cooccurentielle, et relève d'une analyse réalisée sur un écrit totalement lemmatisé en langue française ainsi qu'à des manières de s'exprimer par écrit. Le coeur du lexique profes-

sionnel, ces 123 termes appartiennent selon notre hypothèse au coeur d'un lexique professionnel parfaitement narratif. Ces lemmes ne sont qu'une toute petite partie des termes utilisés pour l'écriture des documents, en fait à peine 1%. Bien que ce vocabulaire soit réduit, il est et reste parfaitement stable dans son existence. Il est amusant de noter que cette stabilité s'exerce à partir de sa non-normalisation dans l'exercice d'une mission constante sur les dernières décennies. Mieux encore, au moins du point de vue des structures de la stabilité, les 14 arêtes du tableau 15 représentent la partie liée de cette stabilisation des entrées langagières.

Si notre recherche porte sur des écrits, les textes analysés ont été produits par des personnes dont l'oral est le média principal de travail. L'agir en situation de travail est peu séparable de l'échange langagier, car il faut bien s'entendre un peu. La communication dans ce travail s'exerce dans un contexte composé de multiples modalités d'organisation entre oral et écrit. Les éducateurs en prévention spécialisée connaissent une pratique basée sur de l'expression orale, mais encadrée par une astreinte régulière d'écriture. Nous avons pu constater dans le cadre de la recherche que les productions écrites professionnelle des bilans ont régulièrement été amorcées par des productions orales préalables ou périphériques. Les deux modalités de production se chevauchent dans le cadre de professions qui alternent échanges oraux et écrits, et c'est le cas ici. Ce phénomène de tension entre l'oral et l'écrit est malgré tout de nature inégale tant du point de vue de sa codification que de sa validation sociale.

Le langage verbal est définitivement incontournable dans la profession d'éducateur de rue et son exercice est conditionné à un échange en cheminement, c'est-à-dire dans un contexte urbain non spécifique. C'est un type de contexte toujours partiellement informel au regard d'une prédiction. L'échange verbal dans le travail de prévention se matérialise sous la forme de liens établis avec les populations contactées. Les personnes visées par cette mission habitent sur des territoires géographiquement et sociologiquement référencés. L'échange verbal qui est la « *caractéristique humaine la plus spécifique* » [9], et qui s'accompagne de mimiques ou d'indices paraverbaux, peut toujours se trouver sous la menace d'une interruption dépendante

des aléas du contexte. Une réelle forme d'incertitude potentielle fait partie de la poursuite des objectifs professionnels.

Le produit de l'astreinte d'écriture est adressé à l'institution détentrice de la compétence légale. Cette obligation est la même pour toutes les équipes de prévention de la Gironde. C'est la stabilisation temporelle avérée de notre corpus qui valide localement sa représentativité narrative. L'astreinte est ritualisée dans ses formes, car quand vient l'heure de l'écriture des bilans d'activité ou des comptes rendus de mission, la prévalence verbale des échanges entre les professionnels est forte, tout le monde en parle au moins un peu. Les échanges verbaux ont tendance à être orientés dans ce sens. Les écrits produits par chaque territoire d'affectation sont proposés à la lecture de tous en interne pour échange, critique ou appréciation. On peut noter que cette modalité de relecture collective ou partagée n'est pas incompatible avec une certaine vitalité lexicale. L'existence pourtant récente des comptes rendus de mission ne s'y oppose probablement pas non plus, bien que l'on puisse envisager qu'elle pourrait avoir tendance à réduire cette propension à la création lexicale. Il est valide de s'interroger sur les raisons de ce renouvellement régulier de vocabulaire qui peut difficilement s'expliquer par la simple répétition d'une astreinte à décrire les résultats de missions qui restent globalement identiques dans leurs objectifs. Nous avons pu pointer l'existence d'un facteur lexical novateur permanent et stable au moment de l'écriture. Tout au long de la période prise en compte, on peut en effet se rendre compte que la diversité du vocabulaire utilisé a toujours été conséquente. Ce vocabulaire sous-tend en fait un lexique adapté à la description d'une problématisation de la question éducative par des praticiens qui interviennent verbalement auprès d'une jeunesse ou de sa parentèle.

La construction par notre recherche de ce lexique le limite à une dimension narrative excluant forcément des termes techniques tels que sigles ou acronymes, éléments non présents dans le corpus généraliste choisi pour la comparaison. La confrontation au corpus généraliste présente automatiquement un biais, chaque lemme ne pouvant pas forcément être présenté à la mesure s'il n'est pas présent dans le corpus de comparaison. Ce redoutable biais peut être évalué à 21% de l'ensemble des termes, car c'est le pourcen-

tage de lemmes qui n'a pu être apprécié sachant que le vocabulaire global travaillé est composé de termes différents, disparates et de natures diverses. Il inclut pour partie des noms propres, des acronymes, des sigles ou des chiffres. La méthodologie de notre traitement a utilisé des suppressions spécifiques de vocabulaire qui nous privent d'une analyse fine, mais ce gros grain montre une nette robustesse à la fois par la permanence temporelle de ses composants et de ses associations internes. En outre, on peut supposer que l'aboutissement à un tel lexique est dépendant de biais dus à une pratique professionnelle essentiellement verbale, incluant des habitudes d'expressions ou une résistance à l'écriture. Pour finir la description de cet inventaire des biais, il est possible de considérer comme discutable le fait d'envisager un ensemble composé de multiples auteurs comme un auteur unique pour la production d'une expression écrite étudiée. Il en va de même dans le fait de considérer le décalage entre le temps de la rédaction et le temps de l'exécution professionnelle comme négligeable dans ses effets. La société évolue de façon continue pendant ce décalage et influence probablement des représentations chez les auteurs. Les auteurs évoluent dans leurs perceptions, dans leurs représentations des problématiques auxquelles ils sont quotidiennement confrontés. Il est d'ailleurs probable que des formes d'habituations puissent émerger. Les représentations propres, pour chacun des auteurs, des difficultés auxquelles sont mis en présence les jeunes gens visés par le travail évoluent également.

La clôture de l'écriture du bilan par les auteurs est finalement celle de la mise en équilibre de tensions antagonistes et multiples, relatives à la société et aux individus, par la production écrites d'actes de langage qui prennent en compte une « réalité matérielle, sociale et idéale » [24].

6.3 Les graphes

L'extraction des cooccurrences lexicales a été réalisée avec une méthodologie grossière quoique systématique. Elle est grossière parce qu'elle ne prend pas « parfaitement » en compte l'importance de l'ordre des mots dans la phrase. Elle se désengage du pied de la lettre, mais elle est systématique parce qu'elle est appliquée à l'ensemble du matériau textuel. Les graphes

triennaux constitués présentent des résultats avec des caractéristiques « petit monde » propres à leur terrain d'extraction et à leur validité temporelle. Ces caractéristiques de petit-monde sont visibles par la distance caractéristique (ou le diamètre) et le coefficient d'agrégation. La méthodologie est robuste dans la mesure où elle présente ces critères sur chaque ensemble triennal tout au long de la période.

Les graphes construits dans notre recherche ne sont en aucune façon des réseaux de dépendance syntaxique. Ils ne sont pas non plus des réseaux sémantiques, ni des ontologies destinées à représenter des connaissances. Notre recherche a construit un réseau pour modéliser le rapport existant entre une équipe de personnes (des professionnels, des auteurs singuliers) et les mots choisis en moyenne par cette équipe pour dire ce qui a été réalisé au cours de l'année précédente. Il est possible de considérer que cette approche dans notre recherche comme une perspective pragmatique, prise au sens de « mode d'appréhension du langage [31] », des résultats mettant en valeur la force des signes captés.

La taille de la fenêtre de capture des cooccurrences lexicales est déterminante pour les propriétés qui sont obtenues sur les graphes. Nous n'obtiendrions sans doute pas les mêmes propriétés avec une fenêtre plus large. Une augmentation de la taille de la fenêtre prendrait le risque de capturer des termes peu voisins dans la phrase. Nous nous sommes donc limités à une capture de mots très voisins. La cooccurrence des mots ou des lemmes n'est pas de nature triviale. Elle ne dérive pas simplement de la distribution ordinaire des mots par leur fréquence d'occurrence comme dans la loi de Zipf. Si nous avions aléatoirement mélangés des lemmes de la langue française dans les phrases en modifiant les fréquences d'apparition dans les bilans d'activité, nous n'aurions pas obtenu des arêtes qui s'approcheraient réellement d'un coeur de langage de spécialité, de fonction ou profession. Le mérite de la fenêtre de capture réside dans sa capacité à pouvoir saisir en souplesse un peu de ce jeu de voisinage entre les mots, tout en n'étant pas limité à l'ordre strict des mots dans la phrase. Les captures obtenues sont différentes du hasard, elles sont relatives au choix des mots à l'intérieur de l'expression écrite.

6.4 Perspectives

La mise en ligne du corpus narratif des bilans d'activité de l'équipe de prévention spécialisée de Bègles (APSB) est prévue pour être réalisée sous une forme anonymisée dans les semaines à venir. Cette mise à disposition publique vise à atteindre des enrichissements à l'aide d'études complémentaires qui pourraient en sortir, à partir du corpus et pour tout champ disciplinaire intéressé.

Le travail réalisé met en évidence un lexique professionnel propre à des intervenants du travail social de rue. Ce lexique a été obtenu à l'aide d'une procédure quasi-automatique sur l'ensemble du corpus. C'est un vocabulaire d'usage, usage professionnel, dont la constitution présente un sens documentaire pour qui s'intéresse au travail de rue. La méthodologie de cette captation sémantique présente des possibilités d'extrapolation vers d'autres espaces professionnels. Extrapolation qui peut être tentée soit pour des condensations textuelles dans des domaines extérieurs au travail social, soit pour des perspectives de captation de constellations lexicales et professionnelles spécifiques. De même, il est largement envisageable d'utiliser cette procédure pour procéder à des comparatifs entre textes et entre méthodes d'appréhension des textes.

Nous avons fait le choix de recourir exclusivement aux formes canoniques de l'écrit. Cette modalité dans la phrase a fatalement fait souffrir toute forme d'appréciation du contexte des mots. La lemmatisation a été nécessaire pour effectuer des comparaisons de fréquences d'usages sur des corpus fondamentalement différents. Le premier corpus étant spécifique d'un exercice professionnel et le second provenant d'un autre corpus textuel mais ne relevant d'aucune spécification professionnelle. Pour limiter les effets négatifs propres à la lemmatisation, il aurait fallu envisager une prospection plus fine des contextes à travers les concordances présentes autour des lemmes professionnels ayant été pointés comme identitaires.

Références

- [1] A. Abeille, Traitement automatique des langues, Encyclopédie Universalis
- [2] R. Albert, A-L. Barabási, Statistical mechanics of complex networks, Reviews of modern physics, Volume 74, 2002
- [3] A-L. Barabási, E. Bonabeau, Scale-free networks, Scientific American, May 2003
- [4] R. Barthes, Eléments de sémiologie, Communications, 4, 1964, p 91-135
- [5] J-M. Berthelot, Epistémologie des sciences sociales, PUF, Les sciences du langage et la communication, 149-201, 2001
- [6] D. Bounoux, La communication par la bande, Introduction aux sciences de l'information et de la communication, Editions La Découverte, 1998
- [7] J. Boutet, « La part langagière du travail : bilan et évolution », Langage et société, 2001/4 n°98, p 17-42
- [8] J. Boutet, « Quelques propriétés des écrits au travail », Réseau Langage et Travail, Cahier n°6, 1993
- [9] J. Boutet, La vie verbale au travail, Octares éditions, 2008
- [10] E. Brunet, Lexicometrica n°2, Qui lemmatise dilemme attise, 2000
- [11] P. Charaudeau, D. Maingueneau (sous la direction de), Dictionnaire d'analyse du discours, Seuil, 2002
- [12] M. Cori, J. Léon, La constitution du TAL, Etude historique des dénominations et des concepts, TAL, Volume 43, n° 3/2002, p 21-55
- [13] P. Delcambre, « Ecrire : conditions de travail et place professionnelle », Etudes de communication, 13 | 1992, 47-72
- [14] P. Delcambre, Reconstruire des pratiques de communication à partir d'« objets », documents d'archives, Presses Universitaires de Rennes, 2000

- [15] J. C. Domenjoz, L'approche sémiologique, Ecole des arts décoratifs, Contribution présentée dans le cadre de la session 1 du dispositif de formation 1998-1999 « catégories fondamentales du langage visuel »
- [16] S.N. Dorogovtsev, J.F. Mendes, Language as an evolving word web, The royal society, 2001
- [17] O. Ducrot, J-M. Schaeffer, Nouveau dictionnaire encyclopédique des sciences du langage, Points, Essais, 1999
- [18] R. Eluerd, La lexicologie, PUF, Que sais-je, 2000
- [19] S. Evert, The statistics of word cooccurrences : Word pairs and collocations, PhD dissertation, IMS, University of Stuttgart, 2005 (<http://www.collocations.de>)
- [20] R. Ferrer i Cancho, R. V. Solé, The small-world of human language, Proceedings of the Royal Society of London, B 268, 2261- 2266, 2001
- [21] R. Ferrer i Cancho, The structure of syntactic dependency networks : insights from recent advances in network theory, Problems of Quantitative Linguistics, 60-75, 2005
- [22] B. Gaume, Information interaction intelligence, Journal, vol. 4, no2, pp. 39-96, 2004
- [23] B. Hayes, Graph theory in practice : Part 1 & Part 2, A reprint from American Scientist, Computing Science, Vol 88, Number 1, January-February, 2000, 9-13, Number 2, March-April, 2000, 104-109
- [24] D. Jodelet (sous la direction de), Les représentations sociales, PUF, 1999
- [25] N. Kalampalikis, S. Moscovici, Une approche pragmatique de l'analyse Alceste, Presses universitaires de Liège, Les cahiers internationaux de psychologie sociale, n° 66, 2005
- [26] P. Laburthe-Tolra, J.P. Warnier, Ethnologie Anthropologie, PUF, 1997
- [27] L. Lebart & A. Salem, « Statistique textuelle », 1994, <http://lexicometrica.univ-paris3.fr/livre/st94/st94-tdm.html>
- [28] I. Léglise, Les travailleurs sociaux entre écrits professionnels contraints et oral débridé, Pratiques, langues et discours dans le travail social, Paris : L'Harmattan, 3-8, 2004

- [29] B. Lemaire, Limites de la lemmatisation pour l'extraction de significations, 9eme Journées Internationales d'Analyse Statistique des Données Textuelles, Lyon, 2008
- [30] D. Maingueneau, L'analyse du discours, introduction aux lectures de l'archive, Paris, Hachette Université, 1991
- [31] D. Maingueneau, Les termes clés de l'analyse du discours, Editions du Seuil, Points, 2009
- [32] M. Markořova, Network model of human language, Physica A, 2007
- [33] D. Mayaffre, « L'entrelacement lexical des textes, cooccurrences et lexicométrie », revue électronique Texte et corpus, n°3, 2008, Actes des journées de la linguistique de corpus 2007, p. 91-102
- [34] A. Mehler, Large Text Networks as an Object of Corpus Linguistic Studies, An International Handbook of the Science of Language and Society, De Gruyter, Berlin/New York, pp. 328–382, 2008
- [35] A.E. Motter, A.P.S. de Moura, Y-C. Lai, P. Dasgupta, Topology of the conceptual network of language, Physical Review E, Vol 65, (2002)
- [36] New B., Pallier C., Ferrand L., Matos R. (2001) Une base de données lexicales du français contemporain sur internet : LEXIQUE, L'Année Psychologique, 101, 447-462. <http://www.lexique.org>
- [37] M. E. Newman, The structure and function of complex networks, Department of Physics, University of Michigan and Santa Fe Institute
- [38] A. Polguère, Notions de base en lexicologie, Observatoire de Linguistique Sens-Texte (OLST), Département de Linguistique et traduction Université de Montréal Montréal (Québec) — Canada version préliminaire 2002
- [39] G-E. Sarfati, Eléments d'analyse du discours, Armand Colin, 2007
- [40] K. A. Seaton, L. M. Hackett, Stations, trains and small-world networks, Physica A, 2004
- [41] M.M. Soares, G. Corso, L.S. Lucena, The network of syllables in Portuguese, Physica A, Vol 355, 678-684, 2005

- [42] D. J. Watts, S.H. Strogatz, Collective dynamics of «small-world » networks, Nature, Vol 393, 441-442, 1998
- [43] D. J. Watts, Networks, Dynamics, and the Small-World Phenomenon, Santa Fe Institute, AJS, Volume 105, Number 2, 1999, 493-527
- [44] P. Watzlawick, J. Helmick Beavin, Don D. Jackson, Une logique de la communication, Points, 1979
- [45] F. Yvon, Introduction au traitement automatique des langues naturelles, Limsi, CNRS, 2007
- [46] La prévention spécialisée : enjeux actuels et stratégies d'action. Rapport du groupe de travail interinstitutionnel sur la prévention spécialisée. Janvier 2004
- [47] Prévention spécialisée. [http://fr.wikipedia.org/wiki/Prévention _spécialisée](http://fr.wikipedia.org/wiki/Pr%C3%A9vention_sp%C3%A9cialis%C3%A9e)
- [48] Conseil Général de la Gironde. http://www.gironde.fr/cg33/jcms/c_5217/accueil (faire une recherche avec le mot clé « prévention spécialiALLIERsée »)
- [49] Recommandations de bonnes pratiques professionnelles, Mise en oeuvre de l'évaluation interne dans les établissements et services vidés à l'article L 312-1 de code de l'action sociale et des familles, ANESM, Agence Nationale de l'Evaluation et de la Qualité des Etablissements et Services Sociaux et Médico-Sociaux
- [50] Analyse et traitement informatique de la langue française, <http://www.atilf.fr/>
- [51] Dictionnaire historique de la langue française, sous la direction d'Alain Rey, Le Robert, 2006
- [52] <http://academie-francaise.fr/la-langue-francaise>
- [53] http://sirao.kgf.uni-frankfurt.de/linguistic-networks/pdfs/Mehler_2008_TexT_Networks.pdf
- [54] <http://www.cnrtl.fr/lexiques/morphalou/> (documentation Morphalou 2.0)
- [55] <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/>
« TreeTagger - a language independent part-of-speech tagger »

- [56] <http://www.ansm.sante.gouv.fr>
- [57] <http://www.R-project.org/Licenses/>
- [58] <http://www.ims.uni-stuttgart.de/~schmid/french-tagset.html>
- [59] <http://www.atala.org/-Revue-TAL->
- [60] <http://tulip.labri.fr/TulipDrupal/>
- [61] <http://www.lexique.org/>
- [62] M. Berthet, D. Cru, Travail prescrit, travail réel et santé au travail, revue Travail et Emploi, n° 96, 2003
- [63] E. Benveniste, Problèmes de linguistique générale, II , Tel, Gallimard
- [64] B. New, M. Brysbaert, J. Veronis, C. Pallier, The use of film subtitles to estimate word frequencies, Applied Psycholinguistics, Volume 28, Issue 04, October 2007, pp 661 - 677
- [65] A. Borzeix, B. Fraenkel, Langage et travail, Communication, cognition, action, Ouvrage collectif CNRS Éditions, 2001

Annexes

Présentation des annexes

Les annexes sont destinées à faciliter la compréhension des éléments abordés dans le corps de la thèse.

La première annexe présente les intitulés et les types de présentation ayant eu cours pendant l'ensemble de la période.

Les bilans ont été répartis en huit annexes successives sur l'ensemble de la période étudiée. L'annexe 2 appartient au tout premier bilan, texte dont la nature est très administrative. L'annexe 3 présente un extrait concernant une partie des modalités d'intervention de la prévention spécialisée. L'annexe 4 présente un paragraphe en rapport avec la question de l'incarcération. L'annexe 5 aborde un mode d'intervention avec la tranche d'âge des 16-18 ans. L'annexe 6 présente une manière d'intervenir auprès des 16-21 ans qui occupent l'espace public commun. L'annexe 7 concerne des problématiques éducatives au sein d'une opération de réhabilitation urbaine. L'annexe 8 aborde la question de la visibilité actuelle de la prévention spécialisée

L'annexe 9 concerne le compte rendu de mission officiel de l'année 2010. Celui-ci présente une structure narrative cantonnée aux commentaires libres. Les tableaux qui le composent en partie n'ont pas été joints. Ils sont sous forme de listes ou de chiffres.

Enfin, les trois dernières annexes présentent les éléments suivants :

L'annexe 10, la typologie des classifications morpho-syntaxiques de Tree Tagger.

L'annexe 11, les réponses des éducateurs à la première enquête réalisée en 2010.

L'annexe 12, les réponses des éducateurs à la seconde enquête 2011.

Annexe 1 - Intitulés et présentation des documents étudiés

Année	Titre	Reliure	PE	PB
1972	Rapport administratif 1971-1972	P	23	4
1973	Rapport d'activité 1972-1973	P	21	2
1974	Rapport administratif 1973-74	P	42	2
1975	Rapport administratif 1974-1975	P	63	2
1976	Rapport administratif 1975-76	P	40	2
1978	Rapport administratif 1976-77	P	48	0
1979	Rapport 1978	P	24	0
1980	Rapport 1979	P	37	2
1981	Rapport d'activité 1980	DP	47	0
1982	Bilan de l'activité 1981-1982	DP	51	0
1983	Rapport d'activité 82 - 83	DP	47	0
1984	Rapport d'activité 1984	T	56	2
1985	Bilan d'activité 1985	T	42	0
1986	Rapport d'activité	T	54	1
1986	1986-1987	T	32	1
1987	Bilan 1987	T	72	1
1988	Bilan 1988	T	46	1
1989	Bilan 1989	T	72	1

TAB. 1 – Tableau des intitulés annuels

Reliure : P *Pince*, DP *Double Pince*, T *Thermocollée*, PP *Peigne Plastique*

Nombre de pages écrites : PE

Nombre de pages blanches : PB

Année	Titre	Reliure	PE	PB
1990	Bilan 90	T	65	1
1991	Bilan 1991	T	40	1
1992	Bilan d'activité 1992	T	35	1
1993	Bilan d'activité 1993	T	48	1
1994	Rapport d'activité 1994	T	61	1
1995	Rapport d'activité 1995	T	72	1
1996	Rapport d'activité 1996	PP	62	1
1997	Rapport d'activité 1997	PP	70	0
1998	Rapport d'activité 1998	PP	67	0
1999	Rapport d'activité 1999	PP	60	1
2000	Rapport d'activité 2000	PP	43	1
2001	Rapport d'activité 2001	PP	48	1
2002	Bilan associatif 2002	PP	61	1
2003	Bilan associatif 2003	PP	57	0
2004	Bilan associatif 2004	PP	58	0
2005	Bilan associatif 2005	PP	35	0
2007	Bilan d'activité 2007	PP	39	0
2008	Bilan d'activité 2008	PP	38	0
2009	Bilan d'activité 2009	PP	60	0
2010	Bilan d'activité 2010	PP	59	0

TAB. 2 – Tableau des intitulés annuels

Reliure : P *Pince*, DP *Double Pince*, T *Thermocollée*, PP *Peigne Plastique*

Nombre de pages écrites : PE

Nombre de pages blanches : PB

Annexe 2 - [Extrait Bilan 1972, pages 9-10]

Les équipes. L'équipe médico-psychologique. On a dit, dans le précédent rapport, l'accent que l'on souhaitait mettre en ce qui concerne l'apport de l'équipe médico-psychologique. La preuve est faite de l'importance de l'assistance psychothérapique - spécificité de l'action des équipes de prévention - en fonction des clientèles auxquelles elle s'adresse. La même équipe médico-psychologique assiste les équipes des Clubs de Bègles et Talence. L'entrée en fonctions de Mr XXXX, psychologue aux côtés de Mr XXXX permet de donner aux études de cas le temps indispensable facilitant la personnalisation de l'action des éducateurs. L'apport du sociologue s'avère fructueuse au niveau de l'appréhension des problèmes dans leur contexte global, la théorisation des problèmes vécus et la réflexion sur les enquêtes sociologiques. L'équipe regroupe donc : - Mr le Dr G. XXXX, psychiatre animateur de l'équipe. - Mr M. XXXX, psychothérapeute, superviseur. - Mr XXXX, psychologue. - Mr XXXX, sociologue. - Mr le Dr F. XXXX, médecin généraliste. L'équipe technique de supervision. Elle regroupe : - Les membres de l'équipe médico-psychologique. - Les membres du bureau. - L'équipe éducative. - Stagiaires et moniteurs. Les réunions du bureau, prévues pour être mensuelles, sont très vite devenues bimensuelles. Leurs membres se réunissent, en réalité, aussi souvent que nécessaire. L'équipe éducative. Il apparaît évident que c'est à ce niveau Annexe 4 [Extrait bilan 2008] au que doivent désormais porter les efforts. On sait que les éducateurs ont - très volontairement - limité leur action, depuis le démarrage, à la cité d'urgence du Haut-Verduc primitivement dénommée Cité du crime". La dégradation de la vie sociale dans ce secteur justifiait amplement ce choix. La politique suivie a indiscutablement porté ses fruits. Les statistiques fournies par Monsieur le Chef de la Sûreté et le Chef de la Brigade des Mineurs (4 février 1972) font apparaître un abaissement extrêmement encourageant du taux de délinquance juvénile (0,02 contre 0,04) plaçant la commune de Bègles, en ce domaine, au premier rang parmi l'ensemble des communes et secteurs de la Communauté Urbaine (le taux cité est calculé par rapport à la densité globale de la population du secteur communal). Ce sont aujourd'hui les grands ensembles voisins (les cités du Dorat entre autres) qui sollicitent l'intervention de l'équipe éducative dont l'action devra bien dans les années à venir s'étendre aux dimensions de la commune. L'équipe, dans sa composition actuelle, ne saurait sérieusement y prétendre. Il appartient aux Ministères de Tutelle et aux membres du Conseil Général d'apprécier si la sollicitation d'un poste d'éducateur spécialisé complémentaire est - ou non - justifiée. Les conventions de stages passées avec l'École d'Éducateurs Spécialisés de Bordeaux d'une part, l'Institut Universitaire de Technologie d'autre part, permettront, si le poste sollicité est accordé, de présenter cette année

une équipe très charpentée ainsi composée : 3 éducateurs spécialisés dont un directeur pédagogique, 1 éducatrice spécialisée diplômée, 1 animateur socio-culturel diplômé, 2 stagiaires d'école (2ème année) conventionnés, 2 animateurs stagiaires IUT conventionnés. *

Annexe 3 - [Extrait bilan 1990, pages 25-26]

Si la prévention spécialisée est un mode d'approche spécifique d'une population très marginalisée, elle se caractérise aussi par la diversité de ses pratiques. Cette diversité, d'ailleurs, soulève l'ambiguïté d'être à la fois une richesse d'intervention (l'institutionnalisation des pratiques ne serait-elle pas la mort de la Prévention Spécialisée ?) et d'établir un "flou" certain au niveau même de la notion de Prévention Spécialisée (cf le rapport de la Commission des maires de 1982 qui n'explique à aucun moment le concept de Prévention Spécialisée). Au niveau de Service de Prévention de Bègles cette interrogation sur nos pratiques, et notamment nos pratiques d'actions partenariales, nous semble intéressante à affiner. On peut noter, en effet, notre présence dans des domaines multiples, à savoir : les OPE, la réhabilitation d'ensembles sociaux dégradés, les missions locales d'insertion professionnelle et sociale des jeunes, les actions d'insertion, les entreprises intermédiaires, le Conseil Communal de Prévention de la délinquance, les actions de soutien scolaire, etc. Si, par l'étendue du champs de notre intervention, nous devenons des interlocuteurs reconnus par l'ensemble de nos partenaires, ne risquons-nous pas d'y perdre un peu de notre identité et de tout ce qui fait la spécificité de la Prévention Spécialisée ? "Une équipe de Prévention, c'est d'abord et avant tout une équipe de permanents auxquels se joignent différents vacataires (psychiatre, sociologue) et dont le travail consiste à mener auprès des jeunes et particulièrement des plus marginalisés, une action éducative visant à faciliter leur insertion sociale tout en les aidant à accéder à l'autonomie" Circulaire 17 octobre 1972. Nous intervenons donc à trois niveaux : l'action éducative menée auprès des adolescents (14-20 ans) les plus marginalisés des quartiers reste, bien entendu, notre priorité. L'article 5 de l'arrêté du 4 juillet 72 précise, au préalable, que "cette insertion doit se faire par des moyens spécifiques supposant notamment leur libre adhésion". L'article 2 du projet de convention adopté le 4 mai 1973 stipule que "l'anonymat des jeunes doit être respecté". Nous tentons d'élargir notre population vers des tranches d'âge plus jeunes (8-13 ans) en vue de prévenir la rupture sociale vécue au moment de l'adolescence. On observe, en effet, que les plus jeunes sont souvent amenés à reproduire le comportement des aînés. De plus, le travail de prévention étant, par définition, "d'agir avant" qu'un problème se pose, il nous paraît intéressant de pouvoir travailler par la suite, avec des adolescents qu'on a connus plus jeunes. Ceci nous amène à prendre en compte les difficultés spécifiques de ces enfants (échec scolaire etc.) mais aussi leur cadre familial.

Annexe 4 - [Extrait bilan 1995, pages 45-46]

La prison, du peu que j'en connais au cours de mes visites à de jeunes détenus est un monde de répétition. Les murs, les portes, les sas, les bruits se répètent sans fin. Comme les heures et les minutes de ceux qui y purgent leur peine. Une seconde d'oubli et qui êtes-vous devenu ? Celui dont plus personne ne se souvient. Relégué. Abandonné. Que pensez-vous de l'accessibilité à la peine ? N'est-elle pas fondamentale ? Celui qui n'y peut parvenir va en prison sans espoir d'en sortir positivement. J'ai vu les réactions de jeunes qui sont incarcérés une première fois. Celles de ceux qui sont incarcérés une seconde fois, ou même une troisième. La prison est un monde. Un jeune détenu peut y passer des mois sans y rencontrer une personne de l'extérieur. Des jeunes n'y reçoivent pas de visites de leur famille. Peut-être, mais rarement la visite des éducateurs du service social de la prison, qui eux-mêmes sont en perpétuel sous-effectif. Les niveaux scolaires sont très faibles, c'est-à-dire plus un jeune de quartier a un faible niveau scolaire, plus il existe de probabilités pour lui d'atterrir en prison. Les jeux ne sont pas égaux. Si un jeune de quartier est incarcéré, jugé, puis condamné, il ne lui est pas si facile de jouir des droits de libération conditionnelle ou de semi-liberté que la loi lui accorde. Je m'occupe d'un jeune bèglais, incarcéré à la maison d'Arrêt de Neuvic depuis plusieurs mois, qui a largement dépassé la moitié de sa peine d'incarcération, et pour qui un projet de sortie a été construit en accord avec la PAIO, et qui se trouve encore en prison et surtout n'a toujours pas reçu (trois mois plus tard !) de réponse sur sa situation. Enquête, enquête, enquête. Peut-être n'arrivera-t-elle que trois semaines avant sa libération définitive ? mais que voudra-t-elle dire alors ? Certains pensent probablement, au nom du confort de tous les jours, qu'il est bien qu'il se trouve là-bas. mais que fera-t-on après ? D'une façon ou d'une autre, il sortira.

Annexe 5 - [Extrait bilan 2008, page 16]

Le travail de rue reste pour les jeunes de cette tranche d'âge [16-18 ans] un outil central. Il s'agit d'un groupe très présent sur le quartier à toutes heures de la journée. Ceci pour la simple raison que les situations individuelles vis-à-vis de l'école ou du travail sont multiples. C'est un âge où la demande d'espace de liberté est de plus en plus importante. C'est aussi un âge de recherche de sens à sa vie liée à une perte de repères qui se manifeste sous plusieurs formes (déscolarisation, errance, formation interrompue, intérim, problème de justice). La relation que nous entretenons avec ces jeunes parfois depuis plusieurs années fait que nous tentons de les accompagner dans cette recherche de sens. C'est un âge où la confrontation des points de vue par la parole est centrale. Ils nous poussent donc régulièrement à nous positionner sur des thèmes tel que la religion, le travail, l'homosexualité, la délinquance, la sexualité, la politique. Ils tentent, par ces échanges, de se faire une idée et de se construire un discours propre. Notre rôle est de les amener à penser par eux-mêmes en proposant une vision décalée de leurs références habituelles. Nous nous sommes également appuyés sur des activités de groupe : chantiers, activités sportives, camps à la montagne. Néanmoins, ces supports ne sont que des prétextes à une relation individuelle qui nous permettront d'aborder certaines problématiques dont notamment celles concernant l'emploi, la formation, l'école, la famille, par des recherches, des accompagnements et des médiations. D'autres rencontres peuvent être axées principalement sur leurs questionnements liés à l'avenir. La question sous-jacente et récurrente à ces échanges peut se résumer par un "Qu'est-ce que tu peux dire de ça, car moi je ne sais pas ce que je veux ni où je vais".

Annexe 6 - [Extrait 1 bilan 2009, page 24]

Souvent rencontré sur le même lieu [les 16–21 ans], c'est un groupe important, composé essentiellement de garçons issus pour quelques uns du quartier, les autres venant de différents quartiers bèglais. Ce groupe est incontournable sur le quartier et il occupe un espace précis et repéré. Cette année, ces jeunes sont très présents sur le quartier et ce tout au long de la journée, la majorité étant sans emploi ou sortis du système scolaire. La consommation de haschich et d'alcool se pose particulièrement, celle-ci devient de plus en plus importante et semble complètement banalisée. Ils passent leur temps dans les halls sans but précis et ne montrent pas de volonté dans la recherche de quelconques activités. Ce groupe évolue dans sa composition mais pas dans son nombre (une vingtaine d'individus). L'arrivée d'un nouvel éducateur sur le quartier a modifié la relation existante avec ce groupe. Nous avons donc consacré une partie de l'année à tisser un nouveau lien par une présence et une attention régulière. Ce travail s'est essentiellement basé sur la parole mais aussi avec des échanges vifs justifiant notre présence, notre utilité et parfois notre personnalité, moments préalables à tout début de relation éducative. Nous avons fait le choix avec ce groupe de ne pas proposer et faire d'activités de loisirs. En effet, nous privilégions l'accès au travail et l'insertion sociale (recherche logement, démarches administratives, transport et mobilité ?). D'autre part ce groupe a la possibilité de se rapprocher du Service Jeunesse et de participer aux différentes activités proposées. De plus, ils ont une certaine autonomie dans l'organisation de leurs loisirs, matérielle et financière. Au fil de l'année, ce travail de rue nous a permis d'enclencher des démarches d'insertion sociale et professionnelles avec le tiers de ce groupe. Le chantier environnement à Hostens a été un outil pertinent pour quatre d'entre eux qui étaient éloignés ou en rupture avec l'emploi. En effet, durant cette période nous leur avons permis de retrouver un rythme de vie plus adapté au monde du travail, à modérer et à évoquer leur consommation de cannabis et à consolider notre relation. Le chantier a permis pour certains de redynamiser leur parcours de formation scolaire et pour d'autres à enclencher une recherche d'emploi. Pour l'autre partie du groupe le seul outil dont nous disposons reste la parole et l'objectif est la mise au travail.

Annexe 7 - [Extrait 2 bilan 2009, page 33]

2009 entérine une réelle transformation des différents lieux d'intervention. Cette transformation se joue à travers deux paramètres. Le premier concerne l'augmentation du nombre des quartiers dont sont issus les jeunes auprès de qui nous sommes intervenus et souligne l'accroissement de la superficie de la zone concernée. Le second pointe un amoindrissement relatif de notre visibilité en raison de cette extension de surface. Le quartier Yves Farge / Terres neuves est une zone urbaine en cours d'opération de renouvellement urbain (ORU) qui regroupe des îlots d'habitations jouxtant des entreprises et une école d'infirmières, ainsi qu'un ensemble de salles thématiques (concerts, expositions, sports, créations musicales, restaurant) tous hébergés dans d'anciens entrepôts militaires rénovés ou en cours de rénovation. Il est économiquement et culturellement dynamique, mais à notre connaissance aucun habitant du quartier ne se trouve être dans les bénéficiaires des emplois induits par ces entreprises. Ce nouveau quartier brasse beaucoup de monde en raison du terminus du tramway qui favorise la mobilité et la circulation des personnes. L'ouverture du quartier et la facilité d'accès aux transports collectifs font qu'il devient passant, traversant. On peut voir circuler des personnes que l'on ne connaît pas, qui ne sont pas de là. Il y a de nouveaux axes routiers, et toujours des travaux en cours. L'année dernière, nous écrivions in extenso XXXX et moi : "Outre les désagréments causés par les travaux, nous croyons que le plus difficile est la situation des habitants qui attendent d'être relogés et qui vivent dans des bâtiments quasiment vidés des trois quarts de leurs locataires (tour F). Les multiples paliers avec des portes d'appartement scellées ne font rien pour améliorer la perception des habitants". Cette année, je peux écrire la même chose pour la tour E. Tour actuellement habitée par au plus dix familles avec plusieurs étages sans aucun habitant. Les jeunes vivent cette période sur un mode d'appropriation exclusif, au strict détriment du sentiment d'habiter chez soi pour les derniers locataires, voire de se sentir en danger. Yves Farge reste un quartier attractif notamment pour les jeunes et beaucoup de quartiers environnants qu'ils soient béglais ou bordelais proches de Bègles (Paul Eluard, La Maye de Bernet, le Treuil, Carle Vernet, Le Renard). De même, il attire des jeunes venant d'autres secteurs plus éloignés (gare, rive droite, Les Aubiers) via le tram. Je suis attentif à ces nouveaux jeunes. J'apprends à les connaître quand ils reviennent régulièrement.

Annexe 8 - [Extrait bilan 2010, pages 46-47]

Le visible et l'invisible. Il est évident que le fonctionnement d'un service comme le nôtre est possible dans une société qui a les moyens pour l'entretenir. Actuellement on constate que nous allons peu à peu vers un système libéral où les choix paraissent clairs, faire des économies, rentabiliser et produire. On a l'impression dans le social de devenir un "outil marchand" qui doit être clair, précis, rentable, visible et évaluable. Sommes-nous visibles ? Je ne pense pas que notre visibilité puisse contenter la vision libérale. Nous sommes visibles dans ce que nous faisons concrètement. Notre bilan d'activité parle d'activités socialisantes pour les jeunes, des camps, sorties diverses, chantiers, visites, activités, fêtes de quartier et tant d'autres. Selon notre discours d'éducateur nous prenons appui dans ces temps partagés pour construire une relation dans laquelle ces suivis éducatifs se construisent. Que deviennent-ils ces suivis ? Peut-on les rendre visibles ? Si on se limite à ce qui est visible étant donné l'évolution de la société, il ne serait pas étonnant à mon avis que certains politiques demandent de faire plus avec moins de moyens et on verra apparaître des services apparentés au social qui chercheront la concurrence et pourquoi pas le profit. Face à ce problème les éducateurs peuvent être tentés de faire plus de visible et plus de concret en se servant de ce concret comme d'une finalité et pas comme un moyen, courant ainsi à notre propre perte ? Mais notre travail ne s'arrête pas là. Nous sommes confrontés la plupart du temps à des problématiques qui demandent du temps et des réponses qui doivent être élaborées par les jeunes eux-mêmes. L'éducateur rencontre un jeune qui a un comportement exécrable à l'école, il se sent dévalorisé et à la maison ne trouve pas un terrain qui puisse l'aider. Un autre jeune ressent des angoisses il est souvent violent avec l'entourage. Un autre se trouve seul ses parents sont partis il est livré à lui-même. Un autre pense qu'il ne vaut rien, il traîne, ne veut rien faire. Un autre a un certain niveau d'études mais ne trouve pas de travail. Un autre vole, casse et n'a confiance en personne. Toutes ces souffrances produisent de la violence. Les réponses ne peuvent pas être rapides, l'éducateur est submergé dans ces problématiques en essayant d'aider et de soutenir ces tranches de vie difficiles. mais comment parler de ça ? Comment marchandiser ce travail ? Un travail qui a besoin d'être connu pour exister et en même temps qui a besoin de silence. Cette position nous situe toujours à la marge du niveau institutionnel et la marge est notre place ? mais jusqu'à quand ? Se battre et résister ce n'est pas que défendre nos fondamentaux mais lutter pour que le pouvoir politique comprenne cette dimension. La notion de ne pas être visible dans ce travail fait appel, questionne, à mon avis à la notion de confiance, car face à l'invisible et à des notions de libre adhésion, secret professionnel, on peut mettre et faire n'importe quoi. Mais la question est la

suivante : est-ce que dans notre société on peut croire encore que des hommes et des femmes peuvent avoir envie d'être empathiques avec d'autres qui souffrent ou essayent de tricoter des solutions et des accompagnements qui peuvent être aidants ? En somme soutenir quelque chose d'un peu utopique ?

Annexe 9 - Compte rendu de mission 2010

Compte rendu de mission, Année : 2010, Association de prévention spécialisée de Bègles

Les zones grisées sont destinées à recevoir vos commentaires. Pour le reste des informations reportez-vous aux tableaux Excel.

I- TERRITOIRE D'INTERVENTION, remplir tableau1.xls

Commentaire : *Nous intervenons sur la commune de Bègles, partagée en trois secteurs dont deux sont en très grande évolution. Le quartier des Lacs se compose de quatre quartiers distincts : Le Verduc, Le Paty, Les Sècheries et Le Dorat. Nous intervenons principalement sur les deux premiers. Pour l'année civile 2009, le secteur Yves Farge a confirmé une réelle extension, passant de deux quartiers côte à côte (Yves Farge et Terres Neuves) à un ensemble de cinq quartiers repartis sur un secteur géographique de superficie au moins trois fois supérieure. Les résidences Maurice Thorez et Goélands (3ème secteur d'intervention) connaissent un début de transformation notable (ORU), qui a des incidences sur la composition de la population et au niveau de l'ambiance du quartier. L'intervention sur Mussonville décidée fin 2009 a été suspendue en septembre 2010 car elle n'a pas suscité de besoins précis.*

II- POPULATION CONCERNÉE (jeunes rencontrés au cours de l'année)

1. Nombre de jeunes rencontrés par sexe et âge, Remplir tableau2.xls 2. Situation des jeunes vis à vis de la scolarité ou de l'emploi, remplir tableau3.xls 3. Situation des jeunes vis à vis des mesures éducatives et du soin, remplir tableau4.xls 4. Situation des jeunes vis-à-vis du logement, remplir tableau5.xls

Commentaire : *Les jeunes que nous rencontrons et avec qui nous avons une relation éducative habitent sur les territoires prioritaires. Cela signifie que leurs situations (scolarité, emploi, ?) sont pour partie influencées par leur milieu de vie et leur classe sociale. Il s'agit de personnes qui sont moins pourvues en capital financier par rapport à d'autres lieux d'habitations. Nous avons constaté cette année une stabilité dans le nombre de jeunes suivis dans les tranches d'âge 13/17 et 18/24 ans. Par contre, on note une augmentation du nombre de jeunes rencontrés dans la tranche d'âge 10/12 ans.*

III- PARTENARIAT

1. Partenariat institutionnalisé, remplir tableau6.xls 2. Partenariat conjoncturel (lié à un projet particulier), remplir tableau7.xls Il n'est pas nécessaire de détailler le nombre de rencontres de travail entre les éducateurs et les autres travailleurs sociaux autour de situations particulières de jeunes (synthèses, CTP)

Commentaire : *Le partenariat institutionnalisé n'est pas lié à un projet précis mais permet, à partir d'un intérêt commun pour des difficultés spécifiques de la jeunesse*

habitant des secteurs prioritaires, à partir également de rencontres soit régulières, soit à la demande, de renforcer notre travail auprès des jeunes. Le partenariat conjoncturel peut se mettre en place à l'initiative des partenaires ou de nous-mêmes de manière plus ou moins formalisée. Il s'agit d'un outil de travail au service de la relation élaborée avec des groupes ou des individus. Il permet de promouvoir la dynamique du quartier.

IV- PROBLÉMATIQUE(S) DES JEUNES REPÉRÉE(S) SUR LE TERRITOIRE D'INTERVENTION, remplir tableau8.xls

Commentaire : Nous avons repéré diverses problématiques qu'il nous semble difficile de classer par ordre d'importance. Notre intervention éducative tient compte de ce que nous dit le jeune et de ce que nous pouvons analyser de sa demande. Il s'agit d'une intervention éducative globale. Les problématiques peuvent être illustrées notamment par : - Un sentiment de victimisation. - Une difficulté à l'engagement. - Un excès à la consommation. Le travail mené autour des questions de mixité, d'inter quartier a des effets sur la place, l'importance des groupes de référence.

V- PROBLÉMATIQUE(S) PRIORISÉE(S) POUR L'ANNÉE, remplir tableau9.xls

Commentaire : L'ordre est peu important car nous travaillons avec l'analyse de la prégnance de chaque problématique pour les jeunes.

VI- RÉPONSES DÉVELOPPÉES PAR L'ASSOCIATION

1. Travail de rue : Temps de présence des éducateurs sur les lieux de vie des jeunes, sur l'environnement social et les groupes, remplir tableau10.xls

Commentaire : Le temps de travail de rue, base de notre action, est variable selon les projets en cours. Il reste malgré tout régulier sur l'année. Lieux priorités : Résidence Yves Farge, Terres Neuves, Près Lacoste, Maye de Bernet et Paul Eluard, les résidences Maurice Thorez, Goélands et Mussonville, le Verduc, le Paty, les Seycheries et le Dorat. Groupes de jeunes ciblés : groupes naturels de différentes classes d'âge, groupes d'école, rares groupes inter quartiers (filles), groupes construits autour d'un projet. Temps de présence des éducateurs/activité totale : pied des bâtiments, dans la commune, dans la rue, avec un groupe ou un individu en accompagnement éventuellement à l'extérieur de la commune représente environ 60 % du temps.

2. Actions collectives ; Habituelles, remplir tableau11.xls, Liées à un besoin identifié, remplir tableau12.xls

Commentaire : Pour nous, la dimension collective d'une action se définit par l'effet qu'elle génère sur le territoire concerné. Les actions collectives habituelles sont préférentiellement mises en place en lien avec les partenaires qui interviennent sur le territoire. Même lorsqu'elles sont habituelles, ces actions s'adressent à des jeunes différents d'une année sur l'autre et prennent en compte leurs "besoins". L'objectif principal est de créer et développer le lien social. Nous mettons en place des actions de groupe qui

sont en lien direct avec la relation éducative entretenue avec les jeunes qui composent ces groupes. Cela se traduit par diverses actions (sorties, camps, chantiers) qui de ce fait sont toujours liées à un repérage des problématiques. Nous nous appuyons pour cela sur des ressources locales (jeunes, familles, autres ?) ou extra locales. Les objectifs de ces actions sont multiples, à la fois individuels et collectifs : Objectifs plutôt d'ordre éducatif/socialisation pour les plus jeunes (10-15 ans). Objectifs visant plutôt l'insertion sociale et professionnelle pour les plus âgés (16-25 ans).

3. Travail individuel ; Permanences ? ?Oui ?Non Pourquoi ? Si permanences, Remplir tableau13.xls

L'histoire de l'APSB nous a conduit à privilégier le travail de rue et la présence sur les quartiers, sans s'inscrire de façon définitive sur une permanence. Notre intervention n'est pas une réponse dans l'urgence mais dans la régularité. Néanmoins le local administratif (qui est excentré) fait parfois office de lieu d'accueil ou d'écoute, et permet aux jeunes de bénéficier de la possibilité d'accéder à Internet, de faire quelques photocopies, d'être aidés pour l'écriture de courriers, de dossiers ? Nous pouvons également utiliser ce lieu pour des réunions avec des jeunes (en individuel, en petit groupe) pour la préparation d'examens, de sorties, de projets. Il permet aussi de formaliser un accompagnement en dehors de notre territoire d'intervention, lui donnant davantage de sens dans l'objet de l'Association. Suivant les secteurs d'intervention (Goélands, Yves Farge, ?), nous pouvons utiliser ponctuellement une salle sur place pour des réunions avec les jeunes.

Suivis individuels ; * Origine de la rencontre, remplir tableau14.xls * Problématiques dominantes traitées dans la relation éducative, remplir tableau15.xls

Commentaire : Tableau 15 : *Une comptabilité analytique des problématiques traitées dans les suivis individuels est très difficile à réaliser, sachant que la rencontre avec les jeunes se fait principalement à travers le travail de rue, ou plus ponctuellement par le biais d'actions de groupe. Nous insistons sur la notion d'accompagnement global où les problématiques sont intimement liées. Notre intervention dans le temps permet de suivre le jeune autour de plusieurs problématiques, qui bien souvent se cumulent. Nous constatons un manque en ce qui concerne l'insertion sociale et la socialisation au sens large. C'est pour cela que nous privilégions cette porte d'entrée. Nous insistons sur la notion d'accompagnement global où les problématiques sont intimement liées. Notre intervention dans le temps permet de suivre le jeune dans cette imbrication.*

4. Répartition de l'activité globale selon les réponses développées, remplir tableau16.xls

Commentaire : *Le comptage du temps est plutôt incompatible avec l'attention nécessaire à une relation éducative sérieuse. Cette difficulté se double d'un recouvrement très réel des divisions proposées.*

5. Travail avec les familles

Commentaire : *Le travail avec les familles se fait de façon ponctuelle et en particulier en fonction des classes d'âge (plus les enfants sont jeunes, plus le travail avec les familles est important). Le travail avec les familles se fait par des rencontres qui peuvent avoir plusieurs origines : soit à notre demande dans le cadre d'une activité et pour une signature d'autorisation parentale, soit avec l'accord du jeune dans le cadre d'un travail autour d'une problématique spécifique, soit encore à la demande du jeune ou de sa famille dans le cadre d'une médiation, entre eux, ou avec une institution. Le conflit est rarement absent d'un travail fort avec les familles. La question de la parentalité et de la place des parents fait partie intégrante et a une place très importante dans notre réflexion d'équipe (analyse des pratiques notamment).*

VII- REMARQUES PARTICULIÈRES, Commentaire : néant

Annexe 10 - Classification tags Tree Tagger

ABR	abbreviation
ADJ	adjective
ADV	adverb
DET :ART	article
DET :POS	possessive pronoun (ma, ta, ...)
INT	interjection
KON	conjunction
NAM	proper name
NOM	noun
NUM	numeral
PRO	pronoun
PRO :DEM	demonstrative pronoun
PRO :IND	indefinite pronoun
PRO :PER	personal pronoun
PRO :POS	possessive pronoun (mien, tien, ...)
PRO :REL	relative pronoun
PRP	preposition
PRP :det	preposition plus article (au,du,aux,des)
PUN	punctuation
PUN :cit	punctuation citation
SENT	sentence tag
SYM	symbol
VER :cond	verb conditional
VER :futu	verb futur
VER :impe	verb imperative
VER :impf	verb imperfect
VER :infi	verb infinitive
VER :pper	verb past participle
VER :ppre	verb present participle
VER :pres	verb present
VER :simp	verb simple past
VER :subp	verb subjunctive present

Annexe 11 - Première enquête (8 répondants)

Educ_1 -> 12 items Le travail de rue
La notion de temps
Les rencontres
Le travail individuel et collectif
Les situations de précarité
La place des filles
Les difficultés et le décrochage scolaire
Les notions d'errance
Les supports activités/chantiers
Intérêt des camps et des temps de vie collective
Les notions de prise de risque

Educ_2 -> 9 items Travail de rue
Accompagnement social et éducatif
Travail auprès des groupes
Actions collectives
Partenariat
Présence sociale
Travail avec les familles
Travail avec les filles
La scolarité

Educ_3 -> 12 items Scolarité
Formation
Violence
Incivilité
Délinquance
Conduite à risque
Emploi
Autonomie
Rapport avec les autres
Partenaires
Travaux chantiers

Activités loisirs

Educ_4 -> 12 items Précarité

Délinquance
Scolarisation / Formation
Parentalité
Emploi
Errance
Incivilité
Addiction
Mixité
Santé
Isolement
Violence

Educ_5 -> 6 items Autonomie

Déscolarisation
Investissement des jeunes
Violence
Travail
Emploi

Educ_6 -> 9 items Errance

Difficulté des jeunes filles
Investissement des habitants
Evolution des territoires d'intervention
Difficultés scolaires
Insertion professionnelle
Conduites addictives
Incivilités
Violences des jeunes

Educ_7 -> 12 items Addiction

Déscolarisation

Comportements d'échec - Conduite à risque

Insertion

Problématique des filles face à la famille

Problématiques face à l'école

Violence

Incivilités

Implication des habitants dans la vie sociale

Gestion période de vacances

Logement pour les jeunes

Insertion après période carcérale

Educ_8 -> 12 items Le travail de rue

Accompagnement éducatif

Les sorties, les camps

Les chantiers

Le travail avec les familles

Evolution des quartiers

Partenariat

Problématique scolaire

Le travail avec les filles

La promotion du milieu (actions collectives)

Le contexte économique, politique et social global

La notion de temps et d'histoire dans la relation

Annexe 12 - Seconde enquête (6 répondants)

Educ_1 TRAVAIL DE RUE :

éducatif, présence, rencontre, lien, discussion, parole, déambuler, marcher, observation, régularité, base

ACTIONS COLLECTIVES

partenariat, quartier, fête, repas, animations, habitants, participation, réunions, préparation, dynamisation

FILLES

isolement, rapport parent, mixité, diversité, réussite, séparation, fuyante, contraception, relation amoureuse

PRECARITE

emploi, situation, contrat, affective, logement, familiale, intérim, avenir, projection, financière

SCOLARITE

échec, réussite, collège, accompagnement, devoir, difficulté, avenir, formation, orientation

ERRANCE

pas de solution, psychologique, présence, groupe, isolement, déambuler, désorienter, repère, projection, absence/présence

ACTIVITES

sportives, culturelles, nature aventure, consommation, souvenir, histoire, éducative, ouverture, diversité, changement

CHANTIERS

éducatif, travail, argent, échanges, apprentissage, horaires, expérience, sécurité, tremplin, ouverture

FAMILLES

monoparentale, précarité, travail, médiation, conflit, séparation, relationnel, autorité

PARTENARIAT

MDSI, action collective échange, élaboration, coordination, conflit, médiation, animation, orientation, droit commun

Educ_2 TRAVAIL DE RUE

régulier, groupe, temps, quartier, doute, jeunes, échange, rencontre, seul, marcher

ACTIONS COLLECTIVES

socialisation, partage, échange, richesse, différence, projet, animation

FILLES

mixité, seule, garçon, orientation, mère, relation, culture, école

PRECARITE

logement, travail, argent, jeune, affective

SCOLARITE

école, échec, orientation, violence, difficulté, adolescence

ERRANCE

solitude, rue, affective, alcool, drogue

ACTIVITES

camps, échange, culture, groupe, consommation

CHANTIERS

débroussaillage, hostens, groupe, majeurs, travail, salaire, contrat, rythme, discussion, effort, savoir-faire

FAMILLES

parents, mère, relation, travail, éducation, monoparentale, difficulté

PARTENARIAT

différence, complémentarité, projet, action, réunion

Educ_3 TRAVAIL DE RUE

rencontre, discussions, salutations, confidences, promenade, écoute, incruste, tranche de vie, échange, confiance

ACTIONS COLLECTIVES

dynamiser, rencontre, animation, vivre ensemble, apprentissage, envie

FILLES

difficiles à approcher, fuyantes, confidences, accroche, groupe, discussions, accompagnement

PRECARITE

argent, logement, carences, éducatif, aide, discussion, orientation

SCOLARITE

collège, orientation, mixité, difficultés, réussite

ERRANCE

rue, précarité, argent, logement, boulot, chien, toxicomanie

ACTIVITES

partage, ludique, envie, découverte, sport

CHANTIERS

travail, dépassement, cohésion, budget, financement, projet, effort, limites

FAMILLES

discussion, jeunes, enfants, éducatif, écoute, accompagnement

PARTENARIAT

entente, projet, missions, objectifs, animation, rue, divergences, convergences

Educ_4 TRAVAIL DE RUE

quotidien, primordial, base, rencontre, jeunes, groupes, vide, proximité, échange, solitude

ACTIONS COLLECTIVES

ensemble, groupe, travail avec l'autre, socialisation, partage, mixité, différence, préados

FILLES

mixité, groupe, sortie, absence, culture, maison, intérieur

PRECARITE

Saemcib, emploi, jeunes, logement, santé, loyer

SCOLARITE

echec, difficile, garçon, apprentissage, orientation, stage, relation avec l'institution

ERRANCE

temps, jeune, précarité, souffrance, relation à l'autre, groupe d'appartenance

ACTIVITES

loisir, groupe, mer, jeunes, nature, sortie, voyages, cultures

CHANTIERS

été, travail, découverte, ailleurs, socialisation, effort, patience, manuel, Hostens, outil

FAMILLES

travail, mère isolée, famille nombreuse, relation ados-parents, autorisation parentale

PARTENARIAT

difficile, service jeunesse, coordination de proximité, mairie, Remue-Ménages, complexité, positionnement

Educ_5 TRAVAIL DE RUE

égalité, fraternité, douleur, proximité, fatigue, amitié, espoir, réalité, pudeur, froid, chaud

ACTIONS COLLECTIVES

espoir, fraternité, construction, lutte, sociabilité, gain, amitié, dépassement, ouverture, richesse

FILLES

fleur, amour, féminité, fragilité, force, espoir, vie, complément, aide, sexualité

PRECARITE

appauvrissement, solitude, désocialisation, tristesse, attente, espoir, jalousie, désespoir, rebondir, mort

SCOLARITE

socialisation, désocialisation, espoir, désespoir, futur, ouverture, fermeture, amitié, apprentissage, solitude

ERRANCE

fermeture, ouverture, espoir, possibilité, chemin, douleur, méconnaissance, pudeur, dépassement, luxe

ACTIVITES

sociabilité, fraternité, amitié, ouverture, dépassement, apprentissage, égalité, amour, découverte, construction

CHANTIERS

découverte, dépassement, construction, argent, danger, fatigue, accomplissement, plaisir, souffrance, apprentissage

FAMILLES

union, amour, haine, équilibre, souffrance, jalousie, solitude, ouverture, enfermement, dictature

PARTENARIAT

complément, difficulté, besoin, vie, richesse, ouverture, sociabilité, fortifiant, indispensable, souffrance

Educ_6 TRAVAIL DE RUE

déambulation, posture, quotidien, présence, absence, proximité, échange, regard, rencontre, relation

ACTIONS COLLECTIVES

partage, groupe, expérience, support, histoires, projet commun, individualités, confrontation, concession, vie

FILLES

protégées, parentalisées, esseulées, autonomes, enfermées, groupées, mixité, éducation, place, société

PRECARITE

argent, travail, logement, orienter, humilité, peur, angoisse, dévalorisation, soutien, contexte

SCOLARITE

école, écriture, échec, réussite, obligatoire, expérience, collectivité, éducation, enseignement, référence

ERRANCE

itinérance, perdu, mouvement, alcool, drogue, délinquance, spirale, survie, illusion, espoir

ACTIVITES

consommation, centre de loisir, différence, ludique, outil, séduction, sens, animation

CHANTIERS

travail, expérience, argent, support, savoir faire, outil, lien, échange

FAMILLES

enfants, habitants, institution, éduquer, autoriser, autorité, partenaire, lien, filiation, repère

PARTENARIAT

réseaux, diversité, complexe, local, différence, incontournable, choix, pertinence, vitrine, orientation