

Université de Montréal
et
Université Paris Diderot (Paris 7)

**Modèle de structuration des relations lexicales fondé sur le formalisme
des fonctions lexicales**

par
Anne-Laure Jousse

Département de linguistique et traduction
Faculté des Arts et des Sciences
Université de Montréal
et
UFR de linguistique
Université Paris Diderot (Paris 7)

Thèse de doctorat effectuée en cotutelle et présentée à la
Faculté des études supérieures de l'Université de Montréal
et à
l'Université Paris Diderot (Paris 7)
en vue de l'obtention des grades de Philosophiæ Doctor (Ph.D.) et
de docteur de l'Université Paris Diderot (Paris 7)
en linguistique

Avril 2010

© Anne-Laure Jousse, 2010.

Université de Montréal
Faculté des études supérieures
et
Université Paris Diderot (Paris 7)
UFR de linguistique

Cette thèse intitulée :

**Modèle de structuration des relations lexicales fondé sur le formalisme
des fonctions lexicales**

présentée par:

Anne-Laure Jousse

a été évaluée par un jury composé des personnes suivantes:

Lyne DA SYLVA	présidente-rapporteure
Alain POLGUÈRE	directeur de recherche (Université de Montréal)
Sylvain KAHANE	directeur de recherche (Université Paris Diderot (Paris 7))
Margarita ALONSO RAMOS	examinatrice externe
Marie-Claude L'HOMME	membre du jury
Laurence DANLOS	membre du jury
Lorna HEATON	représentante du doyen de la FES

Résumé

Cette thèse porte sur l'élaboration d'un modèle de structuration des relations lexicales, fondé sur les fonctions lexicales de la Théorie Sens-Texte [Mel'čuk, 1997]. Les relations lexicales considérées sont les dérivations sémantiques et les collocations telles qu'elles sont définies dans le cadre de la Lexicologie Explicative et Combinatoire [Mel'čuk et al., 1995]. En partant du constat que ces relations lexicales ne sont pas décrites ni présentées de façon satisfaisante dans les bases de données lexicales, nous posons la nécessité d'en créer un modèle de structuration.

Nous justifions l'intérêt de créer un système de fonctions lexicales puis détaillons les quatre perspectives du système que nous avons mises au point : une perspective sémantique, une perspective axée sur la combinatoire des éléments d'une relation lexicale, une perspective centrée sur leurs parties du discours, ainsi qu'une perspective mettant en avant l'élément sur lequel se focalise la relation. Le système intègre l'ensemble des fonctions lexicales, y compris les fonctions lexicales non standard, dont nous proposons une normalisation de l'encodage.

Le système a été implémenté dans la base de données lexicale du DiCo. Nous présentons trois applications dans lesquelles il peut être exploité. Premièrement, il est possible d'en dériver des interfaces de consultation pour les bases de données lexicales de type DiCo. Le système peut également être directement consulté en tant qu'assistant à l'encodage des relations lexicales. Enfin, il sert de référence pour effectuer un certain nombre de calculs sur les informations lexicographiques, qui pourront, par la suite, être implémentés pour automatiser la rédaction de certains champs de fiches lexicographiques.

Mots clés fonctions lexicales, système de fonctions lexicales, fonctions lexicales non standard, relations lexicales, dérivations sémantiques, collocations, structuration des bases de données lexicales.

Abstract

This thesis proposes a model for structuring lexical relations, based on the concept of lexical functions (LFs) proposed in Meaning-Text Theory [Mel'čuk, 1997]. The lexical relations taken into account include semantic derivations and collocations as defined within this theoretical framework, known as Explanatory and Combinatorial Lexicology [Mel'čuk et al., 1995]. Considering the assumption that lexical relations are neither encoded nor made available in lexical databases in an entirely satisfactory manner, we assume the necessity of designing a new model for structuring them.

First of all, we justify the relevance of devising a system of lexical functions rather than a simple classification. Next, we present the four perspectives developed in the system: a semantic perspective, a combinatorial one, another one targeting the parts of speech of the elements involved in a lexical relation, and, finally, a last one emphasizing which element of the relation is focused on. This system covers all LFs, even non-standard ones, for which we have proposed a normalization of the encoding.

Our system has already been implemented into the DiCo relational database. We propose three further applications that can be developed from it. First, it can be used to build browsing interfaces for lexical databases such as the DiCo. It can also be directly consulted as a tool to assist lexicographers in encoding lexical relations by means of lexical functions. Finally, it constitutes a reference to compute lexicographic information which will, in future work, be implemented in order to automatically fill in some fields within the entries in lexical databases.

Keywords: lexical functions, system of lexical functions, non-standard lexical functions, lexical relations, semantic derivations, collocations, structuration of lexical databases.

Table des matières

Résumé	iii
Abstract	iv
Table des matières	v
Liste des tableaux	x
Liste des figures	xi
Liste des abréviations	xiv
Symboles et conventions typographiques	xv
Dédicace	xvi
Remerciements	xvii
Chapitre 1 : Introduction	1
1.1 Problématique de la recherche	1
1.2 Objectifs de la thèse	5
1.3 Organisation de la thèse	7
Chapitre 2 : Ressources et modèles lexicaux relationnels	9
2.1 Introduction	9
2.1.1 Modélisation du lexique en linguistique et ses disciplines connexes	10
2.1.2 Propriétés des relations lexicales	15
2.1.3 Les différents types de ressources lexicales relationnelles .	23
2.2 Panorama des ressources lexicales relationnelles	25
2.2.1 Le DiCo	26
2.2.2 FrameNet	34

2.2.3	WordNet	39
2.2.4	MindNet	47
2.2.5	Autres ressources lexicales relationnelles	51
2.3	Deux modèles psycholinguistiques	57
2.3.1	Modèle computationnel de l'approximation sémantique par analogie [Duvignau, 2002]	57
2.3.2	Modèle fonctionnel du lexique mental [Aitchison, 2003]	62
2.4	Conclusion	65
Chapitre 3 : Les fonctions lexicales.: aspects.formels.et notionnels		68
3.1	Introduction	68
3.2	La notion de fonction lexicale	69
3.2.1	Nature et rôle des fonctions lexicales	69
3.2.2	Caractère graduel des fonctions lexicales	73
3.3	Les encodages de fonctions lexicales	78
3.3.1	L'encodage vulgarisé	78
3.3.2	L'encodage traditionnel	79
3.3.3	Discussion critique sur l'encodage traditionnel	91
3.3.4	L'encodage explicite et l'encodage algébrique	105
3.4	Conclusion	109
Chapitre 4 : Normalisation des fonctions lexicales non.standard.		111
4.1	Introduction	111
4.1.1	Enjeu d'une normalisation de l'encodage des FLNS	112
4.1.2	Présentation du corpus de travail	115
4.1.3	Organisation du chapitre	116
4.2	Travaux antérieurs	117
4.2.1	Approche empirique	118
4.2.2	Approche logique	119
4.2.3	Approche par champ sémantique	120
4.3	Propositions d'après l'observation des FLNS du DiCo	122

4.3.1	FLNS candidates au statut standard	122
4.3.2	FLNS normalisées	127
4.4	Recherche de régularités via la dérivation morphologique	131
4.4.1	Corrélation entre les fonctions lexicales et la dérivation morphologique	131
4.4.2	Recherche d’analogies entre liens lexicaux à partir de marqueurs morphologiques	137
4.5	Organisation des FLNS selon des attributs de description	142
4.5.1	Présentation	142
4.5.2	Organisation des FLNS selon des attributs de description	147
4.6	Conclusion	152
	Chapitre 5 : Élaboration d’un système de fonctions lexicales	155
5.1	Introduction	155
5.1.1	Regard critique sur les classifications des fonctions lexicales	156
5.1.2	Vers une organisation systémique des fonctions lexicales	167
5.2	Perspective sémantique sur le système de fonctions lexicales	171
5.2.1	Décomposition des fonctions lexicales en diverses composantes sémantiques	172
5.2.2	Organisation des composantes sémantiques dans des classes sémantiques	177
5.2.3	Bilan et illustration	187
5.3	Perspective combinatoire sur le système de fonctions lexicales	192
5.3.1	Objectif	192
5.3.2	Les relations syntaxiques modélisées par les fonctions lexicales	193
5.3.3	Établissement des différents patrons de dépendance syntaxique associés aux fonctions lexicales	195
5.3.4	Bilan et illustration	202
5.4	Perspective sur le système de fonctions lexicales orientée sur les parties du discours	205

5.4.1	Objectif	205
5.4.2	Les parties du discours dans l'approche Sens-Texte	206
5.4.3	Les différents couples de parties du discours associés aux liens lexicaux représentés par les fonctions lexicales	210
5.5	Perspective communicationnelle sur le système de fonctions lexicales	213
5.5.1	Objectif	213
5.5.2	Comment déterminer la cible communicationnelle d'une FL?	215
5.5.3	Structuration des FL selon leurs cibles communicationnelles	221
5.6	Conclusion	223
Chapitre 6 : Applications dérivées du système de fonctions lexicales		224
6.1	Introduction	224
6.2	Implémentation du système de fonctions lexicales	225
6.3	Assistant à l'encodage des liens lexicaux	227
6.3.1	Sélection des différents paramètres	227
6.3.2	Interrogation sous forme de requêtes SQL	229
6.4	Calculs de données lexicographiques pour l'ébauche automatique de fiches BLEC	231
6.4.1	Objectif	231
6.4.2	Présentation des différents champs	233
6.4.3	Parties du discours	235
6.4.4	Marques d'usage	237
6.4.5	Étiquettes sémantiques	240
6.4.6	Calculs pour l'encodage des relations lexicales	244
6.4.7	Reconstitution de fiches lexicographiques	249
6.5	Conclusion	255
Chapitre 7 : Conclusion générale.		256
7.1	Bilan	256

	ix
7.2 Perspectives futures	259
Bibliographie	262
Index	280
Annexe I : Liste des FL citées	xix
Annexe II : Applications réalisées avec les FL	xxix
II.1 Ressources lexicales « générales »	xxix
II.2 Ressources lexicales bilingues et multilingues	xxxii
II.3 Ressources lexicales spécialisées	xxxvi
II.4 Ressources lexicales à visée pédagogique et outils d'apprentissage du lexique	xlii
II.5 Applications des FL en Traitement Automatique des Langues .	xlvii
Annexe III : Codification des classes du système de FL	lv

Liste des tableaux

2.I	Comparaison des ressources lexicales relationnelles	56
2.II	Exemples de concepts d'action : 'DÉTÉRIORER' et 'RÉPARER'	61
4.I	FLNS liées aux lexies d'animaux avant et après normalisation .	121
4.II	Correspondance entre FL et dérivatèmes	135
5.I	Classes des couples de parties du discours associés aux liens lexicaux encodés par les FL	211
6.I	Distinction de deux lexies à l'aide des marques d'usage	240
6.II	Distinction de trois lexies à l'aide des marques d'usage	240
6.III	Paires de FL à inférence simple	245
6.IV	FL réfléchies	246
6.V	Paires de FL inverses	247

Liste des figures

2.1	Définition des unités lexicales HAPPY et SAD au moyen de primitifs sémantiques [Wierzbicka, 1996 : 180 et 215]	12
2.2	Description de l'unité lexicale CAR dans le lexique génératif [Pustejovsky, 1995]	13
2.3	Extrait du vocable GLOOMY dans WordNet	44
2.4	Graphe sémantique obtenu à partir de la définition de MOTORIST	47
2.5	Représentation du graphe de ÉCORCER	59
2.6	Unités lexicales verbales liées à ÉCORCER	59
2.7	FL exprimant les relations sémantiques '(se) détériorer' et 'réparer'	62
3.1	Correspondance entre FL et unités lexicales	72
3.2	Exemple de règle de paraphrasage	73
3.3	Comparaison de l'article LAF et de la fiche DiCo pour ABOIEMENT	80
3.4	Typologie de l'encodage traditionnel	92
3.5	Liste des FL présentées dans Zholkovsky et Mel'čuk [1970]	93
3.6	Fiches d'identité des FL S_1 , S_{loc} , $IncepReal_1$ et V_0	104
3.7	Illustration de l'encodage explicite	106
4.1	Classes sémantiques de FL associées aux lexies étiquetées animal	121
4.2	FL standard adjectivales sous l'angle d'attributs polarisés	143
4.3	Le cadre <ACCOUTREMENTS> et ses différents éléments de cadre	146
4.4	Les vingt éléments de cadre les plus fréquents dans FrameNet	147
4.5	FLNS adjectivales normalisées de la lexie CHAPEAU	151
4.6	FLNS adjectivales normalisées de la lexie WAGON	152
5.1	Classification de Steele et Meyer [1990]	163
5.2	Classification de Grimes [1990]	164
5.3	Classification de Alonso Ramos et Tutin [1996]	165
5.4	Fiche de la FL S_{loc} dans Alonso Ramos [1993 : 377]	169
5.5	Liste des classes sémantiques	188

5.6	Relations lexicales de BATEAU organisées par classes sémantiques	191
5.7	Niveaux de représentation de la Théorie Sens-Texte	193
5.8	Valence de Fact ₁	196
5.9	Liste des valences syntaxiques associées aux valeurs de FL . . .	203
5.10	Collocatifs de RANCUNE organisés par patrons de dépendance syntaxique	204
5.11	Correspondance entre les parties du discours profondes et les par- ties du discours de surface	208
5.12	Structures syntaxiques profonde et de surface de <i>Marie casse les pieds à Jean</i>	209
5.13	Relations lexicales de RANCUNE organisées par parties du dis- cours profondes des valeurs de FL	212
5.14	Liste des cibles communicationnelles visées par les FL	221
5.15	Relations lexicales de RANCUNE organisées en fonction des cibles communicationnelles	222
6.1	Système de FL représenté sous forme de table de données	226
6.2	Fiche DiCo de la lexie ANCRE	234
6.3	Extrait de la hiérarchie d'étiquettes sémantiques	242
6.4	Structures de traits constituées à partir des relations dont CRE- VER est la valeur	253
6.5	Fiches de trois lexies potentielles CREVER	254
II.1	Tableau de régime de BOURRASQUE#II.1 dans le DEC	xxx
II.2	Fonctions lexicales et exemples pour BOURRASQUE#II.1 dans le DEC	xxxii
II.3	Macrostructure du dictionnaire Papillon	xxxv
II.4	Article RÉVER des dictionnaires ELRA et FeM	xxxvi
III.1	Codes associés aux classes sémantiques	lvi
III.2	Codes associés aux classes de valences syntaxiques	lvii

III.3 Codes associés aux classes de couples de parties du discours des liens lexicaux encodés par les FL	lviii
III.4 Codes associés aux cibles communicationnelles des FL	lviii

Liste des abréviations

BD	Base de donnée
BLEC	Base de données Explicative et Combinatoire
BLF	<i>Base Lexicale du Français</i>
DAFA	<i>Dictionnaire d'Apprentissage du Français des Affaires</i>
DAFLES	<i>Dictionnaire d'Apprentissage du Français Langue Étrangère ou Seconde</i>
DEC	Dictionnaire Explicatif et Combinatoire
DECFC	<i>Dictionnaire Explicatif et Combinatoire du Français Contemporain</i>
DICE	<i>Diccionario de Colocaciones del Español</i>
FE	Frame Element
FL	Fonction Lexicale
FLNS	Fonction Lexicale Non Standard
FLSS	Fonction Lexicale Standard Simple
ILEC	<i>Introduction à la Lexicologie Explicative et Combinatoire</i>
LAF	<i>Lexique Actif du Français</i>
LEC	Lexicologie Explicative et Combinatoire
Lefff	<i>Lexique des Formes Fléchies du Français</i>
SQL	Structured Query Language
TAL	Traitement Automatique des Langues
TST	Théorie Sens-Texte

Symboles et conventions typographiques

LEXIE	Lexie ou unité lexicale, ex. : CASSER
「LOCUTION」	Locution, ex. : 「CASSER LES PIEDS」
VOCABLE	Vocable, ex. : CASSER
‘signifié’	Signifié d’une unité lexicale, ex. : ‘casser’
‘CONCEPT’	Nom de concept auquel renvoie une unité lexicale, ex. : ‘CASSER’
<CADRE>	Cadre sémantique dans FrameNet, ex. : <ACCOUTREMENTS>
ÉLÉMENT	Élément de cadre sémantique dans FrameNet, ex. : WEARER
Fonction Lexicale	Fonction lexicale, ex. : Magn
→ Glose	Glose de fonction lexicale, ex. : → Intense
~	Mot-clé dans les FL ou les régimes de valeurs, ex. : $Adv_1(crainte) = avec [\sim]$
ART	Article ou autre déterminant dans les régimes des valeurs de fonctions lexicales, ex. : $Real_1(bateau) = piloter [ART \sim]$

NB : Cette thèse suit le modèle typographique français concernant les espaces avant les ponctuations doubles.

À Frédéric et Marianne Osterrath

Remerciements

J'aimerais tout d'abord remercier mes deux directeurs de thèse, Alain Polguère et Sylvain Kahane, qui ont su me guider dans cette recherche tout en me laissant une très grande liberté de manœuvre. Je leur suis sincèrement reconnaissante pour leurs conseils, leurs encouragements, pour leur façon de travailler toujours stimulante, rigoureuse et chaleureuse et pour la grande confiance qu'ils m'ont accordée. Je remercie Lyne Da Sylva pour avoir accepté de présider le jury, les deux rapporteuses Marie-Claude L'Homme et Margarita Alonso Ramos, ainsi que Laurence Danlos pour leur participation en tant que membres du jury. Un grand merci à Lucie Barque, Marie-Pierre Jousse, Marie-Andrée Mauger, Frédéric Osterrath et Aurélie Picton, qui ont bien voulu relire des sections de ma thèse et m'éclairer de leurs remarques.

Je dois encore mentionner Myriam Bouveret, à qui je dois ma découverte coup de foudre des fonctions lexicales et Marie-Claude L'Homme, qui m'a offert mon baptême lexicographique il y a presque dix ans. Je ne voudrais pas non plus oublier Igor Mel'čuk et sa passion communicative de la linguistique : une fois qu'on a été contaminé par le Virus Explicatif et Combinatoire, il semble qu'il n'y ait pas de guérison possible.

En tant qu'étudiante en cotutelle, j'ai eu la chance d'être accueillie dans deux laboratoires : l'OLST et le Lattice. Je remercie les membres de l'OLST, Patrick Drouin, Marie-Claude L'Homme, Alain Polguère et tous les étudiants avec qui j'ai lynché autour de la table ronde. Mon année passée au Lattice a été très stimulante, je tiens également à remercier Laurence Danlos et les étudiants latticiens de l'époque pour la bonne ambiance et l'enthousiasme scientifique qui régnaient dans le laboratoire. Je suis également reconnaissante à la Faculté des Études Supérieures, au département de linguistique et traduction de l'Université de Montréal et à l'organisme

Egide pour toutes les bourses dont j'ai pu bénéficier.

Je dois encore remercier les étudiants, collègues et amis aux côtés de qui j'ai vécu toutes ces années, notamment Lucie Barque, Mathieu Boulanger, Ivan Carel et Fanny Lemetayer, Sara-Anne Leblanc, Carolina Navarette, Étienne Nguyen, Sylwia Ozdowska, Céline Raynal, Bénédicte et Jacques Steinlin et Ophélie Tremblay ; ma mère qui m'a rendu un grand service pendant le sprint final en s'occupant de Marianne et en m'allégeant de bien des tâches ; Éric Buist pour son aide précieuse en \LaTeX et François Lareau, qui a su décrypter pour moi certaines procédures cabalistiques liées à la cotutelle. Enfin, mes amis de France et de Montréal, mes frères et sœurs, mes parents, Frédéric et Marianne m'ont apporté soutien, encouragement et réconfort tout au long de ma thèse, je les en remercie infiniment.

Chapitre 1

Introduction

1.1 Problématique de la recherche

Cette thèse vise l'élaboration d'un modèle de structuration des relations lexicales implémentable dans une base de données lexicale. La problématique de la thèse s'inscrit dans le cadre des nouveaux défis posés par l'évolution de la lexicographie. Au cours du siècle dernier, celle-ci a subi une modification de ses méthodes et objectifs. Cerguilini (cité dans De Schryver [2003]) distingue trois grandes phases :

1. la lexicographie papier assistée par ordinateur ;
2. l'informatisation des dictionnaires papier ;
3. la confection directe de dictionnaires électroniques ou bases de données.

L'arrivée du support informatique a donné lieu à une véritable révolution de la lexicographie et ouvert la porte à de nombreuses réalisations, dont la constitution semi-automatisée de ressources lexicales, la correction systématique des articles de dictionnaire ou la création d'interfaces pour la recherche d'informations. Notons que la troisième phase qui vient d'être mentionnée se trouve elle-même en mutation vers une quatrième : les bases de données lexicales (telles que WordNet [Fellbaum, 1998], FrameNet [Fillmore et al., 2003], le DiCo [Polguère, 2000], etc.) sont maintenant conçues sous forme de graphes d'entités lexicales interconnectées dont les dictionnaires ne sont qu'une vue partielle. Ce mode de structuration invite à repenser en profondeur la structuration et la présentation des données lexicographiques. C'est précisément cet aspect qui retient notre attention dans cette thèse.

Parmi l'ensemble des données lexicographiques habituellement présentes dans les ressources lexicales, nous nous intéressons plus particulièrement aux relations lexicales. Nous entendons, par relations lexicales entre lexies (ou unités lexicales), les

dérivations sémantiques et les collocations telles qu'elles sont définies dans la Lexicologie Explicative et Combinatoire [Mel'čuk et al., 1995]. Les dérivés sémantiques d'une lexie sont définis comme des lexies gravitant autour d'elle au sein d'un même champ sémantique et partageant une composante sémantique forte avec elle. Les liens entre une lexie et ses dérivés sémantiques sont paradigmatiques. Ainsi, pour la lexie ABEILLE, on observe les dérivés sémantiques suivants :

ABEILLE → APICULTEUR, ESSAIM, RUCHE, etc.

Les collocations, quant à elles, unissent, sur le plan syntagmatique, une unité lexicale et son collocatif, sélectionné d'après certaines contraintes. Par exemple, pour la lexie ABEILLE, on relève les collocatifs verbaux suivants :

ABEILLE → BUTINER, PIQUER, POLLINISER, etc.

Les notions de dérivation sémantique et de collocation seront définies en profondeur dans le chapitre suivant à la section 2.2.1.

Dans un dictionnaire ou une base de données lexicale, la présentation des relations lexicales issues d'une lexie va être différente selon le « parcours » du lexique que l'on choisit de représenter. Par exemple, il est possible d'organiser les relations lexicales associées à ABEILLE en fonction des parties du discours (les noms : ESSAIM, RUCHE, etc. ; les verbes : BUTINER, PIQUER, ÉLEVER, etc. ; les adjectifs : OUVRIÈRE, etc.), en fonction des dépendances syntaxiques (les verbes dont la lexie ABEILLE est le sujet : BUTINER, POLLINISER, PIQUER ; ceux dont elle est l'objet : ÉLEVER, etc.), ou encore en fonction du sens véhiculé par la relation (les actions effectuées par une abeille : ESSAIMER, POLLINISER, BUTINER, etc. ; les productions des abeilles : MIEL, 「GELÉE ROYALE」, CIRE, ALVÉOLE, etc. ; les parties du corps d'une abeille : TÊTE, AILES, PATTES, DARD, etc.). Or, du fait de leur structuration, la plupart des bases de données ne permettent pas de tels modes de représentation, dont le potentiel est pourtant très riche en matière de recherche d'information.

Notre travail s'articule autour de la problématique suivante : comment peut-on modéliser les relations entre unités lexicales pour mettre en valeur des parcours cohérents et pertinents à travers le lexique ? Quelle alternative aux simples listes de

relations lexicales peut-on proposer pour souligner leur organisation au sein d'un système relationnel ?

Le lexique est en interaction avec d'autres champs d'étude, dont la sémantique, la morphologie, la syntaxe, la phonologie ou encore l'analyse du discours. Il nous semble donc pertinent d'exploiter quelques-unes de ces facettes pour proposer une structuration des relations lexicales rendant compte de diverses perspectives sur le lexique. On recense quelques travaux de ce type réalisés manuellement ; c'est le cas du DAFLES [Selva et al., 2002], par exemple, qui propose plusieurs interfaces de consultation en fonction de critères syntaxiques, sémantiques, morphologiques, etc. Toutefois, nous pensons qu'un tel objectif à l'échelle du lexique d'une langue ne peut se faire sans une bonne formalisation des liens lexicaux. En effet, la formalisation permet, à partir d'un échantillon de relations lexicales, de traiter les relations de façon systématique à l'échelle du lexique d'une langue voire d'un ensemble de langues. La formalisation s'avère également fondamentale dans une optique d'implémentation pour la manipulation des données et la réalisation de nombreux traitements automatiques. Le formalisme de description utilisé dans la Lexicologie Explicative et Combinatoire (= LEC), les fonctions lexicales, nous semble pour cela tout adapté. Celles-ci se présentent comme des fonctions mathématiques modélisant des relations lexico-sémantiques particulières entre une unité lexicale, appelée mot-clé, et un ensemble d'unités lexicales, appelées valeurs. Par exemple, l'antonymie est représentée par la fonction *Anti* (ex : *Anti(joyeux) = triste, maussade*) ; la relation entre un nom et le verbe décrivant son utilisation ou sa réalisation typique est modélisée par les fonctions *Real_i*, *Fact_i* et *Labreal_{ij}*¹ ; par exemple :

¹Nous devons apporter quelques précisions concernant les conventions d'écriture utilisées pour les fonctions lexicales et les unités lexicales. Les fonctions lexicales sont notées à l'aide de la fonte suivante : *S_{loc}*. Lorsqu'elles ne sont pas expliquées dans le contexte de citation, nous les accompagnons de gloses explicatives, notées comme suit : → *Nom d'endroit typique*, par exemple, pour la fonction lexicale *S_{loc}*. Les valeurs de fonctions lexicales sont parfois accompagnées d'un régime qui précise la façon dont elles s'emploient : par exemple, *Real₁(bateau) = piloter [ART ~]*, signifie que le verbe *PILOTER*, employé dans ce contexte, doit être suivi d'un déterminant (= ART) et du mot-clé (= ~).

Les unités lexicales figurant dans les exemples de fonctions lexicales, qu'elles soient mots-clés ou valeurs, sont notées en italique. Hormis ce cas-ci, elles sont écrites en PETITES MAJUSCULES. La numérotation des lexies correspond à celle utilisée dans le DiCo. Comme nous sommes consciente que le lecteur ne peut accéder au DiCo pour vérifier de quelle acception il est question, nous citons,

$\text{Real}_1(\textit{bateau}) = \textit{naviguer}$ [sur ART \sim], $\textit{piloter}$ [ART \sim],
 $\text{Fact}_0(\textit{ciseaux}) = \textit{couper}$,
 $\text{Labreal}_{12}(\textit{chiffon}\#a) = \textit{essuyer}, \textit{frotter}, \textit{nettoyer}$ [N=Y avec ART \sim].

La motivation initiale des fonctions lexicales est la traduction automatique. Toutefois, il s’agit également de données essentielles pour la modélisation des relations lexicales d’une langue donnée. Les fonctions lexicales sont utilisées dans diverses applications mono- ou multilingues de traitement automatique des langues (TAL), telles que la génération de texte et le paraphrasage. Nous souhaitons, dans ce travail, leur donner une nouvelle perspective d’ordre lexicographique : la topographie des relations lexicales d’une langue.

Nous nous situons dans le cadre théorique de la LEC et utilisons le DiCo [Polguère, 2000] comme corpus de travail. Il s’agit d’une base de données lexicales du français développée par Igor Mel’čuk et Alain Polguère à l’Observatoire de Linguistique Sens-Texte de l’Université de Montréal. Notons que, par analogie avec la distinction entre un **modèle de dictionnaire de type explicatif et combinatoire** (un DEC) et les différents dictionnaires réalisés selon ce modèle, tels que le DECFC (*Dictionnaire Explicatif et Combinatoire du Français Contemporain* [Mel’čuk et al., 1984, 1988, 1992, 1999]), nous distinguons, d’une part, ce qu’on peut appeler une **base lexicale explicative et combinatoire** (BLEC, /belɛk/), c’est-à-dire un modèle de base de données lexicale élaboré selon les principes théoriques de la LEC et, d’autre part, les différentes bases de données réalisées selon ce modèle. Il en existe plusieurs exemples : le DiCo, le DiCoInfo [L’Homme, 2007], le DICE [Alonso Ramos, 2004], la base de données bilingue dérivée du Robert & Collins [Fontenelle, 1997b], etc. Contrairement à un DEC, une BLEC est élaborée avec et en vue d’un format électronique. Son objectif est d’offrir une grande liberté pour la structuration et la consultation des données. Dans cette thèse, bien que nous utili-

dans les cas ambigus, l’étiquette sémantique de la lexie, qui permet de caractériser brièvement son sens. Ainsi, dans l’exemple :

$\text{ImperLiqu}_1 \text{Oper}_1(\textit{cinéma}\#II \text{ ‘comportement’}) = \ll \textit{Arrête ton } \sim ! \gg$,

l’étiquette **comportement** permet d’identifier qu’il ne s’agit ni de l’activité artistique (= CINÉMA#I.1), ni de l’ensemble d’œuvres d’un réalisateur (= CINÉMA#I.2), ni encore de l’établissement de spectacle (= CINÉMA#I.3), mais du comportement (= CINÉMA#II).

sons comme corpus de référence les données du DiCo, notre travail vise n'importe quelle BLEC, quelle que soit la langue décrite et le type de nomenclature choisi (termes d'un domaine spécialisé, unités lexicales d'un champ sémantique particulier, etc.).

1.2 Objectifs de la thèse

Les **fonctions lexicales** (dorénavant FL) constituent un mode d'identification des relations lexicales dont la vocation est d'en systématiser le repérage et la description. Cependant, elles ne sont pas organisées en système, c'est-à-dire que leurs correspondances, les liens qui les unissent, ne sont pas tous systématiquement explicités. Les FL modélisent plusieurs aspects d'une relation lexicale : son sens ainsi que la configuration syntaxique du mot-clé avec la valeur. Par exemple, Real_1 dans la relation suivante : $\text{Real}_1(\textit{bracelet}) = \textit{porter}$ [ART \sim], indique que *porter* correspond à une utilisation typique que l'on fait d'un bracelet, qu'il s'agit d'un verbe prenant BRACELET pour complément d'objet et, pour sujet, le premier actant de BRACELET². Cependant, ces informations se trouvent encapsulées dans un même encodage, ce qui ne permet pas de les dissocier les unes des autres. Kahane et Polguère [2001] ont développé un encodage explicite des FL distinguant les structures sémantique et syntaxique associées aux FL. C'est dans la continuité de ces travaux que s'inscrit notre thèse.

L'objectif principal de la thèse est d'élaborer un système relationnel des FL pour mettre en valeur les différents types de liens qui les connectent. Ce système vise, d'une part, à souligner les rapprochements que l'on peut faire entre les FL et, d'autre part, à mettre en évidence leur valeur les unes par rapport aux autres selon un point de vue différentiel.

Ceci suppose un certain nombre de sous-objectifs plus particuliers :

²Une liste illustrée de toutes les FL citées dans la thèse est fournie à l'annexe I.

- Dans un premier temps, il nous faut remettre en question les différentes façons de considérer les relations lexicales et de les encoder. Il existe plusieurs types de FL distingués notamment en fonction de la fréquence des relations qu’elles encodent. Nous opérons des modifications sur l’encodage de certaines FL dites non standard de façon à les rendre davantage manipulables.
- Comme nous venons de le mentionner, les formules de FL encapsulent différents types d’information d’ordre sémantique et syntaxique. Nous cherchons à effectuer une « lecture » des FL afin de dégrouper ces caractéristiques et de les rendre explicites. Une fois celles-ci isolées, il est alors possible de faire émerger des rapprochements inédits entre les différentes FL et de créer un système représentant ces différentes formes de regroupement.
- Le système ainsi élaboré, nous pouvons l’exploiter pour réaliser plusieurs applications :
 - dans un premier temps, nous présentons un outil d’aide à l’encodage des relations lexicales directement dérivé du système ;
 - dans un second temps, nous déterminons dans quelle mesure il est possible, en s’appuyant sur le système de FL, d’inférer certaines généralités et de les implémenter en vue d’automatiser partiellement la rédaction de fiches lexicographiques du DiCo ou d’autres BLEC.

Nous pensons que cette thèse peut apporter une contribution dans plusieurs domaines. Tout d’abord, vis-à-vis de la LEC, le travail réalisé permettra d’organiser les FL en un modèle cohérent et hiérarchisé. Il offrira également la possibilité d’intégrer les FL non standard jusqu’ici écartées des traitements automatiques et de mettre à profit les FL pour servir de nouveaux objectifs. Enfin, il permettra d’ouvrir le DiCo (ou autre BLEC) à une utilisation plus étendue. En ce qui concerne la lexicographie computationnelle, notre projet propose une structuration novatrice des relations lexicales pouvant donner lieu à des possibilités de consultation plus adaptées aux besoins des utilisateurs. Les travaux réalisés trouveront également une utilité dans les domaines de la didactique des langues et des technologies de

l'information et de la communication au service de l'enseignement, dans le cadre desquels on pourra développer des outils d'aide à la rédaction.

1.3 Organisation de la thèse

Les différents aspects traités dans la thèse sont répartis en sept chapitres, le premier étant la présente introduction.

Le deuxième chapitre est destiné à poser le cadre conceptuel de la thèse en définissant les concepts phares de notre travail, notamment les relations lexicales prises en compte. Il vise également à situer notre cadre d'étude, à justifier la pertinence de notre démarche et à souligner les points forts des autres approches théoriques dont nous pourrions tirer profit.

Le troisième chapitre dresse un bilan des FL. Nous prenons soin de présenter, de façon indépendante, la notion de FL et les différentes formes d'encodage sous lesquelles elle se matérialise. Nous discutons de certains aspects théoriques et formels qui mettront en avant notre démarche et la contribution apportée par la thèse.

Le quatrième chapitre est destiné à la normalisation d'un ensemble particulier de FL, auxquelles on attribue un statut non standard. Après une discussion sur les différents statuts assignés aux FL, nous proposons deux méthodologies pour le repérage d'analogies entre relations lexicales, l'une destinée aux dérivations sémantiques, l'autre aux collocations. Nous reconsidérons l'encodage des fonctions lexicales non standard et en suggérons une normalisation permettant de mieux les identifier et de mieux les organiser dans les bases de données lexicales.

Dans le cinquième chapitre, nous proposons une vue d'ensemble des différentes classifications élaborées sur les FL pour justifier la nécessité d'en créer un système organisationnel. Nous présentons le système que nous avons réalisé en détaillant

chacune de ses perspectives.

Dans le sixième chapitre, nous détaillons la façon dont le système a été implémenté et présentons les applications que nous en avons dérivées : un outil d'aide à l'encodage des relations lexicales, ainsi qu'un algorithme pour le calcul d'informations lexicographiques destiné à l'automatisation partielle de la rédaction de fiches lexicographiques.

Enfin, le septième chapitre constitue la conclusion générale de la thèse. Nous proposons un bilan des aspects théoriques et applicatifs développés tout au long de ce projet en mettant en valeur la contribution apportée. Nous présentons également les perspectives futures que nous souhaitons développer dans le prolongement de la thèse.

Chapitre 2

Ressources et modèles lexicaux relationnels

2.1 Introduction

Cette thèse vise l'élaboration d'un modèle d'organisation des relations lexicales en vue d'une implémentation dans une base de données lexicale relationnelle. L'objectif de ce chapitre est, d'une part, de mettre en perspective notre travail vis-à-vis des autres ressources et modèles lexicaux relationnels et, d'autre part, de poser le cadre conceptuel de la thèse en définissant les concepts clés qui y sont traités.

En premier lieu, nous devons introduire la distinction entre un *modèle lexical relationnel* et une *ressource lexicale relationnelle*, distinction sur laquelle repose en partie l'organisation du chapitre. Un modèle relationnel est une représentation théorique du lexique visant à mettre en valeur les différentes relations existant entre les unités lexicales. Certains modèles relationnels donnent lieu à des ressources lexicales relationnelles. Ces dernières permettent de matérialiser et de valider les premiers à travers la description des propriétés relationnelles des unités lexicales d'une langue donnée.

Dans la suite de ce chapitre (section 2.2), nous présentons plusieurs ressources lexicales relationnelles. Parmi elles, certaines sont construites d'après des modèles théoriques particuliers, c'est le cas notamment du DiCo [Polguère, 2000], de FrameNet [Fillmore et al., 2003] ou encore de WordNet [Fellbaum, 1998]. D'autres sont élaborées sans arrière-plans théoriques spécifiques. Pour chacune des ressources, nous tâchons de déterminer :

1. les positions théoriques sous-jacentes, le cas échéant ;
2. les choix opérés vis-à-vis des relations lexicales prises en compte ;
3. les modes d'accès aux données proposés.

Dans la section 2.3, nous présentons deux modèles psycholinguistiques du lexique : celui de Aitchison [2003] et celui de Duvignau [2002]. Bien qu'ils n'aient pas encore fait l'objet d'une implémentation dans des ressources lexicales, ils nous semblent pertinents pour la problématique que nous traitons dans la thèse.

Pour bien comprendre l'enjeu de chacun de ces modèles et ressources et la façon dont ils se positionnent les uns par rapport aux autres, nous devons tout d'abord discuter certaines notions phares liées à la problématique de la thèse. Dans la suite de cette introduction, nous présentons donc les différentes façons dont le lexique est modélisé en linguistique et dans ses disciplines connexes, les propriétés liées aux relations considérées, ainsi que les types de ressources lexicales élaborées.

2.1.1 Modélisation du lexique en linguistique et ses disciplines connexes

Le lexique intéresse plusieurs champs disciplinaires. Evens et al. [1980] puis Murphy [2003] ont mis en évidence les motivations des disciplines telles que la philosophie, l'anthropologie, la linguistique, la psychologie et l'informatique pour l'étude des relations sémantiques. Nous nous intéressons particulièrement à la représentation du lexique et des relations entre unités lexicales dans les domaines de la linguistique, de la psycholinguistique et de l'intelligence artificielle. Ce qui peut rapprocher ces différentes disciplines, c'est que le lexique y est modélisé selon un même mode relationnel. Elles relèvent pourtant de traditions différentes : alors que l'intelligence artificielle hérite du modèle aristotélicien de la hiérarchisation des concepts, le paradigme relationnel en linguistique est davantage hérité de la linguistique structurale saussurienne, qui voit l'unité lexicale comme étant partie intégrante d'un système et n'étant définissable qu'au moyen de ses relations avec les autres éléments du système.

Ces disciplines se différencient également les unes des autres par leurs objectifs et méthodologies. En linguistique, on distingue couramment deux approches du lexique : une approche componentielle (également appelée décompositionnelle) et une approche relationnelle. Celles-ci ne sont pas véritablement en opposition, mais plutôt complémentaires. Les modèles componentiels (ou décompositionnels) définissent les unités lexicales en décomposant leur sens en unités sémantiques plus petites. Cette approche regroupe des écoles théoriques très variées : aussi bien les partisans de l'analyse componentielle comme Pottier [1992] et Rastier [1987], qui décomposent le sens des unités lexicales en sèmes, que les linguistes optant pour une décomposition en termes de primitifs sémantiques comme Wierzbicka [1996], ou encore ceux proposant des décompositions en termes d'expressions typées pouvant entrer dans des calculs sémantiques [Pustejovsky, 1995]. Dans toutes ces approches, le sens d'une unité lexicale est défini en termes de composantes isolables au moyen d'un ensemble prédéfini de symboles. Il peut s'agir de métalangues comme celles utilisées dans les dictionnaires classiques, ou d'ensembles déterminés de primitifs sémantiques, comme la Métalangue Sémantique Naturelle (= MSN) de Wierzbicka [1996]. Ainsi HAPPY et SAD sont définies au moyen de propositions écrites à l'aide de primitifs sémantiques, comme *someone, something, think, feel, like, person, etc.* (cf. figure 2.1, p. 12).

Le Lexique Génératif [Pustejovsky, 1995] décrit, quant à lui, une unité lexicale à travers trois structures (illustrées à l'aide de l'unité lexicale CAR (= 'voiture') dans la figure 2.2, p. 13) :

1. une structure argumentale (ARGSTR), qui déclare les différents actants ou participants de la situation dénotée par une unité lexicale et leur attribue des types (tel que **person**, **vehicule**, **thing**, etc.);
2. une structure événementielle (EVENSTR), qui représente les sous-événements de la situation dénotée ;
3. ainsi qu'une structure de qualias, dans laquelle sont mis en relation les arguments et événements des deux premières structures dans les rôles assignés à l'unité lexicale décrite. On note quatre rôles : **formal** correspond au type

HAPPY

X feels something

sometimes a person thinks something like this :

{something good happened to me ; I wanted this ;
I don't want anything more now}

because of this a person feels something good

X feels something like this

SAD

X feels something

sometimes a person thinks something like this :

{something bad happened ; if I didn't know that it hap-
pened,
I would say : I don't want it to happen ; I don't say this
now
because I know : I can't do anything}

because of this, this person feels something bad

X feels something like this

Figure 2.1 – Définition des unités lexicales HAPPY et SAD au moyen de primitifs sémantiques [Wierzbicka, 1996 : 180 et 215]

de l'unité décrite, **const** énumère les parties constitutives de l'entité dénotée, **telic** décrit l'utilisation prototypique qui en est faite et **agentive** l'événement causant l'existence de cette entité.

Les modèles décompositionnels n'ont pas pour vocation de décrire directement les liens sémantiques entre les unités lexicales. Même si la décomposition du sens des unités lexicales permet d'induire un certain nombre de relations lexicales comme la synonymie, l'antonymie ou encore l'hyponymie, l'étude des relations sémantiques entre lexies n'est globalement pas concernée par ce type de modèles.

À l'inverse, les modèles purement relationnels envisagent l'unité lexicale comme une unité atomique, dont le sens se définit par les relations lexicales qu'elle entretient avec les autres unités du lexique. Ces modèles ne fournissent pas de décomposition sémantique des unités lexicales, ils ne considèrent pas que le sens fait partie des unités lexicales, mais qu'il se situe parmi elles. On pourrait également qualifier ces

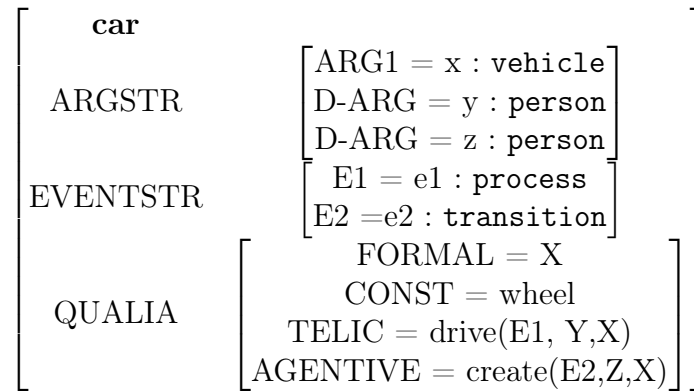


Figure 2.2 – Description de l’unité lexicale CAR dans le lexique génératif [Pustejovsky, 1995]

approches de holistes [Murphy, 2003 : 104], puisqu’elles considèrent le sens comme émergeant du système relationnel d’une langue, le sens d’une unité lexicale étant dépendant des autres unités lexicales avec lequel elle entre en relation.

Ces deux approches ne sont pas incompatibles. Il existe des approches mixtes qui reconnaissent le besoin de représenter le sens des unités lexicales de façon individuelle tout en maintenant la nécessité d’une représentation explicite des relations lexicales. C’est le cas, par exemple, du modèle de Cruse [1986], qui considère que les relations syntagmatiques et paradigmatisques contribuent au sens des unités lexicales, ou de la LEC [Mel’čuk et al., 1995]. Toutefois, même si la LEC se positionne à la fois comme un modèle décompositionnel, en fournissant des définitions analytiques détaillées en composantes sémantiques¹, et comme un modèle relationnel, nous ne traitons dans cette thèse que de cette deuxième facette.

En psycholinguistique, l’étude des relations sémantiques entre en jeu dans les travaux portant sur la mémoire à long terme et sur l’acquisition du langage. Notons que, dans cette discipline, ce ne sont plus seulement les relations entre unités lexicales qui sont concernées, mais plus largement, les relations entre concepts, ceux-ci étant directement reliés ou non aux unités lexicales. Les modèles élaborés visent

¹On trouve ces définitions dans les quatre volumes du DECFC [Mel’čuk et al., 1984, 1988, 1992, 1999] ainsi que dans la base de données de définitions formelles BDéf [Altman et Polguère, 2003].

à reproduire l'organisation du lexique mental, c'est-à-dire la façon dont le lexique est stocké dans le cerveau humain et non plus la façon dont il est employé dans les textes. Les observations sur les unités lexicales ne se font généralement pas à partir de l'étude de corpus, mais à l'aide de recherches du type stimulus-réponse ou encore en étudiant certains dysfonctionnements linguistiques comme l'aphasie. Les méthodes et objectifs se différencient donc très clairement de ceux de la linguistique.

Dans le domaine de l'intelligence artificielle, le lexique occupe une place centrale. Il est représenté dans des ontologies ou des réseaux sémantiques, utilisés pour la réalisation d'applications en TAL, telles que la désambiguïsation sémantique, l'élaboration de résumés automatiques ou la fouille de texte. D'après Sowa [1987], bien qu'il existe plusieurs types de réseaux sémantiques, ce qui leur est commun est une représentation graphique déclarative qui est utilisée pour représenter la connaissance. Dans ces modèles, la connaissance linguistique est représentée au même titre que la connaissance du monde. Pourtant, ces deux types de connaissance jouent des rôles tout à fait distincts. À travers l'analyse de plusieurs systèmes de traitement de la langue médicale, Zweigenbaum et al. [1996] ont démontré que plus l'analyse linguistique est fine, plus la base de connaissances est riche. Par ailleurs, les systèmes ne convertissant pas les réseaux linguistiques obtenus par une analyse de texte en format de représentation conceptuelle normalisée (du type graphe conceptuel) obtiennent des résultats nettement moins bons pour la recherche d'informations dans une base de connaissances.

Les grands projets de représentation des connaissances adoptent généralement un modèle ontologique et privilégient les relations de subsomption (relation mère-fille). Ce sont, par conséquent, les relations paradigmatiques (hyponymie, synonymie, méronymie, etc.) qui sont généralement représentées dans ces modèles. La plupart des travaux actuels sont centrés sur le développement de méthodes d'acquisition de connaissances lexicales à partir de textes écrits en langue naturelle. On peut citer, par exemple, MindNet [Richardson et al., 1998], qui cherche à extraire des informations relationnelles à partir de définitions et de textes libres.

Comme nous l'avons évoqué plus haut, bien que différentes dans leurs objectifs et méthodologies, ces trois disciplines envisagent le lexique selon un mode relationnel. Cependant, les relations dont il est question ne se situent pas sur le même plan. Nous examinons cette question dans la section suivante.

2.1.2 Propriétés des relations lexicales

D'un domaine scientifique à un autre, et au sein d'une même discipline, les recherches effectuées sur les relations lexicales reposent sur des principes si divers qu'il est parfois difficile d'établir des comparaisons sur les types de relations retenues. Nous tâchons donc d'explicitier un certain nombre de propriétés associées aux relations lexicales dans différentes approches du lexique, soit :

1. leur nature, c'est-à-dire si elles sont conceptuelles, sémantiques ou contextuelles ;
2. leur dimension : paradigmaticque ou syntagmaticque ;
3. leur caractère prototypique et idiomatique ;
4. leur degré de figement ;
5. les degrés de granularité dans le système qu'elles forment ;
6. leur identification et modélisation.

Tout au long de ce chapitre, nous verrons que ces propriétés ne convergent pas d'un modèle à un autre.

2.1.2.1 Nature des relations

Plusieurs termes sont employés pour désigner les relations entre unités lexicales (ou entre concepts) traitées dans les différentes approches : *liens lexicaux*, *relations sémantiques*, *relations lexicales*, *relations sémantico-lexicales*, *relations lexico-sémantiques*, *relations conceptuelles* ou encore *relations textuelles* ou *contextuelles*. Voici la terminologie que nous adopterons tout au long de la description des différents modèles. Nous parlerons de *relations lexicales* pour désigner des relations

récurrentes (habituellement dites *sémantiques*, *sémantico-lexicales* ou encore *lexico-sémantiques*) entre deux unités lexicales, généralement identifiées par leur sens. Par exemple, le lien entre les unités lexicales GRAND et PETIT est une occurrence de la relation lexicale d’antonymie.

Nous devons alors poser une différence entre une *relation lexicale*, d’une part, et les *liens lexicaux* qui en constituent des occurrences, d’autre part. Par exemple, la relation lexicale ‘nom de lieu où se trouve/se déroule \sim ’ est instanciée par une multitude de liens lexicaux dont : CRIME \rightarrow LIEU [du CRIME] ou encore ÉVIER \rightarrow CUISINE, VACHE \rightarrow ÉTABLE.

On parle de *relation conceptuelle* pour désigner une relation entre deux concepts. On dira, par exemple, que le concept ‘VENDEUR’ est lié au concept ‘ACHAT’. Ces deux concepts peuvent être exprimés au moyen de diverses unités lexicales dans un énoncé, y compris par des noms propres, par exemple : *Pierre a vendu une voiture ; Mon voisin a fait de grosses ventes cette semaine ; Ce vendeur sait vendre*, etc. Ce n’est donc plus l’unité lexicale qui est retenue mais le concept appartenant à une représentation mentale à laquelle il renvoie dans la situation de discours. Ces relations intéressent tout particulièrement les modèles cognitifs, dont l’objectif est de représenter le fonctionnement du lexique mental. On peut également observer des relations mixtes conceptuelles et lexicales dans WordNet, par exemple, où certains ensembles de synonymes (= synsets) peuvent être associés à des unités lexicales. Ainsi, les concepts ‘SEDUCER’ et ‘POLLUTER’ sont associés aux unités lexicales SEDUCER et POLLUTER et ont pour concept hyperonyme ‘BAD PERSON’, qui n’est pas lexicalisé en anglais.

On parle, par ailleurs, de *relations contextuelles* ou *textuelles* pour désigner les relations qu’une unité lexicale entretient avec les autres unités au sein d’un même texte ou d’un même paragraphe, c’est-à-dire les unités lexicales constituant son contexte d’emploi.

2.1.2.2 Dimension des relations lexicales

Les relations entre unités lexicales se situent sur deux dimensions. La première dimension est paradigmatique ou associative. Elle unit des unités lexicales ayant la propriété de se substituer l'une à l'autre. C'est le cas, par exemple, des relations de synonymie, d'antonymie ou encore d'hyponymie. Toutefois, les relations paradigmatiques ne se situent pas toutes *in absentia*. De nombreuses relations dites paradigmatiques s'expriment au moyen d'un syntagme combinant les deux unités lexicales. C'est le cas de la relation entre un nom (ou un verbe) d'action et le nom de son premier actant ; par exemple, le premier actant de CRIME peut être exprimé au moyen de la lexie ASSASSIN ou du syntagme *auteur du crime*.

La seconde dimension est syntagmatique. Elle se situe *in praesentia* sur l'axe linéaire du discours. Les expressions formées sur ce plan sont communément appelées collocations. Cependant, le terme *collocation* recouvre des notions différentes selon les théories linguistiques. La tradition anglo-saxonne du contextualisme, présentée dans les travaux de Firth et Halliday, considère les collocations davantage sous l'angle de la simple cooccurrence textuelle, si, toutefois, elle se révèle statistiquement significative. Les collocations jouent un rôle important dans l'interprétation linguistique dans la mesure où elles définissent le contexte d'emploi des unités lexicales et permettent la seule appréhension possible de leur sens dans ce contexte précis. Dans cette optique, c'est donc avant tout les notions de contexte et de fréquence d'occurrence qui définissent les collocations. Ces travaux ont donné naissance à toute une série de dictionnaires de collocations dont le BBI [Benson et al., 1997] ou l'*Oxford Collocations Dictionary for Students of English* [McIntosh et al., 2002].

D'autres approches, notamment la LEC, considèrent les collocations comme obéissant à des contraintes d'idiomaticité. Selon ce principe, si la collocation est bien une cooccurrence, elle ne l'est qu'en tant que cooccurrence restreinte, la seule fréquence n'étant pas un critère suffisant pour déterminer si un syntagme est collocationnel ou non. Nous reviendrons en détail sur la définition des collocations dans l'approche de la LEC à la section 2.2.1.

2.1.2.3 Prototypicité et idiomatité des liens lexicaux

Examinons à présent la question de la prototypicité des liens lexicaux. Des associations entre unités lexicales semblent être considérées comme « meilleures », c'est-à-dire plus prototypiques que d'autres. Cette hypothèse a été maintes fois testée par des cognitivistes au moyen de tests de type stimulus-réponse sur des locuteurs, pour certaines relations typiques comme l'antonymie, la synonymie ou encore l'hyponymie. Par exemple, les deux unités lexicales GRAND et PETIT fonctionnent comme une paire lexicale prototypique : l'adjectif GRAND appelant l'adjectif PETIT plus spontanément que l'adjectif COURT. Herrmann et al. [1986] ont tenté d'évaluer le degré de prototypicité en dressant une échelle de 1 à 5 pour des liens allant du moins prototypique au plus prototypique. Par exemple, alors que la paire IMMACULATE/FILTHY totalise une moyenne de 4.62, le couple MAXIMIZE/MINIMIZE est considéré antonyme au plus haut degré de prototypicité, soit 5.

Par ailleurs, certaines expressions sont porteuses d'une charge idiomatique plus importante que d'autres. Par exemple, dans le DiCo, la relation d'intensification appliquée au verbe REGRETTER#I.1 présente les collocatifs suivants : *beaucoup*, *énormément* et *infiniment*. Ces trois valeurs forment un crescendo du moins vers le plus idiomatique. La notion d'idomaticité est un continuum et il est bien souvent difficile de trancher entre un syntagme libre et un syntagme collocationnel. Aussi, si certaines expressions semblent totalement libres pour un locuteur natif d'une langue, on observe par le biais de révéléateurs, tels que les systèmes de traduction automatique, qu'elles sont pourtant idiomatiques et contraintes dans leur formation. Par exemple, comme le souligne Orliac [2004 : 2] l'expression transparente *to brush one's teeth* est une collocation en anglais. Pour désigner la même action, on ne peut employer une expression équivalente telle que **to wash one's teeth* (alors que les deux expressions sont possibles en français : *se brosser/laver les dents*). Selon F. Hausmann, les expressions idiomatiques constituent la majeure partie des textes en langues naturelles, à tel point qu'un locuteur natif n'a pas conscience de l'idomaticité enchâssée dans sa propre langue maternelle.

Même dans les dictionnaires traditionnels où beaucoup de locutions sont traitées de manière insuffisante, et où le regard de l'étranger n'intervient pas, les emplois idiomatiques prennent plus de place que les emplois non idiomatiques. Si tous les emplois idiomatiques étaient traités selon leur dignité lexicale et selon les besoins d'un étranger, leur supériorité serait écrasante. Car le lexique est beaucoup plus idiomatique qu'on ne pense. Sa caractéristique principale n'est pas sa compositionnalité mais son idiomatité. [Hausmann, 1997 : 280]

Le caractère à la fois omniprésent et inconscient de l'idomaticité d'une langue la rend donc difficile à saisir et à caractériser.

2.1.2.4 Degré de figement des liens lexicaux

La phraséologie d'une langue, c'est-à-dire l'ensemble de ses expressions idiomatiques, comporte un caractère graduel. Mel'čuk et al. [1995] et Mel'čuk [2008] ont présenté plusieurs typologies des expressions idiomatiques. Nous ne rentrons pas dans le détail de ces classements, mais attirons l'attention du lecteur sur la distinction fondamentale entre deux types d'expressions : les **locutions** et les **collocations**. Pour les définir, nous devons introduire les notions de compositionnalité et de choix. Un syntagme libre comme *soulever un rocher* est formé d'unités lexicales choisies librement et de façon indépendante. Son signifié est compositionnel, c'est-à-dire qu'il résulte de la composition des signifiés de ses éléments constituants : le signifié de *soulever un rocher* comprend celui de SOULEVER et celui de ROCHER. À l'inverse, les collocations et les locutions sont soumises à une contrainte dans la sélection d'au moins un de leurs constituants. Les locutions doivent être interprétées de façon globale : par exemple, la locution 「MONTER UN BATEAU」 est interprétée en un seul bloc comme DUPER. Elle constitue une unité lexicale à part entière au même titre que MONTER ou BATEAU. En d'autres termes, le signifié d'une locution n'est pas compositionnel ; il n'équivaut pas à la combinaison des signifiés de ses éléments constitutifs. Une autre caractéristique fondamentale d'une locution est qu'aucun des éléments individuels la constituant ne résulte d'un choix. Ainsi, dans 「CASSER SA PIPE」, la lexie CASSER n'est pas choisie par rapport à BRISER ou ABÎMER, tout comme PIPE ne l'est par opposition à CIGARETTE ou CIGARE. De

même, les éléments de la locution «CHEMIN DE FER» ne résultent pas d'un choix, FER n'étant pas sélectionné parmi un ensemble de lexies comme TERRE ou SABLE, ni CHEMIN par opposition à ROUTE ou VOIE.

Au contraire, une collocation est composée d'un élément librement choisi parmi un ensemble de lexies, qu'on appelle *base*, par exemple, AIMER, et d'un *collocatif* choisi en fonction des unités lexicales associées préférentiellement à cette base. On choisira ainsi, pour exprimer l'intensité auprès de AIMER, l'un des adverbes (ou locution adverbiale) suivants : «À LA FOLIE», PASSIONNÉMENT, ÉPERDUMENT plutôt que l'adverbe INTENSÉMENT, ce dernier étant moins naturellement ou spontanément utilisé par les locuteurs du français. Le comportement des signifiés des éléments d'une collocation est décrit par Kahane [2008 : 7] de la façon suivante :

Dans une collocation, la base possède son sens usuel, tandis que le collocatif est généralement dénaturé et exprime avant tout le sens de la fonction lexicale. Il en résulte que si l'on prend une collocation comme *une colère aveugle*, on voit qu'il est facile de commuter le collocatif avec de nombreux éléments collocationnels ou non (*une colère froide*, *une colère sourde*, *une colère homérique*, etc.) et en particulier avec une expression périphrastique : *une colère qui s'exprime sans contrôle* ; il est par contre difficile de commuter proprement la base (*aveugle* se combine surtout avec des noms de sentiments proches de COLÈRE comme HAINE), bien que la fonction lexicale puisse se combiner avec d'autres noms de sentiment et donner d'autres valeurs comme *une peur panique*.

Notons que, tout au long de la thèse, seules les collocations seront étudiées, les locutions n'entrant en ligne de compte que dans la mesure où elles participent à une collocation. Par exemple, dans le syntagme collocatif : *peler une «pomme de terre»*, la locution «POMME DE TERRE» joue le rôle de la base de la collocation et PELER, de son collocatif.

2.1.2.5 Granularité du système des relations lexicales

Pour aborder la question de la granularité du système des relations lexicales et de leur dénombrement, nous devons rappeler la différence précédemment mentionnée

entre une relation lexicale et les liens lexicaux qui en constituent des occurrences. Le choix de rassembler deux liens lexicaux sous une même relation lexicale ou sous deux relations différentes tient au degré de granularité que l'on choisit. Plus le degré de description est précis et plus le nombre de relations risque d'augmenter. Prenons le cas de la méronymie. Dans certains modèles, on la considère comme une simple relation « partie-tout » sans plus de précision. Au contraire, dans la base de données WordNet, trois types de méronymie sont distingués : « member meronymy » (ex. : *association* → *associate*), « substance meronymy » (ex. : *steel* → *iron*) et « part meronymy » (ex. : *table* → *leg*). Winston et al. [1987] poussent plus loin encore la granularité en proposant six types de relations méronymiques, soit : « component-integral object » (ex. : *pedal-bike*), « member-collection » (ex. : *ship-fleet*), « portion-mass » (ex. : *slice-pie*), « stuff-object » (ex. : *steel-car*), « feature-activity » (ex. : *paying-shopping*), et « place-area » (ex. : *Everglades-Florida*). La méronymie est donc représentée dans ces approches à travers trois niveaux de granularité différents, décrivant respectivement une, trois et six relations distinctes. Le nombre de relations lexicales identifiées dans les modèles est loin de réunir un consensus. Selon T. Fontenelle, cette question est intimement liée au cadre théorique à l'intérieur duquel on se place et à l'utilisation que l'on compte faire des relations identifiées :

Les écoles peuvent être en total désaccord sur ce nombre, d'autant plus qu'il peut varier en fonction de l'utilisation que l'on compte faire du modèle relationnel. Ainsi, Werner (1988) postule l'existence de trois relations de base, la modification, la taxonomie et la séquence, pour exprimer n'importe quel type de connaissance lexicale. À l'opposé, Ahlswede & Evens (1988) [...] identifient et codent plus de 100 relations différentes rien que pour les adjectifs. On s'accorde généralement sur la nécessité de tenir compte, dans les lexiques à orientation sémantique, des quatre relations suivantes, que l'on peut qualifier de fondamentales : [synonymie, antonymie, hyperonymie, hyponymie]. [Fontenelle, 1997a : 67]

Recenser les différentes relations lexicales d'une langue consiste à établir des analogies entre paires d'unités lexicales. Nous traitons en détail de cette problématique

au cours du chapitre 4, dans lequel nous cherchons à explorer deux méthodologies permettant d'étudier les mécanismes de mise en valeur d'analogies entre relations lexicales.

2.1.2.6 Identification et modélisation des relations lexicales

Selon les modèles du lexique et les ressources qui en sont issues, les relations représentées ne sont pas toujours identifiées de la même façon, voire ne sont pas identifiées du tout. Le *Redes*, dictionnaire des collocations de l'espagnol [Bosque, 2004], fournit, pour chaque unité lexicale décrite, un grand nombre de collocations sans les identifier, ni d'un point de vue sémantique, ni sur le plan syntaxique (c'est-à-dire en indiquant la structure syntaxique impliquant le mot-clé et le collocatif donné). La plupart des dictionnaires de collocations sont construits de cette façon. D'autres modèles, comme WordNet, identifient simplement les relations sur le plan sémantique en les nommant : synonymie, antonymie, hyperonymie, etc. Certaines ressources proposent une classification selon des patrons de structures de dépendances syntaxiques : N + Adj, N+ V, etc., sans toutefois fournir de description sémantique. C'est le cas, par exemple, de la base de données lexicale Elexico [Klosa et al., 2006] ou de Antidote [Charest et al., 2007]. La LEC fournit, quant à elle, une modélisation des relations au moyen des fonctions lexicales. Celles-ci indiquent à la fois la partie du discours du collocatif, la configuration syntaxique dans laquelle il figure avec le mot-clé et une brève caractérisation sémantique de la relation. Par exemple, la fonction lexicale S_{loc} : $S_{loc}(abeille) = ruche$, indique que l'unité lexicale *ruche* est un nom et qu'elle signifie approximativement 'endroit en relation avec les abeilles'. Les fonctions lexicales sont représentées selon un encodage formel. Cette formalisation est motivée à des fins d'application en TAL et notamment en traduction automatique. La formalisation apparaît nécessaire lorsque l'on fait face à un grand ensemble de relations : elle permet d'apporter une forme de systématisation et de contrôle sur la bonne identification des relations. Elle permet également de

rendre compte de façon explicite du système que forment les relations lexicales.

Nous avons précisé, jusqu'à présent, les grands modes de modélisation des relations lexicales en linguistique et dans ses disciplines connexes. Par la suite, nous avons discuté un certain nombre de propriétés associées aux relations lexicales. Ceci nous conduit donc, à présent, à détailler les différentes ressources lexicales tirées des modèles du lexique et décrivant lesdites relations.

2.1.3 Les différents types de ressources lexicales relationnelles

Nous appelons *ressources lexicales relationnelles* les dictionnaires, bases de données lexicales, graphes ou réseaux mettant en valeur les connexions entre unités lexicales. Toutefois, ce n'est pas tant la forme d'une ressource que son contenu qui en détermine l'appellation relationnelle. Les dictionnaires papier mettant en valeur les relations lexicales, qu'il s'agisse des dictionnaires de collocations ou des dictionnaires analogiques, peuvent, eux aussi, être qualifiés de relationnels.

La distinction précédemment mentionnée entre les modèles décompositionnels et les modèles relationnels se matérialise à travers les ressources chargées de décrire le lexique. Alors que les dictionnaires classiques sont des modèles componentiels (= décompositionnels) du lexique, les ressources lexicales plus modernes, conçues sous la forme de graphes ou de bases de données relationnelles, de même que les thésaurus, sont considérés comme des modèles relationnels ou holistes. Comme le souligne Murphy [2003 : 21], les modèles de type dictionnaire papier traditionnel et ceux de type thésaurus se situent sur un continuum. À l'extrême componentialiste, on retrouve des dictionnaires ne contenant aucune information relationnelle, à l'autre extrême — l'extrême holiste —, des modèles de type thésaurus ne comportant aucune information de nature définitionnelle sur les unités lexicales, dont le sens est censé émerger des relations avec les autres unités du réseau lexical.

On peut distinguer plusieurs types de ressources, chacune d'entre elles étant le reflet de la position théorique qui a présidé à son élaboration.

Dans le domaine du TAL et de l'intelligence artificielle, le terme *ontologie* est beaucoup employé sans que sa définition ne soit toujours bien établie. *Grosso modo*, on peut le définir comme un inventaire des différents objets et procès caractérisant un domaine et des spécifications de certaines relations qu'ils entretiennent entre eux. Vossen [2003] regroupe sous le terme général d'*ontologie*, plusieurs types d'objets hérités de traditions différentes.

- Premièrement, les ontologies formelles, héritées de la tradition philosophique, définissent des classes sur la base des genres et espèces, et donnent lieu à une hiérarchie constituée de nœuds reliés entre eux par subsomption. On parle également de taxonomies ou de hiérarchies.
- Deuxièmement, Vossen distingue les réseaux et cadres conceptuels. Ceux-ci proviennent de la tradition des sciences cognitives et de l'intelligence artificielle, et ont pour objectif de structurer la connaissance. Chaque nœud représente un concept et est associé à un certain nombre de traits typiques, comme les parties constituantes, les fonctions ou les propriétés physiques liées au concept. En combinant des relations d'héritage, ces réseaux servent de systèmes de représentation de connaissances.

La tradition linguistique et la sémantique lexicale ont donné lieu à des ressources représentant non plus la connaissance mentale, mais le sens des unités lexicales.

Deux grandes approches sont à distinguer :

- l'approche par traits sémantiques, qui présente des unités lexicales associées à des traits sémantiques (ex. : /animés/, /inanimés/) permettant de prédire certains comportements ou structures syntaxiques ;
- l'approche par réseaux lexico-sémantiques, dans lesquels les unités lexicales sont définies au moyen des relations qu'elles entretiennent entre elles.

Enfin, en dernier lieu, on distingue les thésaurus employés dans les sciences de l'information. Un thésaurus est un ensemble organisé de termes représentatifs d'un domaine de connaissances. Le plus souvent, il présente pour chaque terme un en-

semble d'informations facilitant la recherche documentaire, des indications sur les conditions d'emploi d'un terme, ainsi que des termes génériques et spécifiques qui permettent d'élargir ou de restreindre le champ de la recherche.

Dans la section suivante, nous présentons plusieurs ressources relationnelles héritées de ces différentes traditions.

2.2 Panorama des ressources lexicales relationnelles

Le panorama que nous présentons à présent a pour objectif de mettre en perspective les principales ressources lexicales relationnelles actuellement disponibles. Les ressources que nous avons retenues nous semblent représentatives de ce qui est réalisé et exploité en lexicographie, en psycholinguistique et en intelligence artificielle. Elles sont élaborées d'après des motivations et méthodologies très diverses. Nous les avons principalement choisies en raison de leur forte pertinence en matière de représentation des informations lexicographiques qu'elles proposent. Nous nous attardons plus particulièrement sur quatre ressources : le DiCo, FrameNet, WordNet et MindNet.

La présentation du DiCo nous permet d'apporter des précisions concernant le projet plus global de la thèse. Nous y voyons l'occasion de présenter notre corpus de travail et de définir en détail les relations qui seront prises en compte dans le modèle que nous souhaitons élaborer, soit les dérivations sémantiques et les collocations.

Le projet FrameNet mérite également une attention particulière dans la mesure où il propose une perspective originale du traitement des unités lexicales et dépasse les approches traditionnelles, centrées sur la définition des sens ou signifiés. En tentant d'expliquer des phénomènes linguistiques par des représentations cognitives à l'aide de cadres conceptuels, il offre un point de vue à la fois linguistique et cognitif sur l'organisation des unités lexicales.

Il semble difficile de passer outre WordNet, qui exerce une hégémonie parmi les ressources lexicales et sert de référence pour la réalisation d'un grand nombre d'applications dans le domaine du TAL.

Nous avons également retenu MindNet en raison de l'originalité de la méthodologie

selon laquelle la ressource est élaborée et de la variété de relations qu'il propose. D'autres ressources ne sont que brièvement présentées. Nous en proposons une synthèse comparative dans le tableau 2.I (p. 55).

2.2.1 Le DiCo

Le DiCo est une base de données lexicale développée par Alain Polguère et Igor Mel'čuk à l'Observatoire de Linguistique Sens-Texte de l'Université de Montréal. Il est élaboré d'après les principes théoriques de la LEC [Mel'čuk et al., 1995]. Ce modèle lexicographique a donné lieu à de nombreuses ressources. Comme nous l'avons mentionné dans l'introduction, on en distingue deux types : les DEC, dictionnaires papier, et les BLEC, bases de données lexicales relationnelles. Ces dernières contiennent quasiment les mêmes données que les DEC, mis à part les définitions, et fournissent une interface avec des possibilités de requêtes plus variées. La plupart de ces ressources sont présentées en détail à l'annexe II. Dans la présente section, nous insistons plus particulièrement sur le DiCo. Notons cependant que les types de données présentées ici figurent dans la majorité des DEC et des BLEC. Le DiCo compte actuellement 538 vocables, soit 1125 unités lexicales (acceptions) et 26 093 liens lexicaux décrits. Sa nomenclature se focalise avant tout sur la description des lexies les plus courantes du français qui contrôlent un certain nombre de dérivations sémantiques et de collocations.

Les informations lexicales représentées dans le DiCo

Le DiCo fournit une caractérisation sémantique des unités lexicales au moyen de brèves formules appelées **étiquettes sémantiques** [Polguère, 2003]. Une étiquette sémantique correspond approximativement au genre prochain d'une définition analytique. Les lexies BAGARRE#a et BAGARRE#b, par exemple, ont les étiquettes sémantiques suivantes, permettant de les identifier et de les différencier rapidement :

BAGARRE#a : échange de coups ;

BAGARRE#b : confrontation verbale.

Le DiCo donne également des informations sur les actants sémantiques et syntaxiques des unités lexicales. Un **tableau de régime** [Miličević, 2009] permet de spécifier la correspondance entre les actants sémantiques et syntaxiques, et la façon dont les actants syntaxiques sont exprimés. La lexie BAGARRE#a, par exemple, compte trois actants sémantiques. Sa forme propositionnelle est la suivante :

~ ENTRE L'individu X ET L'individu Y À PROPOS DE OU POUR Z.

Deux régimes lui sont associés :

Rég. 1 :

X+Y = I = entre N et N, A-poss

Z = II = Prép-propos² N | Z est la cause de B. ; pour N | Z est la finalité de B.

Rég. 2 :

X = I = de N, A-poss

Y = II = avec N

Z = III = Prép-propos N | Z est la cause de B. ; pour N | Z est la finalité de B.

Ces régimes doivent être interprétés de la manière suivante :

– selon le premier régime, il est possible de dire :

bagarre entre N et N à cause de Z (si Z est la cause de la bagarre),

leur bagarre à cause de Z (si Z est la cause de la bagarre),

bagarre entre N et N pour Z (si Z est la finalité de la bagarre),

leur bagarre pour Z (si Z est la finalité de la bagarre) ;

– selon le second régime, les actants de BAGARRE#a peuvent être exprimés comme suit :

bagarre de N avec N à cause de Z (si Z est la cause de la bagarre),

sa bagarre avec N à cause de Z (si Z est la cause de la bagarre),

bagarre de N avec N pour Z (si Z est la finalité de la bagarre),

²Prép-propos désigne une expression de type prépositionnel exprimant *grosso modo* le sens 'à propos de', par exemple : *à propos de*, *au sujet de*, *concernant*, *portant sur*, *sur*, *sur la question de* ...

sa bagarre avec N pour Z (si Z est la finalité de la bagarre).

Nous insistons particulièrement sur les relations lexicales représentées dans le DiCo. Dans le cadre de la LEC, deux grands types de relation sont considérés : d'une part des liens paradigmatiques ou **dérivés sémantiques**, d'autre part, les expressions syntagmatiques, ou **collocations**.

Si la notion de dérivation morphologique est bien connue et définie comme une relation à la fois sémantique et formelle entre deux unités lexicales, relations du type *fumer* → *fumoir* (nom de lieu), *effondrer* → *effondrement* (nom d'action), *noir* → *noircir* (verbe causatif), etc., la notion de dérivation sémantique reste, en revanche, moins usitée en linguistique. Il s'agit d'une extension de la notion traditionnelle de dérivation morphologique, présentée notamment dans Polguère [2000] et Mel'čuk et Polguère [2007]. Cette notion est, toutefois, bien plus large que l'idée que l'on se fait de la dérivation : il s'agit, plus généralement, d'une relation entre deux unités lexicales, fondée sur une parenté de sens. Mel'čuk et Polguère [2007 : 18] donnent trois conditions nécessaires pour qu'une relation entre deux lexies L_1 et L_2 soit considérée comme une dérivation sémantique.

1. L_2 entretient une relation sémantique avec L_1 , c'est-à-dire que les définitions de L_1 et L_2 possèdent des composantes de sens communes. Dans le cas le plus typique, L_2 se définit en termes de L_1 . Par exemple, la lexie HACHE [= L_2] est définie à l'aide de COUPER [= L_1], car HACHE signifie 'artefact servant à couper...'.

Cette première condition fait référence aux définitions des unités lexicales. On pourrait lui reprocher un manque d'objectivité, dans la mesure où les définitions des unités lexicales ne sont pas données, mais construites, et bien souvent, précisément dans le but d'explicitier certains liens sémantiques entre unités lexicales. Toutefois, il faut entendre cette condition au sens large, c'est-à-dire que la lexie L_1 ne doit pas forcément apparaître dans la définition de la lexie de L_2 en tant que telle, mais que l'on doit en trouver un synonyme, une unité lexicale plus générique ou plus spécifique. Par exemple, si la lexie HACHE est définie comme un artefact servant à

couper — ce qui en fait un dérivé sémantique de COUPER —, elle est également considérée comme un dérivé sémantique des verbes TRANCHER, FENDRE, etc.

2. La relation sémantique entre L_2 et L_1 est récurrente dans la langue. Par exemple, la relation entre HACHE et COUPER, ‘artefact servant à...’, est récurrente en français : FRAPPER \rightarrow MARTEAU, OUVRIR \rightarrow CLÉ, FUMER \rightarrow PIPE, etc.
3. La relation entre L_1 et L_2 peut s’exprimer au moins marginalement de façon morphologique dans la langue décrite. Dans certaines langues, cette relation s’exprime de manière régulière par de la dérivation morphologique (ou de la polysémie régulière). Par exemple, pour la relation ‘artefact servant à...’, on trouve en français les dérivations suffixales suivantes : *bouch(-er) \rightarrow bouch+on*, *ras(-er) \rightarrow ras+oir*, *décapsul(-er) \rightarrow décapsul+eur*, etc.

D’après cette troisième condition, les dérivés sémantiques d’une lexie L peuvent donc être des dérivés morphologiques de L comme dans l’exemple (1), des lexies sans aucune parenté morphologique avec L comme en (2), ainsi que des lexies en relation de co-polysémie avec L : les lexies MOUTON#I.b ‘viande’ et MOUTON#I.c ‘fourrure ou cuir’, en relation de co-polysémie avec MOUTON#I.a ‘animal d’élevage ou espèce animale correspondante’, en sont sémantiquement dérivées selon les relations représentées par les fonctions lexicales dans les exemples (3) et (4) :

- (1) $S_0(\textit{aboyer}\#I) = \textit{aboiement}\#I$;
- (2) $S_{\textit{instr}}(\textit{chant}\#I.a) = \textit{micro}$;
- (3) Chair de \sim utilisée comme aliment(*mouton*#I.a) = *mouton*#I.b;
- (4) Fourrure/peau de \sim utilisée comme matériau(*mouton*#I.a) = *mouton*#I.c.

La notion de dérivation sémantique étant assez large, Mel’čuk et Polguère [2007 : 19] déterminent trois grandes familles pour mieux la circonscrire. Nous les résumons ci-dessous :

1. Une première famille regroupe les cas où les deux lexies concernées possèdent approximativement le même sens. Il s’agit des relations de synonymie exacte

(par exemple VÉLO et BICYCLETTE) et approximative (par exemple, DÉFENDRE et INTERDIRE), de conversion (X précède Y est presque équivalent à Y suit X), de changement de partie du discours (COUP est le nom d'action équivalent au verbe FRAPPER) ainsi que les termes génériques (COURS D'EAU est le terme générique pour RIVIÈRE).

2. La deuxième famille regroupe les relations d'antonymie (par exemple, entre INTERDIRE et AUTORISER).
3. La troisième famille regroupe les cas plus larges, où l'une des deux lexies désigne un élément de la situation dénotée par l'autre. Il peut s'agir d'un actant (NAGER \rightarrow NAGEUR), d'un circonstant (DORMIR \rightarrow LIT) ou d'une caractéristique d'un actant ou circonstant (COMPRENDRE \rightarrow COMPRÉHENSIBLE).

Les collocations, quant à elles, sont des expressions syntagmatiques semi-figées composées de deux unités lexicales. Nous avons exposé en détail plus haut dans le chapitre (p. 19) les caractéristiques d'une collocation et ce qui la différencie d'une locution selon le point de vue de la LEC. Nous en reproduisons la définition formelle citée dans Jousse et Polguère [2005 : 7] :

Une collocation est une expression $L_1 L_2$ (ou $L_2 L_1$), formée des lexies L_1 et L_2 , que le locuteur produit d'une façon particulière : il sélectionne L_1 librement d'après son sens ' L_1 ', alors qu'il sélectionne L_2 pour exprimer un sens ' s ' en fonction de L_1 . La collocation est donc une expression semi-idiomatique. Voici trois exemples de collocations du français :

1. *grosse* [= L_2] *tempête* [= L_1]
2. *dormir* [= L_1] *profondément* [= L_2]
3. *pleuvoir* [= L_1] *des cordes* [= L_2]

On appelle **base** de la collocation l'élément qui, sélectionné librement par le locuteur, retient son sens dans la collocation et la contrôle. Dans les exemples ci-dessus, les bases des collocations sont les éléments étiquetés L_1 . Les éléments étiquetés L_2 sont appelés **collocatifs** de la base (dans la collocation).

Selon Mel'čuk et Polguère [2006 : 71], les dérivations sémantiques et les collocations constituent des **liens orientés** : le dérivé sémantique et le collocatif sont

choisis respectivement en fonction de la lexie source de la dérivation sémantique et de la base de la collocation : MALADE se combine avec GRAVEMENT, BLESSÉ, avec GRIÈVEMENT, etc. Ces relations étant orientées d’une lexie de départ vers une lexie cible, elles pourront donc être modélisées à l’aide de fonctions lexicales. Les fonctions lexicales prennent en tant que mot-clé la source d’une dérivation sémantique ou la base d’une collocation et retournent respectivement un ensemble de dérivés sémantiques ou de collocatifs. Les fonctions lexicales sont matérialisées sous la forme d’un encodage formel permettant de les manipuler dans des traitements automatiques. Le chapitre suivant étant entièrement consacré à leur présentation, nous ne nous étendons pas plus longtemps sur la question.

Nous souhaitons apporter quelques précisions sur la distinction entre les relations paradigmatisques et les relations syntagmatiques. Nous avons introduit plus tôt dans le chapitre (p. 20) une distinction entre les relations lexicales et les liens lexicaux, ces derniers constituant des occurrences des premières. De cette façon, la relation lexicale d’hyponymie est instanciée à travers un grand ensemble de liens lexicaux, dont les six suivants :

Hypo(<i>bateau</i>) = <i>bac</i> ,	Hypo(<i>grade#1</i>) = <i>capitaine</i> ,
Hypo(<i>bateau</i>) = <i>barque</i> ,	Hypo(<i>grade#1</i>) = <i>commandant</i> ,
Hypo(<i>bateau</i>) = <i>chaloupe</i> ,	Hypo(<i>grade#1</i>) = <i>lieutenant</i> .

Il arrive que des relations lexicales paradigmatisques soient instanciées par des liens lexicaux syntagmatiques et inversement. Ainsi, les relations d’intensification et d’utilisation typique, prototypiquement syntagmatiques, par exemple :

Magn(*cloche*) = *grosse*,
 Magn(*admiration*) = \lceil *sans bornes* \rceil , \lceil *sans réserve* \rceil ,
 Real₁(*frein#I*) = *appuyer* [sur le \sim],

sont exprimées, dans certains cas, par des liens paradigmatisques :

Magn(*cloche*) = // *bourdon*,

$\text{Magn}(\textit{admiration}) = //\textit{adoration}\#a,$

$\text{Real}_1(\textit{frein}\#I) = //\textit{freiner}.$

Dans ces cas là, le sens du mot clé et celui de la fonction lexicale sont compris dans celui de la valeur : un bourdon est une grosse (= **Magn**) cloche, l'adoration est une grande (= **Magn**) admiration, freiner est l'utilisation typique (= **Real**₁) des freins. Dans la LEC, lorsqu'une valeur comprend à la fois le sens du mot-clé et celui de la fonction lexicale, on parle de fusion. Pour en rendre compte dans l'encodage des relations lexicales, on fait précéder la valeur d'une double barre oblique « // » (cf. exemples ci-dessus).

À l'inverse, il arrive bien souvent que certaines relations lexicales, typiquement paradigmatiques, par exemple :

$S_1(\textit{abeille}) = \textit{apiculteur} : \rightarrow \text{Nom du premier actant},$

$S_1(\textit{meurtre}) = \textit{meurtrier},$

$\text{Mult}(\textit{chèque}) = \textit{chéquier} : \rightarrow \text{Ensemble de},$

$\text{Mult}(\textit{concert}) = \textit{tournée},$

retournent des valeurs en cooccurrence avec les mots-clés, par exemple :

$S_1(\textit{abeille}) = \textit{éleveur} [\text{de } \sim\text{s}],$

$S_1(\textit{meurtre}) = \textit{auteur} [\text{de ART } \sim],$

$\text{Mult}(\textit{chèque}) = \textit{carnet} [\text{de } \sim\text{s}],$

$\text{Mult}(\textit{concert}) = \textit{série} [\text{de } \sim\text{s}].$

Cependant, ces deux phénomènes ne doivent pas remettre en cause la répartition profonde des relations lexicales selon les deux axes paradigmatiques et syntagmatiques. La relation entre CLOCHE et BOURDON peut être vue à la fois sous l'angle de la relation d'intensification, c'est-à-dire d'une relation syntagmatique, mais également sous l'angle de l'hyponymie, qui est une relation paradigmatique. Concernant les autres relations données en exemple, nous dirons que les valeurs *série* [*de concerts*], *auteur* [*d'un crime*], *carnet* [*de chèques*] etc. sont des synonymes respectivement de TOURNÉE, MEURTRIER et CHÉQUIER, entretenant avec le mot-clé une relation paradigmatique. Cependant, ces unités lexicales, contrairement à leurs synonymes, nécessitent la réalisation syntagmatique de leur premier actant. On ne

peut, en dehors d'un contexte où les unités lexicales MEURTRE ou CONCERT auraient déjà été exprimées, employer les unités lexicales AUTEUR ou SÉRIE seules. Ce phénomène tient donc à la propriété des valeurs de FL plutôt qu'aux relations elles-mêmes.

Structuration de la base de données DiCo et accès à l'information

Bien que le DiCo soit une base de données sous forme électronique, il possède une structure assez traditionnelle, inspirée des dictionnaires papier d'usage courant. Steinlin et al. [2004] ont développé, au laboratoire Lattice-Talana de Paris 7, une architecture beaucoup plus riche — le DiCobjet — permettant d'améliorer et d'affiner les différents modes d'accès aux données. Le DiCo se consulte actuellement au moyen de l'interface DiCouèbe³. Le DiCouèbe permet d'effectuer des requêtes prédéfinies sur une base de données SQL. Les résultats sont affichés dans des tables HTML. Il est possible d'effectuer les requêtes à l'aide d'une interface de requêtes ou directement en mode SQL et d'exporter les résultats. Un nouveau format de structuration du DiCo est en cours de développement. Il s'agit d'un système lexical dans lequel chaque entité est isolée et connectée à d'autres via un certain nombre de relations [Polguère, 2009].

Une des applications du système de FL que nous élaborons dans la thèse est la confection d'interfaces de consultation pour le DiCo (ou autre BLEC). Ces interfaces doivent permettre plusieurs accès répondant à diverses motivations qui poussent un utilisateur à consulter une base de données lexicale (recherche orthographique, onomasiologique, en fonction de structures syntaxiques spécifiques, de parties du discours, etc.). Nous aurons l'occasion de préciser quels types d'interfaces pourront être créées tout au long du chapitre 5.

³L'interface DiCouèbe peut être consultée à l'adresse suivante : <http://www.olst.umontreal.ca/dicouebe>. Pour une description détaillée du fonctionnement du DiCouèbe, nous invitons le lecteur à se reporter à la documentation [Jousse et Polguère, 2005], téléchargeable à partir de l'adresse <http://idefix.ling.umontreal.ca/dicouebe/DiCoDOC.pdf>.

2.2.2 FrameNet

FrameNet est un projet de lexicographie computationnelle développé à l'Université de Berkeley⁴. Le projet repose sur la théorie de la sémantique des cadres (= *Frame Semantics*) et représente des réseaux de sens auxquels les unités lexicales participent. Les grandes lignes théoriques remontent aux travaux de Fillmore de la fin des années 1970 [Fillmore, 1977, 1982]. Le principe central de la sémantique des cadres est que le sens des unités lexicales doit être décrit au moyen de cadres sémantiques. Ces cadres correspondent à des représentations schématiques des structures conceptuelles et des patrons de croyances, de pratiques, d'institution, d'images, etc., qui fournissent un fondement pour l'interaction compréhensible dans une communauté linguistique donnée.

Les informations dans FrameNet

Dans FrameNet, chaque unité lexicale décrite est associée à (ou évoque) un cadre (= *frame*). Les différents sens liés aux unités lexicales d'un même vocable correspondent aux cadres qu'elles évoquent. La polysémie est donc en partie traitée à travers l'attribution de cadres sémantiques distincts pour les unités lexicales d'un vocable polysémique. Par exemple, l'adjectif GUILTY est associé aux cadres <CONTRITION> et <GUILT OR INNOCENCE> [Barque, 2006], ce qui permet d'en isoler deux sens distincts. FrameNet ne contient pas d'informations phonétiques, morphologiques ni étymologiques. On n'y trouve pas d'informations explicites sur les collocations ni de données statistiques sur les occurrences d'emploi des unités lexicales. En revanche, la base contient des entrées lexicales individuelles pour les unités lexicales décrites, des descriptions de cadres et d'éléments de cadre, ainsi que des corpus annotés. Chaque entrée lexicale contient une définition extraite du *Concise Oxford Dictionary* ou élaborée par des lexicographes de FrameNet. En outre, chaque entrée indique la réalisation syntaxique des éléments de cadre, ainsi que les patrons de valence. Trois grands types de relations sont traitées :

⁴FrameNet est accessible à l'adresse suivante : <http://framenet.icsi.berkeley.edu/~framenet>

- des informations sur la valence sémantique et syntaxique ;
- des relations entre unités lexicales au sein d'un même cadre ;
- des relations entre cadres.

1. Valence

Les informations concernant la valence sémantique d'un mot sont exprimées au moyen d'éléments de cadre (= *Frame Elements* ou *FE* dans la terminologie de FrameNet). Les informations concernant sa valence syntaxique sont exprimées par les syntagmes impliqués et leurs fonctions grammaticales. FrameNet utilise les cadres et leurs éléments pour décrire les propriétés de combinatoire sémantique et syntaxique des prédicats. Par exemple, le verbe GIVE sera décrit au moyen du cadre qu'il évoque : le cadre <TRANSFER>, dont les éléments sont DONOR, THEME et RECIPIENT. GIVE peut être exprimé selon deux patrons syntaxiques différents comme l'illustrent les exemples ci-dessous :

1. *The teacher gave the student a book.*
2. *The teacher gave a book to the student.*

On déduit de ces phrases deux patrons de réalisation grammaticale :

1. Éléments : DONOR - RECIPIENT - THEME
 Syntagmes : NP — NP — NP
 Fonctions syntaxiques : Sujet - Complément - Complément d'objet
2. Éléments : DONOR — THEME — RECIPIENT
 Syntagmes : NP — NP — PP-to
 Fonctions syntaxiques : Sujet - Complément d'objet - Complément

2. Relations entre unités lexicales dans un même cadre

Les verbes GIVE et RECEIVE évoquent tous les deux le cadre <TRANSFER>. Ils diffèrent quant aux perspectives qu'ils imposent sur l'événement de transfert. On observe une différence entre les rôles thématiques et les éléments de cadre. Une

analyse selon les rôles sémantiques attribuerait deux ensembles de rôles distincts à TO GIVE et TO RECEIVE :

TO GIVE : Agent Theme Recipient

TO RECEIVE : Recipient Theme Source

Il est vrai que l'on peut déduire globalement, à partir des rôles sémantiques, des généralisations sur les fonctions syntaxiques. Par exemple, les rôles Recipient sont plus susceptibles d'être sujets que les rôles Source. Cependant, on occulte, d'une certaine manière, la relation de sens dont une analyse en termes de cadres est capable de rendre compte : dans le cadre <TRANSFER>, le DONOR peut être représenté par les rôles Agent et Source, exprimés respectivement par TO GIVE et TO RECEIVE.

3. Relations entre cadres

On distingue quatre types de relations entre les cadres : une relation d'héritage, une relation, dite de sous-cadre, une relation d'utilisation et une relation de renvoi.

1. Relation d'héritage (= *Frame inheritance*)

→ *Héritage complet* : si le cadre B hérite du cadre A, il doit y avoir un élément dans B correspondant à chaque élément de A. Les éléments de B peuvent avoir un nom différent de ceux de A, mais il doit exister un lien entre eux.

→ *Héritage monotone* : si un cadre parent ou un élément parent a un type sémantique, le type sémantique du cadre ou de l'élément enfant doit être le même ou un sous-type de celui du parent.

→ *Héritage multiple* : un cadre enfant peut avoir plusieurs parents.

2. **Relation de sous-cadre** (= *Sub-frame relation*) Beaucoup de cadres expriment des concepts qui ont des sous-parties naturelles et bien définies. Il s'agit notamment des événements complexes. Le cadre <PICNIC>, par exemple, comprend les sous-cadres : <PREPARING FOOD>, <GOING TO A SUITABLE SPOT>,

... Les sous-cadres contiennent des éléments liés aux éléments du cadre complexe et donc indirectement liés entre eux.

3. **La relation d'utilisation** (= *Uses relation*) La relation d'utilisation est moins strictement définie que la relation d'héritage. Si le cadre B utilise le cadre A, le cadre B n'a pas besoin d'avoir un élément correspondant à chaque élément du cadre A. Par exemple, le cadre <CONVERSATION> est clairement relié au cadre plus abstrait <COMMUNICATION>, mais le cadre <COMMUNICATION> contient un élément que le premier n'a pas : MESSAGE. Par conséquent, on dit que <CONVERSATION> n'hérite pas de <COMMUNICATION>, mais qu'il l'utilise.

4. **La relation 'voir aussi'** (= *'See also' relation*) Cette relation sert à expliquer la nature des différences entre les cadres. Il s'agit d'un pointeur d'un cadre vers d'autres cadres similaires.

Structure générale de la base de données FrameNet et modes d'accès aux données

La base de données lexicale contient actuellement environ 9 000 unités lexicales, dont plus de 6 100 sont complètement définies dans 625 cadres sémantiques (= *semantic frames* en anglais) et exemplifiées à l'aide de 135 000 phrases annotées. Elle est structurée en deux bases de données : une base de données lexicale contient des informations sur les cadres et les unités lexicales ; une base d'annotations contient les phrases annotées et les sous-corpus dont elles sont tirées. À chaque mot cible dans la phrase annotée sont associés des patrons de réalisation grammaticale.

Les données de FrameNet sont disponibles en ligne sous la forme de notices navigables, d'un graphe et d'une base de données dans laquelle on peut faire des recherches. On peut avoir accès à plusieurs utilitaires, énumérés ci-dessous.

- **Un index de cadres** présente une liste cliquable de tous les cadres décrits. À chacun est associée une définition, ses différents éléments, des exemples et une liste d’unités lexicales appartenant à ce cadre.
- **Un index d’unités lexicales** présente l’ensemble des unités lexicales. Chacune d’entre elles possède une fiche contenant le cadre qu’elle évoque, une définition, une liste des gouverneurs syntaxiques (ex. : pour CAKE : BAKE, EAT, DECORATE, etc.), ainsi que la liste des éléments du cadre et leurs réalisations syntaxiques.
- **Des textes annotés.** En plus du travail lexicographique, FrameNet a entrepris un projet d’annotation de textes pour démontrer dans quelle mesure le projet peut contribuer au balisage de l’information contenue dans les textes. L’annotation consiste à marquer les éléments des cadres évoqués par les unités lexicales dans chaque phrase.
- **FrameSQL** est un mode de recherche expert grâce auquel on peut effectuer des recherches spécialisées dans toute la base de données. Il est possible de faire des requêtes en combinant tous les champs possibles : cadres, unités lexicales, valence, élément de cadre, etc. Trois interfaces sont disponibles : un mode de recherche globale (FN2full), un mode de recherche orienté vers les cadres (FN2basic), ainsi qu’un mode de recherche orienté vers les unités lexicales (FN2abc).
- **FrameGrapher** est un outil de représentation graphique des cadres. Il représente, sous forme de schémas, les relations entre cadres (ex. : héritage, utilisation, etc.) dans la base de données.

FrameNet est utilisé pour la réalisation de nombreuses applications en TAL [Lönneker-Rodman et Baker, 2009]. On trouve des exploitations de FrameNet pour l’annotation de corpus et des travaux portant sur la compréhension automatique de textes ou encore la désambiguïsation automatique. Par ailleurs, le grand succès de Frame-

Net a donné lieu à la création d'un FrameNet espagnol, un FrameNet japonais, un FrameNet allemand ainsi qu'un « Cross-lingual FrameNet » : groupe de discussion informelle sur la question de l'inter-langue qui se réunit pour parler des questions d'annotation et de structuration des lexiques, ainsi que de l'intégration des données à travers des langues.

Les objectifs de FrameNet et les types de données qui y sont traitées sont très différents de ceux des BLEC. Les éléments de cadre ne peuvent être assimilés aux fonctions lexicales, tant en raison de leur nature conceptuelle qu'en raison des types de valeurs qui les instancient dans les textes. Nous devons toutefois mentionner les travaux de Alonso Ramos et al. [2008] ou encore Bouveret et Fillmore [2008], qui visent à intégrer les fonctions lexicales de la LEC pour identifier et décrire certaines relations syntagmatiques dans FrameNet. Dans notre thèse, nous nous inspirons des éléments de cadre de FrameNet afin d'identifier certains types d'attributs descriptifs liés aux unités lexicales et de proposer une normalisation de l'encodage de certaines fonctions lexicales. L'utilisation des éléments de cadre en parallèle avec les fonctions lexicales est présentée dans le chapitre 4 (section 4.5).

2.2.3 WordNet

WordNet est un réseau lexical développé au laboratoire de sciences cognitives de l'Université de Princeton sous la direction de George A. Miller⁵. Le projet repose sur des conceptions psycholinguistiques de la mémoire humaine. Une des raisons de son développement est de confronter des hypothèses psycholinguistiques non pas à quelques champs sémantiques, mais à l'ensemble du vocabulaire commun. L'hypothèse sous-jacente postule que si les lexicographes s'adaptent mieux à la façon dont l'être humain pense, il sera plus facile à un utilisateur de naviguer dans le dictionnaire. En travaillant sur des associations d'unités lexicales, les psycholinguistes ont découvert des propriétés synchroniques du lexique mental pouvant être employées pour un dictionnaire. D'après Miller [Fellbaum, 1998 : 7], WordNet est

⁵La base de données est accessible à l'adresse suivante : <http://wordnet.princeton.edu/>

une base de données à laquelle on peut accéder par des voies sémantiques : c'est un dictionnaire sémantique conçu comme un réseau, en partie en raison du fait que la représentation des unités lexicales et des concepts en un système inter-relié semble correspondre à la façon dont les locuteurs organisent leur lexique mental.

Structure générale de la base de données WordNet

WordNet est-il un dictionnaire ou un thésaurus ? Cette question a été maintes fois posée pour tenter de définir la nature de cette ressource. Selon Fellbaum [1998 : 7], WordNet n'est ni un dictionnaire traditionnel ni un thésaurus, mais une combinaison des deux types de ressources lexicales de référence. Sa forme ressemble à celle d'un thésaurus, étant donné que ses unités de description sont des synsets (= ensembles de synonymes) constitués de toutes les unités lexicales qui expriment un concept donné, et qu'il permet d'accéder à des mots par les concepts. Mais WordNet est plus qu'un thésaurus, dans la mesure où il lie les concepts par des relations explicites et étiquetées. Les relations lexicales et les relations sémantico-conceptuelles, existant respectivement entre les unités lexicales et entre les synsets, sont clairement distinguées. Certains concepts peuvent exister sans unités lexicales correspondantes. On trouve, en effet, des concepts non lexicalisés (ex. : « bad person »). La structure de WordNet révèle donc un inventaire conceptuel, qui est seulement en partie relié au lexique de l'anglais. En cela aussi il diffère d'un thésaurus qui ne représente, en règle générale, que des concepts lexicalisés. WordNet peut être assimilé à un dictionnaire dans le sens où il fournit des définitions et des exemples pour la plupart des synsets. Tout comme un dictionnaire, il contient également des informations sur les unités lexicales morphologiquement reliées.

Les relations dans WordNet

La relation sémantique de base est la synonymie : chaque unité lexicale est associée à un synset. Il s'agit d'ensembles d'unités lexicales considérées comme synonymes stricts ou très proches. Par exemple, {joke, gag, laugh, jest, jape} forment un synset et {laugh, laughter} en forment un autre. Les unités lexicales figurant dans un

synset ne sont pas forcément synonymes dans tous les contextes. Ainsi, si GREASE et LARD sont synonymes dans le contexte de la cuisine, ils ne le sont pas nécessairement hors de ce contexte particulier.

S'agit-il de liens entre unités lexicales ou entre concepts ? Comme nous l'avons mentionné précédemment, WordNet distingue les relations entre les unités lexicales des relations entre concepts. Evens [Fellbaum, 1998 : 5] a montré que les chercheurs aspirant à modéliser la mémoire et le lexique mental tendent à construire des réseaux sémantiques avec des relations sémantico-conceptuelles, tandis que ceux qui travaillent sur les aspects lexicaux se concentrent exclusivement sur les relations entre unités lexicales. Or, selon les conclusions d'Evens, qui reposent sur des expérimentations réalisées sur des locuteurs, les relations conceptuelles du lexique mental ne coïncident pas avec une organisation purement linguistique. WordNet propose des concepts non lexicalisés résultant de la structure relationnelle de la base, qui peut lier deux concepts par le biais d'un troisième non lexicalisé en anglais. Cette structuration reflète des trous lexicaux (= *lexical gaps*) de la langue. Cependant, ces « trous » ne doivent pas être considérés comme des artefacts structuraux, ils sont souvent lexicalisés dans d'autres langues et révèlent des structures conceptuelles différentes des structures lexicales.

WordNet ne se préoccupe pas des relations syntagmatiques. La description est centrée sur une sémantique des unités lexicales plus que sur une sémantique textuelle ou discursive. Par conséquent, les relations syntagmatiques ainsi que les propriétés de combinatoire des unités lexicales font défaut dans la base. Par ailleurs, aucune information d'ordre étymologique n'est fournie. Miller [id., 35] a observé que certaines relations de notre lexique mental comme la relation IS-NOT-A (ex. : « A whale is not a fish. ») ne sont pas présentes dans WordNet. Il n'existe pas non plus d'indications pour distinguer les noms propres des noms communs, les noms de masse des noms comptables, etc. Pour ce qui est des relations entre unités lexicales ou entre concepts, les rédacteurs de WordNet avouent avoir pris conscience d'un

manque de relations sémantiques au fur et à mesure que la base de données s'est élargie.

Organisation en fonction des parties du discours

D'après des études psycholinguistiques sur l'organisation mentale des relations selon les parties du discours, l'équipe de WordNet émet l'hypothèse que les noms, dans l'esprit humain, ne sont pas organisés comme les verbes, ni comme les adjectifs, ni comme les adverbes. S'appuyant sur ces postulats psycholinguistiques, WordNet choisit de diviser le lexique en quatre classes : noms, verbes, adjectifs et adverbes, formant quatre réseaux différents. Chaque classe a sa propre organisation et présente certains types de relations en adéquation avec la partie du discours concernée. Notons que dans le cadre de la LEC, nous ne considérons pas qu'une telle distinction soit justifiée, nous en exposons la raison un peu plus loin.

WordNet structure les noms sur la base de l'hyponymie. Vingt-cinq grandes classes organisent la hiérarchie de noms, par exemple, `act / activity`, `animal / fauna`, `artifact`, `attribute`, `body`, `cognition/knowledge` ... Les relations entre noms sont d'ordre hyperonymique, méronymique, antonymique et dérivationnel. On note que les formes dérivationnelles ne sont pas forcément morphologiquement apparentées à une unité lexicale donnée en requête, mais à n'importe quelle unité lexicale du synset. Par exemple, `ARROW` appartient au synset `{arrow, pointer}` et a pour dérivé relié le verbe `POINT`, morphologiquement lié au nom `POINTER`.

L'organisation des verbes dans WordNet répond aux expérimentations menées sur des locuteurs. Les relations de cause et de troponymie sont les relations les plus intuitives observées sur les sujets testés. On dit que `V1` est troponyme de `V2` si `V1` veut dire 'V2 d'une manière particulière'; par exemple, `POISON` est un troponyme de `KILL` puisque `POISON` peut être défini comme 'to kill in a certain way'. Les relations proposées pour les verbes sont donc la cause, la troponymie, l'implication (par exemple, `SNORE` implique `SLEEP`), les locutions verbales contenant le verbe en question (par exemple, `⌈KILL OFF⌋` pour `KILLV`), les unités lexicales morphologi-

quement dérivées (par exemple, $KILL_N$, $KILLABLE$, $KILLER$ et $KILLING$), ainsi que le typage des actants des verbes (par exemple, $KILL_V$: Somebody —s something / Somebody —s somebody / Something —s somebody).

Les adjectifs sont répartis en deux grandes classes : les adjectifs descriptifs et les adjectifs relationnels. C'est l'antonymie qui régit l'organisation des adjectifs descriptifs. Deux types d'antonymes sont distingués :

- Les antonymes directs : par exemple, $HEAVY$ et $LIGHT$ sont conceptuellement opposés et constituent aussi une paire lexicale.
- Les antonymes indirects : par exemple, $HEAVY$ et $AIRY$, sont conceptuellement opposés sans être pour autant considérés comme une paire lexicale.

WordNet organise les adjectifs en groupes de synsets associés par similarité sémantique à un adjectif focal, qui relie le groupe à un autre groupe contrasté du pôle opposé de l'attribut. Par exemple, $PONDEROUS$ est synonyme de $HEAVY$, qui est un antonyme direct de $LIGHT$. Par conséquent, on observe l'opposition conceptuelle $PONDEROUS/LIGHT$ par le biais de $HEAVY$. Outre les synonymes et les antonymes, il existe des pointeurs vers les attributs de noms susceptibles d'être modifiés par des adjectifs. Par exemple, l'adjectif $HEAVY$ est associé à l'attribut "weight", $TALL$ pointe vers les attributs : "height" et "stature". Enfin, on observe des formes dérivées : par exemple, $TALLNESS$ pour $TALL$.

Les adjectifs relationnels ne sont pas traités de la même manière en raison de leur différence de comportement : ils ne sont pas gradables et n'ont pas d'antonymes directs. Ils sont donc contenus à part dans la structure de la base de données. La relation proposée pour les adjectifs relationnels est la dérivation. Pour chaque adjectif relationnel, on indique le nom dont il est dérivé : par exemple, $PRESIDENTIAL$ est relié à $PRESIDENT$.

La plupart des adverbes sont des dérivés ou homomorphes d'adjectifs. Ceux-ci héritent des propriétés des adjectifs dont ils sont dérivés. Les adverbes ne sont pas organisés en arbre ou en groupes, mais sont liés aux sens adjectivaux dont ils sont dérivés par le marqueur $DERIVED-FROM$.

Il ne nous semble pas qu'une telle distinction entre les parties du discours soit justifiée, dans la mesure où l'on retrouve quasiment les mêmes types de relations pour les différentes parties du discours. Si la méronymie semble ne concerner que les noms, l'antonymie concerne toutes les parties du discours, tout comme les relations de dérivation morphologique. Par ailleurs, l'hyponymie et la troponymie nous semblent tout à fait similaires.

Les modes d'accès à l'information

Il existe une interface pour naviguer au sein de la base de données, permettant aux synsets et aux relations d'être édités au format texte. Le navigateur permet d'accéder aux données en deux étapes.

Premièrement, après avoir entré une forme, on en visualise les différents sens (cf. figure 2.3 avec le vocable GLOOMY). Les sens sont regroupés par parties du discours. Un aperçu indique également le nombre de sens de chaque partie du discours représenté dans les textes annotés.

Adjective

- S : (adj) glooming, **gloomy**, gloomful, sulky (depressingly dark) *"the gloomy forest"; "the glooming interior of an old inn"; "gloomful' is archaic"*
- S : (adj) **gloomy**, grim, blue, depressed, dispirited, down, downcast, downhearted, down in the mouth, low, low-spirited (filled with melancholy and despondency) *"gloomy at the thought of what he had to face"; "gloomy predictions"; "a gloomy silence"; "took a grim view of the economy"; "the darkening mood"; "lonely and blue in a strange city"; "depressed by the loss of his job"; "a dispirited and resigned expression on her face"; "downcast after his defeat"; "feeling discouraged and downhearted"*
- S : (adj) blue, dark, dingy, disconsolate, dismal, **gloomy**, grim, sorry, drab, dreary, dreary (causing dejection) *"a blue day"; "the dark days of the war"; "a week of rainy depressing weather"; "a disconsolate winter landscape"; "the first dismal dispiriting days of November"; "a dark gloomy day"; "grim rainy weather"*

Figure 2.3 – Extrait du vocable GLOOMY dans WordNet

Deuxièmement, on peut accéder aux informations associées à une unité lexicale sélectionnée. Les relations avec les autres ensembles sont alors visibles et sélectionnables.

Apports et limites

Le projet WordNet de Princeton connaît un succès triomphal dans le domaine du traitement automatique des langues. Il sert de ressource support pour de nombreux projets, aussi bien pour la création de concordances sémantiques [Landes et al., 1998] que pour l’annotation de corpus [Fellbaum et al., 1998] ou encore l’extraction de textes [Voorhees, 1998]. WordNet a donné lieu à la création d’EuroWordNet [Vossen, 1998], une base de données multilingue constituée de réseaux pour plusieurs langues européennes. Les différents réseaux sont structurés sur le modèle de WordNet, c’est-à-dire selon des ensembles de synonymes (ou synsets) reliés par quelques relations sémantiques de base. Chacun des réseaux représente un système interne propre à la langue décrite. Parallèlement à ces descriptions individuelles, les réseaux sont connectés à un index interlangue fondé sur WordNet, permettant d’établir des équivalences d’une langue à une autre. Notons que Warner et al. [2004] ont travaillé sur l’ajout des relations syntagmatiques dans le réseau espagnol d’EuroWordNet. Il existe également un autre projet multilingue fondé sur le projet de Princeton : MultiWordNet. Ce projet est développé à partir d’une base de données de l’italien, dont la structure est rigoureusement identique à celle de WordNet⁶.

Il existe depuis peu une version WordNet libre d’accès pour le français (= WOLF) [Sagot et Fišer, 2008]. Ce projet a été créé dans le but de combler les lacunes du WordNet français élaboré dans le cadre d’EuroWordNet, non libre d’accès et relativement peu développé. La structure de WOLF est très proche de celle du WordNet de Princeton. Certaines de ces données sont, par ailleurs, liées à d’autres ressources lexicales du français comme le *Lefff* [Sagot, 2010] et DicoSyn [Ploux et Victorri,

⁶MultiwordNet est accessible à l’adresse suivante : <http://multiwordnet.itc.it/english/home.php>.

1998] (cf. section 2.2.5).

Malgré l'hégémonie de WordNet, certains linguistes ont émis quelques réserves quant à sa pertinence linguistique. Il s'agit notamment des défenseurs d'une sémantique textuelle et différentielle. Notons que ces critiques s'adressent autant à WordNet que, de façon plus générale, à toutes les approches ontologiques. Pour Slodzian [2000], le choix du paradigme sémantico-cognitif conduit à considérer que le signe est linguistique et que le sens est conceptuel, ce qui entraîne nécessairement une vision réductrice de la langue. Elle formule plusieurs critiques à l'égard de WordNet. Premièrement, l'unité de base étant l'unité lexicale isolée et totalement décontextualisée, les propriétés syntagmatiques sont négligées. Deuxièmement, la structure hiérarchique renforce l'idée d'une linguistique uniquement centrée sur le signe et qui ignore la problématique du sens textuel. Or, elle pense que le sens ne peut être défini que par d'autres unités linguistiques et non pas par un concept. Ceci présuppose un usage en contexte. De son point de vue, seule une sémantique textuelle différentielle est capable d'atteindre ces objectifs. Rastier [2004 : 15] vient renforcer son point de vue dans une analyse des ontologies contre lesquelles il se positionne radicalement :

Les lexiques des langues ne sont pas structurés comme des ontologies. Les relations sémantiques sont en effet plus complexes et variables que ce que prévoient les constructeurs d'ontologies. Pour construire une « dé-ontologie », il faut restituer la diversité des discours et des genres, qui rendent illusoire une ontologie unique ; insister sur le problème de la diversité sémiotique des textes, les corrélations complexes entre contenu et expression, l'incidence constituante du contexte. C'est là une condition pour remplir des tâches de caractérisation, notamment en linguistique de corpus.

Du point de vue de la LEC, l'absence de relations syntagmatiques dans une ressource lexicographique de si grande envergure nous semble une importante lacune qui mériterait d'être comblée.

On note que ces critiques sont symptomatiques des différences de paradigmes par le biais desquels le lexique est traité. Les choix méthodologiques opérés dans une

approche particulière ne peuvent apparaître satisfaisants pour une autre approche s'ils vont à l'encontre de ses principes fondateurs.

2.2.4 MindNet

MindNet [Richardson et al., 1998] est un projet de représentation des connaissances, développé par le groupe de traitement automatique des langues de Microsoft. La méthodologie d'élaboration de MindNet, visant à constituer une base de connaissances de façon automatique, nous semble particulièrement originale. Un analyseur syntaxique à large couverture est utilisé pour construire des réseaux sémantiques à partir de dictionnaires de langue, d'encyclopédies et de textes libres. Les réseaux résultent d'un processus entièrement automatisé, qui prend du texte en entrée, le découpe en phrases, établit un graphe de dépendances sémantiques entre les unités lexicales de chaque phrase, regroupe ces graphes individuels en un grand graphe unique et, enfin, assigne un poids probable à chacun des sous-graphes en fonction de leur fréquence dans les corpus. Par exemple, à partir de la définition suivante de l'unité lexicale MOTORIST :

motorist : a person who drives, and usu. owns, a car

il est possible d'établir le graphe sémantique reproduit à la figure 2.4.

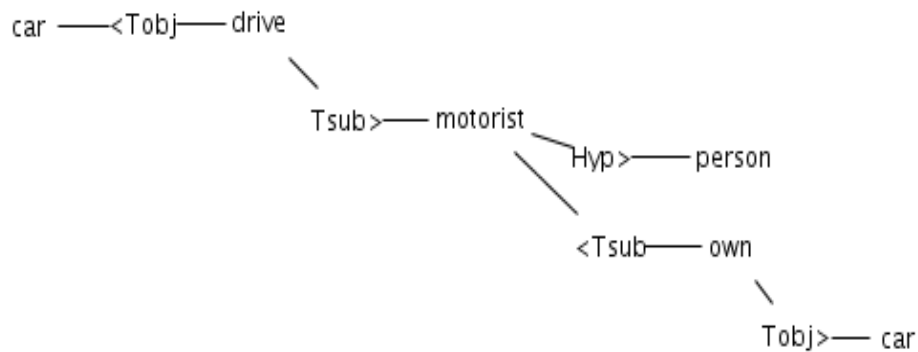


Figure 2.4 – Graphe sémantique obtenu à partir de la définition de MOTORIST

Richardson et al. [1998] décrivent le processus de la façon suivante : une fois les structures sémantiques créées, elles sont propagées à travers l'ensemble de la base de données MindNet et reliées à chaque unité lexicale qui la compose. Les relations, constituées d'un ou plusieurs liens connectés ensemble, représentent des chemins sémantico-relationnels (en anglais *semrel paths*) entre unités lexicales. Par exemple, le chemin entre CAR et PERSON est le suivant :

$$\text{car} \leftarrow \text{Tobj} \text{ --- drive --- Tsub} \rightarrow \text{motorist} \text{ --- Hyp} \rightarrow \text{person}$$

Il est possible d'obtenir plusieurs chemins à partir de diverses structures sémantiques. Par exemple, de l'unité lexicale CAR à l'unité lexicale DRIVER, on obtient l'ensemble de chemins suivants :

1. $\text{car} \leftarrow \text{Tobj} \leftarrow \text{driver}$
2. $\text{car} \leftarrow \text{Mod} \leftarrow \text{driver}$
3. $\text{car} \leftarrow \text{Mod} \leftarrow \text{road_hog} \rightarrow \text{Hyp} \rightarrow \text{driver}$
4. $\text{car} \leftarrow \text{Tobj} \leftarrow \text{crash} \rightarrow \text{Tsub} \rightarrow \text{driver}$
5. $\text{car} \leftarrow \text{Tobj} \leftarrow \text{control} \rightarrow \text{Tsub} \rightarrow \text{driver}$
6. $\text{car} \leftarrow \text{Tobj} \leftarrow \text{broadside} \rightarrow \text{Tsub} \rightarrow \text{driver}$
7. $\text{car} \leftarrow \text{Mod} \leftarrow \text{road_hog} \rightarrow \text{Join} \rightarrow \text{road_hog} \rightarrow \text{Hyp} \rightarrow \text{driver}$
8. $\text{car} \leftarrow \text{Locn(in)} \leftarrow \text{compartment} \rightarrow \text{Tsub} \rightarrow \text{driver}$
9. $\text{car} \leftarrow \text{Unsp(on)} \leftarrow \text{intention} \rightarrow \text{Possr} \rightarrow \text{driver}$
10. $\text{car} \rightarrow \text{Attrib} \rightarrow \text{intention} \rightarrow \text{Possr} \rightarrow \text{driver}$

Il est possible, à partir de sous-chemins de deux structures sémantiques différentes, de créer un chemin étendu. Par exemple, CAR et TRUCK ne sont reliées directement par une relation sémantique ou par un chemin sémantico-relationnel dans aucune structure sémantique. Cependant, si l'on joint les deux relations sémantiques suivantes :

$$\text{car} \text{ --- Hyp} \rightarrow \text{vehicle} \text{ et } \text{vehicle} \leftarrow \text{Hyp} \text{ --- truck,}$$

toutes deux tirées de différentes structures sémantiques, il est possible, partant de l'unité lexicale VEHICLE, d'obtenir le chemin suivant :

car — Hyp → vehicle ← Hyp — truck

On remarque que la polysémie n'est pas traitée dans MindNet : les acceptions ne sont pas différenciées explicitement. Par exemple, rien ne permet de déterminer quelle unité lexicale CAR est visée dans les relations exposées plus haut.

Des poids sont automatiquement assignés aux chemins sémantico-relationnels pour refléter leur saillance. Les relations obtenues sont énumérées ci-dessous ; notons que certaines étant encore en cours d'élaboration, nous ne présentons que les plus abouties.

- **Attributive modification** : cat → {domestic, wild, fierce, ... }
- **Causation** : virus → {disease, AIDS, cancer, ... }
- **Classifier** : species → {type, form, group, ... }
- **Goal** : school → {drive, hurry, tread, ... }
- **Hyperonymy** : bird → {robin, eagle, pheasant, ... }
- **Intensifier** : dark → {very, too, partially, totally, ... }
- **Location** : restaurant → {serve, table, meal, bill, ... }
- **Manner** : run → {quickly, slowly, suddenly, ... }
- **Means** : open → {key, pin, switch, force, ... }
- **Modifier** : quick → {mentally, relatively, unfairly, ... }
- **Modals** : call → {can, shall, go, need, use, ... }
- **Meronymy** : bird → {feather, wing, beak, leg, ... }
- **Possessor** : brush → {artist }
- **Proposition** : call → {return, engage, present, drink, ... }
- **Purpose** : read → {improve, find, decide, ... }
- **Result** : change → {flow, suit, ... }
- **Source** : borrow → {language, library, friend, ... }
- **Synonymy** : hill → {slope, pile, mound, ... }
- **Time/Duration** : war → {damage, fight, surrender, ... }
- **Typical Indirect Object** : pay → {person, worker, employee, ... }

- **Typical Object** : sing \rightarrow {song, note, music, praise, ...}
- **Typical Subject** : sing \rightarrow {people, singer, bird, chorus, ...}

Les relations dans MindNet sont de plusieurs ordres. On observe, en effet, un mélange entre, d'une part, des relations syntaxiques, par exemple : la modification attributive, l'objet typique ou encore le sujet typique, et, d'autre part, des relations sémantiques comme les relations d'intensification, de synonymie ou de but. Nous pouvons approximativement mettre certaines de ces relations en correspondance avec les fonctions lexicales :

Hyperonymy : Gener	Typical Indirect Object : Certains cas de S_2 ou S_3
Location : S_{loc}	
Possessor : Certains cas de S_1	Typical Object : Certains cas de S_2
Source : Certains cas de S_3	Typical Subject : Certains cas de S_1
Synonymy : Syn	

Cependant, les instances de relations obtenues par analyse de textes sont exclusivement retenues selon des critères de fréquence. Par conséquent, on retrouve des liens entre unités lexicales qui seraient radicalement exclus des ressources fondées sur la LEC. Certaines valeurs ne peuvent être considérées comme des véritables collocations (au sens de la LEC), par exemple : SUDDENLY n'est pas un collocatif de RUN. Par ailleurs, le lien entre READ et FIND, instance de la relation **purpose**, est loin d'être considéré comme un lien prototypique (cf. section 2.1.2.4). Les données de MindNet sont très hétérogènes et pourraient difficilement figurer telles quelles au sein d'une ressource lexicographique. Ce n'est toutefois pas l'objectif du projet MindNet ; il n'est donc pas étonnant que les projets de ce type mettent davantage l'accent sur la quantité de données couvertes que sur leur précision.

2.2.5 Autres ressources lexicales relationnelles

Les quatre ressources que nous venons de présenter ne sont pas les seules actuellement disponibles et utilisées. Nous mentionnons, dans cette section, d'autres ressources lexicales réalisées sur l'anglais, le français et l'allemand, que nous jugeons utiles pour notre recherche. Nous nous proposons de les décrire sommairement à l'aide d'un tableau comparatif visant à offrir une vue d'ensemble des principales ressources lexicales relationnelles. Nous envisageons un certain nombre de critères destinés à dégager les principales caractéristiques de ces ressources, soit :

1. les positions théoriques sous-jacentes ;
2. les langues couvertes par les ressources ;
3. le type de relations représentées : si elles sont conceptuelles, lexicales ou contextuelles ; paradigmatiques ou syntagmatiques ;
4. le nombre et l'identification des relations représentées ;
5. la structure des ressources.

Toutefois le tableau étant très succinct, nous devons faire au préalable une brève présentation de chacune des ressources.

VerbNet [Kipper et al., 2006] est l'un des lexiques de verbes les plus riches réalisés sur l'anglais. Il est développé à l'université du Colorado⁷. Le projet repose à la fois sur les classes de verbes de Levin [1993] et la sémantique des cadres. Il est mis en correspondance avec d'autres bases de données lexicales comme WordNet et FrameNet. VerbNet est organisé en classes de verbes, celles-ci étant des extensions des classes de verbes proposées par Levin. Chaque verbe est décrit au moyen de rôles thématiques, de restrictions sur ses arguments, et de cadres conceptuels (*frames* en anglais) constitués d'une description syntaxique et d'un ensemble de prédicats sémantiques. Les données de VerbNet sont directement utilisables pour effectuer des calculs formels.

⁷VerbNet est accessible à l'adresse suivante : <http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.

Antidote [Charest et al., 2007] est un logiciel développé par *Druide Informatique*. La version actuelle du logiciel (*Antidote HD*) comprend plus de dix dictionnaires (synonymes, familles morphologiques, analogies, cooccurrences, conjugaisons, etc.), un correcteur grammatical et orthographique, ainsi qu'un ensemble de guides linguistiques. Les données des dictionnaires de cooccurrences et d'analogies sont collectées de façon automatique, ce qui leur confère une grande quantité de relations lexicales. Toutefois, ces relations ne sont pas aussi précises que celles recensées dans les BLEC. Les liens syntagmatiques relevés, par exemple, sont avant tout des cooccurrents textuels. Nous avons décrit, dans [Jousse et Tremblay, 2006], la façon dont il peut être utilisé pour réaliser des applications pédagogiques d'enseignement du lexique au primaire.

Le **Dictionnaire du Français Usuel** (=DFU) [Picoche et Rolland, 2002] (disponible également en version électronique) est un dictionnaire du français à orientation pédagogique. Il est conçu d'après des travaux menés par Jacqueline Picoche sur les structures sémantiques et les stratégies d'enseignement du lexique [Picoche, 1977, 1986, 1993].

DicoSyn, dictionnaire électronique de synonymes, a été conçu par Ploux et Victorri [1998]. Il est développé actuellement au *Crisco* par Jean-Luc Manguin avec la collaboration de Jacques François et Bernard Victorri, et accessible à l'adresse suivante : <http://www.crisco.unicaen.fr/>. Il est conçu d'après la fusion de plusieurs dictionnaires de synonymes pour servir de support pratique à certains travaux de recherche sur la construction d'espaces sémantiques, dont les fondements théoriques ont été formulés par Victorri et Fuchs [1996]. Pour toute unité lexicale vedette, il offre un mode de représentation de sa polysémie, fondé sur le calcul de la distance sémantique entre tous ses synonymes dans un espace multidimensionnel. Les synonymes sont regroupés par cliques. L'interface d'utilisation permet également de visualiser l'espace sémantique d'une unité lexicale donnée par plans.

ELexico [Klosa et al., 2006] est une base de données lexicale de l’allemand, développée à l’institut de langue allemande de Mannheim. Il combine les traits d’un dictionnaire hypertexte et d’un système d’informations sur des données lexicales. Environ 500 unités lexicales sont pleinement décrites. Elexico se distingue des autres bases de données lexicales de l’allemand par la grande variété d’informations qu’elle fournit. Ces informations sont d’ordre morphologique, syntaxique, sémantique, relationnelle et contextuelle, et sont majoritairement extraites à partir de corpus.

Bien qu’elles se situent à l’écart des autres ressources lexicales mettant l’accent sur les relations sémantico-lexicales entre les unités d’une langue donnée, nous souhaitons mentionner les lexiques syntaxiques comme le lexique-grammaire [Gross, 1975], DicoValence [van den Eynde et Mertens, 2006] ou encore, le *Lefff* [Sagot, 2010] (qui est à la fois un lexique morphologique et syntaxique)⁸. Nous donnons l’exemple du **lexique-grammaire** pour illustrer ces types de ressources lexicales. Il s’agit du premier lexique formel réalisé, toutes langues confondues. Le lexique-grammaire repose sur le distributionnalisme de Harris [1964] et recense, à travers un ensemble de tables, les structures syntaxiques élémentaires du français. Des tables similaires ont été créées pour d’autres langues, dont l’italien, le portugais, le grec et le coréen. Les tables regroupent des verbes ainsi que des substantifs, adjectifs et adverbes prédicatifs sur la base de critères distributionnels et transformationnels. Une table correspond à une ou deux constructions syntaxiques particulières et rassemble toutes les unités lexicales qui entrent dans ces constructions. Par exemple, la table 1 des verbes contient tous les verbes qui admettent en plus d’un sujet, un complément infinitif. Cette ressource lexicale comporte une composante relationnelle dans la mesure où elle permet d’explicitier des liens entre les unités lexicales d’une langue donnée selon une perspective syntaxique. Rappelons que l’une des perspectives du système de FL que nous développons dans le cadre cette thèse consiste précisément à regrouper les FL partageant des propriétés de combinatoire

⁸Certaines données du lexique-grammaire et de DicoValence ont été intégrées récemment dans le *Lefff* [Danlos et Sagot, 2009].

syntaxique.

Nom	Langue(s)	Position théorique	Nature des relations	Relations identifiées	Structure
WordNet - Euro- WordNet - WOLF	anglais, néerlandais, italien, espagnol, allemand, français, tchèque, estonien	expérimentations psycholinguistiques	relations lexicales et relations conceptuelles — paradigmatiques (+ syntagmatiques dans Euro-WordNet espagnol)	synonymie, hyperonymie, antonymie, dérivés + relations encyclopédiques	quatre sous-réseaux pour les quatre parties du discours
FrameNet	anglais	sémantique des cadres	relations lexicales et relations conceptuelles	valence sémantique et syntaxique, relations entre cadres, relations entre unités lexicales et cadres	une BD lexicale + un corpus annoté aligné sur la BD
MindNet	anglais	représentation des connaissances	relations contextuelles et lexicales — paradigmatiques et syntagmatiques	Plus de 25 relations identifiées	graphe constitué de sous-graphes de dépendance sémantique
VerbNet	anglais	classes de verbe de Levin [1993], sémantique des cadres	relations lexicales et conceptuelles	hiérarchie de classes de verbes	base de données connectée à d'autres dont WordNet et FrameNet
DiCo	français	Lexicologie Explicative et Combinatoire	relations lexicales — paradigmatiques et syntagmatiques	plus de soixante relations formalisées au moyen des fonctions lexicales	BD SQL + interface Php

Nom	Langue(s)	Position théorique	Nature des relations	Relations identifiées	Structure
Antidote	français		relations lexicales — et contextuelles — et paradigmatiques et syntagmatiques	synonymie, collocations, relations paradigmatiques et dérivés morphologiques	logiciel sous forme de base de données
DFU	français	travaux de Picoche sur la didactique du lexique	relations lexicales	collocations, dérivés sémantiques	dictionnaire papier et informatisé
DicoSyn	français	construction dynamique du sens, espaces sémantiques	relations lexicales	cliques de synonymes, espace sémantique des unités lexicales	espace métrique multidimensionnel
Lexique-Grammaire	français + italien, portugais, grec, coréen	distributionnalisme de Z. Harris	combinatoire grammaticale des éléments prédictifs	similarités de propriétés de combinatoire grammaticale	ensembles de tables
ELexico	allemand		relations lexicales — et contextuelles — et paradigmatiques et syntagmatiques	cooccurents et unités lexicales thématiquement reliées, pas d'identification des relations (sauf synonymie et antonymie)	base de données relationnelle

Tableau 2.I: Comparaison des ressources lexicales relationnelles

2.3 Deux modèles psycholinguistiques

Dans cette section, nous présentons deux travaux qui nous semblent particulièrement pertinents pour la structuration des bases de données lexicales : le modèle de Duvignau [2002] et celui de Aitchison [2003]. Ces modèles ont en effet pour vocation de fournir une représentation du lexique en adéquation avec la façon dont le lexique mental des locuteurs est organisé. Les éléments proposés, plus ou moins complémentaires avec l’approche par fonctions lexicales, nous semblent tout à fait compatibles avec les interfaces de consultation de BLEC qui seront développées à partir de notre modèle de structuration de relations lexicales.

2.3.1 Modèle computationnel de l’approximation sémantique par analogie [Duvignau, 2002]

Dans sa thèse de doctorat, Duvignau [2002] a cherché à valider une hypothèse postulant une certaine organisation mentale des verbes. Elle souhaitait démontrer que la métaphore verbale reflète, du moins en partie, l’existence d’une organisation du lexique des verbes selon la relation lexico-sémantique d’hyperonymie/hyponymie et la similarité intra et inter domaine (rapprochant, par exemple, les verbes CABOS-SER dans le domaine de la mécanique, MEURTIR dans le domaine de la médecine et DÉLABRER dans le bâtiment). Elle nomme ces deux types de relations, des *relations de proximité sémantique par analogie* [Duvignau, 2002 : 138]. Pour tester cette hypothèse, deux corpus ont été constitués : un premier portant sur des métaphores présentes dans des textes de vulgarisation scientifique et un autre composé d’analyses de lexicalisations précoces verbales (c’est-à-dire au moyen de verbes) produites spontanément par des enfants entre deux et quatre ans. Ces dernières s’apparentent à des métaphores puisqu’elles présentent, elles aussi, des combinaisons inhabituelles de termes sur le plan sémantique. Ces différentes sources se rejoignent dans le sens où les deux types de locuteurs, adultes comme enfants, se situent dans une démarche explicative visant à se faire comprendre par leurs interlocuteurs. Associée à Gaume [Gaume et Duvignau, 2004], Duvignau cherche à créer un lexique métapho-

rique mental d'après un modèle computationnel de l'approximation sémantique par analogie. Leurs travaux combinent une approche psycholinguistique et linguistique à une approche mathématique et informatique. Leur hypothèse est la suivante :

Élaborer des dictionnaires électroniques en s'appuyant sur une théorie linguistique qui s'avère être en adéquation avec des processus d'acquisition précoce du lexique, leur confèrera un caractère ergonomique et devrait ainsi améliorer leur utilisabilité. [Gaume et Duvignau, 2004 : p. 158]

Partant de cette hypothèse, leur objectif est de développer une méthode mathématique susceptible de révéler les traces d'une structuration précoce du lexique mental. Le principe de la méthode est de tirer parti de la structure mathématique du réseau des renvois définitionnels entre verbes. Le modèle computationnel travaille à partir d'un graphe de dictionnaire et calcule des similarités entre ses sommets (les entrées du dictionnaire en question). La méthode choisie pour extraire le graphe du dictionnaire consiste à prendre pour sommets du graphe les entrées du dictionnaire et à admettre l'existence d'un arc de sommet A vers un sommet B, si et seulement si l'entrée B apparaît dans la définition de l'entrée A. Duvignau [2002] en présente une illustration autour du sommet ÉCORCER. À partir de la définition suivante :

ÉCORCER [ekóRse] v. tr. : Dépouiller de son écorce (un arbre).

Décortiquer, peler (le grain, les fruits)

et en appliquant la procédure que l'on vient d'exposer, on arrive au résultat exposé dans la figure 2.5.

En réitérant cette construction pour chacune des entrées du dictionnaire, on obtient le graphe du dictionnaire. Si l'on ne s'intéresse qu'aux seuls verbes apparaissant à l'infinitif dans les entrées verbales, on récupère l'ensemble de verbes représenté dans la figure 2.6.

Cette méthode permet donc d'obtenir des séries de verbes reliées par approximation sémantique. Gaume et Duvignau [2004 : 251] l'expliquent comme suit :

Cet algorithme met au jour les structures « topologico-sémantiques » des graphes de dictionnaires parmi lesquelles figurent la proxémie qui

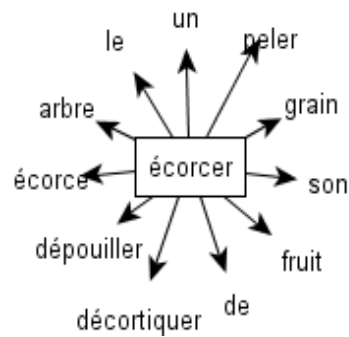


Figure 2.5 – Représentation du graphe de ÉCORCER

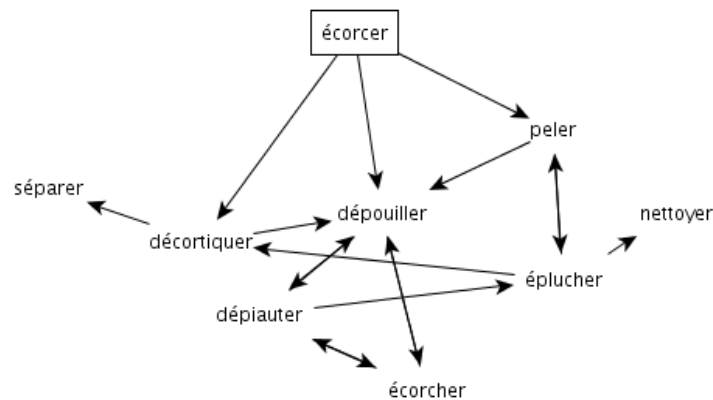


Figure 2.6 – Unités lexicales verbales liées à ÉCORCER

organise dans un continuum l’hyperonymie-hyponymie et la similarité intra et inter domaine.

À partir de ces données, leur objectif est de confectionner un dictionnaire reflétant ce phénomène.

Comparaison avec la LEC et les fonctions lexicales

Nous voudrions formuler certaines remarques par rapport à ces travaux. Il nous semble que les fonctions lexicales de la LEC sont capables de réaliser un travail tout à fait similaire. Les deux approches relèvent de paradigmes différents : alors que la LEC adopte une approche lexicologique, les expériences de Duvignau et Gaume se placent dans une perspective plus cognitive. D’après les exemples de

métaphores relevés, Duvignau émet trois hypothèses [Gaume et Duvignau, 2004 : p. 161] :

1. il existe des concepts d'action comme 'DÉTÉRIORER' ou 'RÉPARER' ;
2. ces concepts véhiculent leur intension (l'essentiel de leur signification) à travers des expressions linguistiquement variées ;
3. ces expressions linguistiques « réalisent » l'essentiel de la signification du concept dans des champs sémantiques différents, qui renvoient à des types d'objets spécifiques (papier, voiture, maison) et, parfois, à des domaines d'activité particuliers (mécanique, bâtiment).

Puis, elle répartit les verbes d'un même concept en termes de domaine [Gaume et Duvignau, 2004 : p. 161] :

Tous ces verbes (comme « rescotcher », « soigner », « recarrosser », « re-crêpir ») sont donc rattachés à un concept qui constitue pour chacun d'entre eux, un noyau de sens commun. De tels verbes constituent, de ce fait, des co-hyponymes, dont on peut distinguer deux types :

- les co-hyponymes intra-domaine : « fissurer », « lézarder », « se délabrer », sont des co-hyponymes intra-domaine car ils relèvent d'un même domaine d'activité, à savoir le bâtiment
- les co-hyponymes interdomaines : « fissurer » et « meurtrir » sont des cohyponymes inter-domaines car ils relèvent de domaines différents, respectivement le bâtiment et la médecine

Duvignau parle de concepts et d'organisation mentale. Selon elle, les mots relevés réalisent un concept et sont unis par une relation de co-hyponymie. La LEC, quant à elle, dans une perspective linguistique et lexicologique, parle non plus d'un concept, mais d'une relation lexicale observable entre plusieurs unités lexicales. Les relations sont orientées d'une lexie de base vers une lexie cible, qu'on appelle dérivé sémantique dans le cas des dérivations sémantiques, et collocatif pour les collocations. Dans le cas des exemples précédemment cités, les relations sont orientées d'une lexie de base vers un ensemble de collocatifs. Autrement dit, une même relation

lexicale est exprimée au moyen de syntagmes collocationnels, le collocatif variant en fonction de l'unité lexicale de base. Pour exprimer la relation de réparation, par exemple, on choisira pour la lexie VERRE, le verbe RECOLLER et non les verbes RAPIÉCER ou REMONTER ; pour la lexie LIVRE, on préférera les verbes SCOTCHER, RECOLLER et RÉNOVER à RAPIÉCER, RESSOUDER ou encore REPRISER.

Aussi, dans un paradigme lexicologique, les fonctions lexicales peuvent se substituer aux « concepts » de Duvignau. Les verbes comme (SE) DÉTÉRIORER ou ÉCORCER sont, en effet, reliés à des noms particuliers par le biais d'une relation lexicale encodée à l'aide des fonctions lexicales $LiquFunc_0$ ou Degrad, exprimant respectivement les sens 'dégrader/causer la fin' et 'se dégrader'.

Gaume et Duvignau [2004 : 161] donnent plusieurs exemples pour les concepts 'DÉTÉRIORER' et 'RÉPARER', que nous reproduisons dans le tableau 2.II. Or,

<i>Objets</i> DOMAINE	'CONCEPTS D'ACTION'	
	'DÉTÉRIORER'	'RÉPARER'
<i>verre, assiette</i>	casser, briser, ébrécher, fêler	recoller, ressouder
<i>livre, papier</i>	déchirer, découper, friper, froisser	scotcher, recoller, rénover
<i>bois, arbre</i> VÉGÉTAL	couper, cisailer, fendre, tronçonner, casser, entailler	traiter
<i>vêtement, habit</i> COUTURE	déchirer, trouer, découdre, détériorer	rapiécer, recoudre, raccommoder, repriser
<i>corps, psychisme</i> MÉDECINE	casser, gercer, blesser, balafre, meurtrir, démoraliser, bouleverser	soigner, panser, traiter, médicaliser
<i>voiture, bus</i> MÉCANIQUE	dysfonctionner, accidenter, cabosser, esquinter, endommager	réparer, recarrosser, remonter
<i>maison, immeuble</i> BÂTIMENT	fissurer, lézarder, détruire, délabrer	restaurer, ravalier, recrépir

Tableau 2.II – Exemples de concepts d'action : 'DÉTÉRIORER' et 'RÉPARER'

avec les FL Degrad, $LiquFunc_0$, ainsi que CausDe_nouveauPred, CausDe_nouveauFunc_{0/1} et CausDe_nouveauFact₀ (\rightarrow remettre en état), il est possible d'obtenir le même genre de descriptions. Nous avons rassemblé, dans la figure 2.7, un ensemble d'exemples tirés du DiCo qui expriment ces relations.

Duvignau parle de différences interdomaines. Dans notre approche, nous voyons ces différences comme relevant de champs sémantiques particuliers : par exemple,

(SE) DÉTÉRIORER

Degrad(*forêt*#I) = *brûler*
 Degrad(*humeur*#1a) = *s'altérer*
 Degrad(*lait*#b) = *aigrir, cailler, coaguler, surir, tourner*
 Degrad(*maison*#I.1) = *« tomber en ruines », se dégrader, se délabrer*
 Degrad(*terre*#I.1) = *s'appauvrir*
 LiquFact₀(*canalisation*) = *crever, éclater, se détériorer*
 LiquFact₀(*tuyau*#I.1) = *bloquer, boucher, obstruer, crever, percer* [ART ~]
 LiquFunc₀(*curiosité*#I.3) = *désarmer, émousser* [ART ~]
 LiquFunc₀(*forêt*#I) = *couper, déboiser, défricher* [ART ~]
 LiquFunc₀(*maison*#I.1) = *démolir, détruire, « jeter bas », raser* [ART ~]
 LiquFact₀-air(*pneu*) = *dégonfler* [ART ~]
 LiquFact₀-paroi(*pneu*) = *crever, percer* [ART ~]

RÉPARER

CausDe_nouveauFact₀(*évier*) = *déboucher* [ART ~]
 CausDe_nouveauFact₀(*pneu*) = *réparer* [ART ~]
 CausDe_nouveauFact₀(*tuyau*#I.1) = *déboucher, dégorger, nettoyer, purger, réparer* [ART ~]
 CausDe_nouveauFunc₀(*silence*#I.2a) = *ramener, rétablir* [ART ~]
 CausDe_nouveauFunc₀(*soupçon*#I) = *raviver* [ART ~]

Figure 2.7 – FL exprimant les relations sémantiques ‘(se) détériorer’ et ‘réparer’

toutes les lexies d’habitation peuvent avoir les mêmes collocatifs que MAISON#I.1 ; les lexies désignant des véhicules, les mêmes collocatifs que VOITURE, etc. Il s’agit, d’ailleurs, d’un phénomène d’héritage entre champs sémantiques, décrit par Mel’čuk et Wanner [1996] à travers l’exemple du champ sémantique des émotions. Notons qu’à partir de notre modèle d’organisation des relations lexicales (présenté au chapitre 5), il sera possible d’accéder à des relations sémantiquement équivalentes, mais encodées par différentes FL.

2.3.2 Modèle fonctionnel du lexique mental [Aitchison, 2003]

Aitchison [2003] cherche à cerner la façon dont les locuteurs « entreposent » les unités lexicales dans leur mémoire et les processus utilisés pour y accéder. L’objectif général de son travail est de produire l’ébauche d’un modèle fonctionnel de l’organisation du lexique mental de l’anglais. Constatant qu’un être humain adulte connaît plus de 50 000 mots et qu’il peut y accéder très rapidement pour s’exprimer,

elle postule qu'il doit exister une organisation systématique du lexique mental. À partir d'indices tels que les troubles pathologiques de la parole, les lapsus ou distorsions langagières, les expérimentations psycholinguistiques et les travaux menés en linguistique de corpus, il est possible de dégager des indices de la structuration mentale du lexique. Cependant, un tel niveau de complexité ne peut être appréhendé et modélisé que de façon superficielle et partiellement valide dans la mesure où il ne s'appuie que sur des fonctionnements hypothétiques et non sur des données tangibles. Elle compare donc son modèle à un plan du métro londonien, c'est-à-dire à des diagrammes simplifiés encapsulant des traits saillants d'un phénomène en vérité considérablement plus complexe. Elle distingue plusieurs problèmes d'organisation du lexique qui devront être interconnectés pour former l'ébauche d'un modèle du lexique mental. Voyons quelles conceptions du lexique sont supposées dans ses travaux :

- **La nature des sens lexicaux** : le sens des unités lexicales est envisagé selon la théorie du prototype [Rosh, 1973] plutôt que d'un point de vue discret, c'est-à-dire en isolant et en décrivant les différents sens pour une unité lexicale donnée.
- **Les relations entre les unités lexicales** : l'approche visant à rechercher des composantes primitives communes aux unités lexicales est écartée au profit d'une approche par réseau, selon laquelle le lexique mental serait un graphe connecté ayant pour nœuds des unités lexicales reliés par des chemins.

Les psycholinguistes utilisent fréquemment des tests d'association de mots de type stimulus-réponse. Or, si les réponses appartiennent très souvent au même champ sémantique que les stimuli, on remarque cependant que les différentes associations de mots produites constituent des liens de nature différente. Elle classe donc ces liens en quatre types principaux dont les deux premiers sont les plus fréquents : la coordination (par exemple : *apple, pear*), la collocation (par exemple : *wide awake, blissfully ignorant*), la superordination (par exemple : *bird,*

fish, insect = animals) et la synonymie (par exemple : *wide, broad*).

- **Les structures phonétiques dans le lexique mental** : il semble que certaines parties des unités lexicales soient plus saillantes que d'autres, c'est-à-dire plus profondément marquées dans l'esprit humain. Il s'agit du début et de la fin des unités lexicales, ainsi que de leur patron rythmique global. En outre, à l'intérieur des unités lexicales, les syllabes ont leur propre structure. Ceci peut conduire à mélanger accidentellement deux unités relativement similaires du point de vue de leur structure phonétique.

Ces réflexions posent les bases de ce dont devrait rendre compte un modèle dans les processus de stockage, de production et de reconnaissance d'unités lexicales. Dans la perspective de la production, trois faits importants doivent apparaître. Premièrement les humains activent beaucoup plus d'unités lexicales qu'ils n'en ont besoin lorsqu'ils parlent. Deuxièmement, le lemme (sens et partie du discours) peut être séparé de sa forme (le son). Troisièmement, il se peut que le contour d'une unité soit correctement sélectionné mais qu'au moment de la production, une autre unité similaire soit prononcée à sa place. En ce qui concerne la reconnaissance des unités lexicales, tout comme pour le processus de production, les locuteurs envisagent beaucoup plus d'unités lexicales que celles qui se retrouvent sélectionnées à la fin. Un grand nombre d'entre elles sont activées puis graduellement filtrées et supprimées.

Aitchison déclare qu'il est impossible, à l'heure actuelle, d'implémenter un tel modèle en raison de sa très haute complexité. Toutefois, cette esquisse du fonctionnement du lexique mental apporte des pistes originales pour l'enrichissement de la structuration d'une base de données lexicale. La prise en compte des proximités phonétiques pour éviter les erreurs semble tout à fait novatrice et pertinente. Par ailleurs, parmi les relations qui ont pu être dégagées lors des tests d'association de mots, on note que, si la synonymie, l'hyponymie et les collocations sont

très répandues, la coordination (permettant d'identifier des séries paradigmatiques telles que *sel, poivre, moutarde*) reste absente des dictionnaires. Un dernier point qui nous semble fondamental en vue d'une application lexicographique, est l'idée d'activer un très grand nombre de liens pour laisser à l'utilisateur la possibilité de filtrer et de sélectionner lui-même les éléments dont il a besoin. D'après nous, cette dynamique de sélection des unités lexicales mériterait d'être implémentée dans une base de données lexicale. C'est en partie ce que nous proposons dans notre modèle de structuration, qui sera présenté au chapitre 5.

2.4 Conclusion

Ce deuxième chapitre a permis de clarifier certains points nécessaires à la compréhension de notre problématique. Premièrement, nous avons évoqué les différentes disciplines scientifiques s'attachant à décrire le lexique de façon relationnelle. Les différents modèles hérités de ces disciplines mettent en avant des structures du lexique qui se différencient par un certain nombre de propriétés. Les relations elles-mêmes sont de diverses natures : les modèles linguistiques considèrent les relations entre unités lexicales attestées le plus souvent en corpus sur les plans à la fois paradigmatique et syntagmatique ; les bases de connaissances et les modèles cognitifs tendent davantage à envisager des relations entre concepts. Les premiers cherchent à représenter le lexique en usage, tel qu'il est exprimé dans les textes, les seconds, à représenter la façon dont le lexique est organisé dans le cerveau humain. Par ailleurs, le nombre de relations considérées varie considérablement. Celles-ci ne sont pas identifiées de façon univoque, mais en corrélation avec les objectifs visés par les modèles.

Nous avons également dressé une synthèse comparative des principales ressources lexicales relationnelles existantes pour l'anglais, le français et l'allemand. Ces ressources sont de différentes formes : dictionnaires (informatisés ou non), bases de données relationnelles et graphes lexicaux. Elles ont en commun une philosophie relationnelle, dans laquelle une unité lexicale se définit par les associations qu'elle

entretient avec d'autres. Elles diffèrent toutefois par leurs arrière-plans théoriques, par les relations qui y sont identifiées, par leurs structures, ainsi que par la façon dont elles ont été créées : alors que certaines sont élaborées selon des méthodes automatiques (cf. MindNet) et mettent l'accent sur la quantité de données traitées, d'autres optent pour une approche semi-automatisée (cf. Antidote) et privilégient la précision parmi les données recueillies. D'autres encore, comme le DiCo et Elexico, effectuent un relevé manuel des données dans le but de présenter des informations très riches et complètes sur chaque unité lexicale, mais ne comptent que peu d'entrées.

Notre intérêt dans cette thèse porte plus particulièrement sur l'étude des relations lexicales. Or, si l'on parcourt l'état de l'art qui vient d'être dressé sur les modèles relationnels du lexique et les différentes ressources lexicales relationnelles, on remarque que peu d'entre eux proposent une étude détaillée des relations lexicales. Seul le modèle théorique de la LEC semble accorder une attention très particulière aux relations qu'il modélise, et ce, non seulement du fait de leur grand nombre, mais aussi en raison de la caractérisation qui est faite de chacune d'elles dans le système formel des données lexicales. L'utilisation de ce modèle comme cadre théorique nous semble, par conséquent, tout à fait adapté à notre objectif, qui consiste, rappelons-le, à proposer un modèle de structuration multi-dimensionnelle des relations lexicales. Le choix de la LEC comme cadre de travail présente un certain nombre d'avantages et suppose certaines positions théoriques vis-à-vis des relations. Énumérons-les en détail. Tout d'abord, en adoptant les principes théoriques de la LEC, nous restons dans une vision linguistique et lexicologique du lexique. Aussi, nous écartons délibérément les considérations conceptuelles ou psycholinguistiques pour nous centrer sur une étude des relations lexicales. Par ailleurs, la LEC a l'avantage de prendre en considération des relations syntagmatiques. À la différence de ses modèles concurrents, l'éventail des relations couvertes est conséquent et permet d'avoir, à travers les relations associées à chaque unité lexicale, une couverture du lexique beaucoup plus importante. Enfin, l'aspect le plus important

est certainement la modélisation des relations au moyen d'un encodage formel exploitable pour des traitements automatisés ; cette modélisation portant à la fois sur les caractéristiques sémantiques et syntaxiques des relations. C'est en ceci que le DiCo se distingue particulièrement de toutes les autres ressources. La modélisation des relations lexicales au moyen des fonctions lexicales s'avère fondamentale pour la poursuite de notre objectif.

Nous consacrons le chapitre suivant à la présentation des fonctions lexicales, qui constituent l'outil théorique à partir duquel nous souhaitons bâtir notre modèle. Ce chapitre consiste plus précisément à dresser un bilan sur les fonctions lexicales. Nous essaierons de déterminer en quoi elles se révèlent un bon outil pour la structuration du lexique, tout en mettant en valeur certaines lacunes que nous tenterons, en partie, de combler tout au long de la thèse.

Chapitre 3

Les fonctions lexicales : aspects formels et notionnels

3.1 Introduction

Nous avons consacré le chapitre précédent à la présentation de modèles théoriques du lexique et de plusieurs bases de données lexicales. Cet état de l’art nous a permis de mettre en perspective la LEC par rapport aux autres modèles relationnels du lexique. Nous consacrons ce troisième chapitre à la présentation de notre outil théorique : les fonctions lexicales. Les fonctions lexicales (dorénavant notées *FL*) ont été définies il y a une cinquantaine d’années. Nous souhaitons en dresser un bilan afin de mettre en valeur leurs forces et leurs faiblesses, et de déterminer si elles se montrent satisfaisantes pour servir les besoins des applications dans lesquelles elles sont exploitées. Les FL ont été utilisées pour la rédaction de plusieurs dictionnaires et bases de données lexicales monolingues, bilingues, didactiques et spécialisées, ainsi que dans diverses applications en traitement automatique des langues. La plupart des ressources et applications développées avec les FL sont présentées en détail à l’annexe II.

Nous avons choisi de présenter les FL en distinguant, d’une part, la notion théorique de FL, et, d’autre part, les différentes formes d’encodage des FL, afin de nous écarter des présentations canoniques qui en sont faites. Les FL sont traditionnellement définies d’un point de vue formel comme des fonctions de type mathématique notées $f(x) = y$, où y représente la valeur — plus précisément un ensemble de valeurs — et x , le mot-clé ou argument de la fonction. Nous écartons temporairement l’aspect formel pour nous concentrer sur l’invariant conceptuel que l’on peut déga-

ger de ces différents encodages. Polguère [2003] a déjà présenté l'intérêt de dévêtir les FL de leur formalisme dans un cadre pédagogique où l'aridité de la forme n'est pas appropriée en tant que support pour l'apprentissage de la notion et pour son utilisation. Séparer la notion de la forme permet une certaine prise de recul pour percevoir les principes sous-jacents à la conceptualisation des FL et les différentes formes qu'elles revêtent, les critiques que nous formulerons portant sur ces deux facettes.

Nous présentons, dans un premier temps (section 3.2), la notion de FL en développant plusieurs aspects théoriques. Dans un deuxième temps (section 3.3), nous présentons les différentes formes d'encodage formalisant la notion de FL, soit l'encodage de vulgarisation utilisé dans le *Lexique Actif du Français* (= LAF), l'encodage traditionnel, ainsi que les encodages algébrique et explicite. Nous mettons l'accent sur l'encodage traditionnel, pour lequel nous proposons une typologie et une discussion critique.

3.2 La notion de fonction lexicale

Afin de fournir une caractérisation complète de la notion de FL, nous envisageons plusieurs aspects liés à leur définition : leur nature et le rôle qu'elles jouent dans le modèle Sens-Texte (3.2.1) ainsi que leur caractère graduel et les conditions permettant la distinction de plusieurs statuts (3.2.2).

3.2.1 Nature et rôle des fonctions lexicales

La notion de FL a vu le jour au début des années 60, plus précisément en 1963-64. Leurs premières descriptions ont été présentées dans Zholkovsky et Mel'čuk [1965, 1966, 1967] en russe, Zholkovsky et Mel'čuk [1970] en français et Mel'čuk et Zholkovsky [1970] en anglais. Mel'čuk [1994b : 194] explique la genèse du concept de FL comme suit :

Nous avons découvert les Fonctions Lexicales [FL] il y a plus de 30 ans, lors de notre participation à une expédition géologique dans la zone

montagneuse semi-désertique, au sud du Kazakhstan — en vue de les utiliser pour la traduction automatique. Nous avons eu l'idée de ce qui est devenu plus tard (avec la collaboration précieuse de A. Zolkovskij) les FL *Magn* et *Oper*₁, quand nous étions en train de chercher une méthode simple permettant d'éviter les milliers de tests ennuyeux nécessaires pour permettre à l'ordinateur de trouver les équivalents russes — déterminés par le contexte lexical — de lexèmes anglais comme HEAVY, IMPORTANT, EXTENSIVE, HIGH. etc., d'une part, et [to] DO, [to] MAKE, [to] GIVE, [to] GET, etc., de l'autre.[...] Les FL devaient permettre d'établir les correspondances nécessaires de façon directe et logique.

Mais dès que le premier jeu des FL a été proposé, il s'est avéré que les FL ont par ailleurs un statut théorique fort important en linguistique. *Primo*, elles constituent le chaînon manquant de la THÉORIE DE LA PHRASÉOLOGIE, parce qu'elles permettent de décrire de façon rigoureuse et systématique les **collocations** [...]. *Secundo*, elles sont essentielles dans la THÉORIE DE LA SYNTAXE car plusieurs régularités syntaxiques ont besoin du concept de FL pour se prêter à une description formelle satisfaisante.[...] *Tertio*, les FL ont ouvert des perspectives fort prometteuses dans la THÉORIE DE LA LEXICOGRAPHIE.

Wanner et Alonso Ramos [2005] ont mis en valeur l'ambiguïté de l'entité FL, qui sert différents types d'usage dans le modèle Sens-Texte et ses applications. Pour la présentation de la notion de FL, deux points nous semblent importants à clarifier par rapport à la nature des FL : le fait qu'elles sont considérées comme des métalexies et comme des fonctions.

Polguère [2003 : 125-126] propose de considérer les FL comme des métalexies :

Pour être bien comprises, les FL doivent avant tout être appréhendées comme des métalexies : elles sont caractérisées par des propriétés de sens et de combinatoire, mais ne sont pas associées de façon directe à des signes, et donc à des signifiants. L'association à des signifiants particuliers se fait, de façon indirecte, par l'application de la métalexie (en tant que fonction) à une lexie particulière, préalablement identifiée par le locuteur. Ces métalexies universelles devraient être enseignées comme telles. La FL *Magn* existe, même si c'est une notion « abstraite ». Elle existe au même titre que la notion de partie du discours, ou de genre grammatical.

La confusion entre les FL et le formalisme qui permet de les identifier est très fréquente. L’encodage traditionnel des FL, jugé trop aride, tend parfois à décourager certains lexicographes ou didacticiens, qui pourraient pourtant bénéficier des potentialités qu’elles ont à offrir pour la représentation et l’enseignement du lexique. Dans la suite de ce chapitre, nous tentons, autant que possible, de présenter la notion de FL en mettant de côté la question de l’encodage afin d’éviter au lecteur de faire l’amalgame entre les deux.

Voyons à présent pour quelles raisons on considère les FL comme des fonctions. Elles modélisent deux types de relations lexicales : la dérivation sémantique et les collocations. Ces deux notions ont été définies et présentées en détail à la section 2.2.1 (p. 28) du chapitre précédent. Les dérivations sémantiques et les collocations sont orientées d’une lexie de départ vers une lexie cible, le choix de la deuxième étant déterminé par la première. Ainsi, dans l’expression collocationnelle : *administrer un médicament*, l’emploi du verbe ADMINISTRER est contraint par celui de la lexie MÉDICAMENT. Comme ces liens sont orientés, ils peuvent être représentés par des fonctions. Ces fonctions sont dites lexicales parce qu’elles prennent en argument (ou mot-clé) et retournent comme valeurs des unités lexicales. Toutefois, alors que le mot-clé d’une FL est une unité lexicale unique, sa valeur est constituée d’un ensemble d’unités lexicales, par exemple, la FL Oper_1 , représentant le verbe support d’une unité lexicale, appliquée à $\text{COMMÉRAGE}\#A$, retourne l’ensemble d’unités lexicales {DIRE, FAIRE, SE LIVRER} : $\text{Oper}_1(\text{commérag}\#a) = \text{dire, faire} [\text{ART} \sim s]; \text{se livrer} [\text{à ART} \sim s]$.

Les FL identifient deux aspects d’une relation lexicale. D’une part, son sens (intensification, réalisation, etc.) et, d’autre part, son patron syntaxique (mot-clé-sujet + verbe intransitif; verbe + mot-clé-complément, etc.). Elles synthétisent donc le lien entre la sémantique et la syntaxe. C’est en ceci qu’elles s’intègrent dans les modèles de représentation Sens-Texte. La TST comporte plusieurs niveaux de représentation, dont les premiers sont les représentations sémantique, syntaxique profonde et

syntactique de surface¹. Les FL apparaissent dans les représentations syntaxiques profondes, qui font le lien entre les représentations sémantiques et les représentations syntaxiques de surface. Elles constituent alors des unités lexicales profondes, dont le rôle est de pointer vers une famille d'unités lexicales qui remplissent le rôle sémantico-syntaxique correspondant [Mel'čuk, 1996 : 80]. Ainsi, dans la représentation de la figure 3.1, la fonction AntiReal_2 appliquée à la lexie BARRICADE (et vulgarisée par la glose : $\rightarrow [Y]$ surmonter l'obstacle que représente \sim) peut prendre la forme, en syntaxe de surface, des unités lexicales *enfonce*, *franchir*, *forcer*, etc.



Figure 3.1 – Correspondance entre FL et unités lexicales

Les FL permettent également de modéliser des opérations de paraphrasage. On entend par paraphrasage, la production de phrases (quasi-)synonymes ou paraphrases. Le système de paraphrasage dans la Théorie Sens-Texte a été proposé par Zholkovsky et Mel'čuk [1965] puis réévalué à plusieurs reprises, notamment par Miličević [2007]². En bref, un ensemble de règles sont formulées pour effectuer le passage d'une représentation syntaxique profonde à une autre. Par exemple, la fonction V_0 fournit l'équivalent verbal d'une unité lexicale nominale, la fonction Oper_1 permet, quant à elle, d'exprimer une unité lexicale nominale au moyen d'un verbe support, par exemple :

$$V_0(\textit{soupir}) = \textit{soupirer} ;$$

$$\text{Oper}_1(\textit{soupir}) = \textit{pousser} [\text{ART } \sim].$$

On peut alors formuler la règle suivante (fig. 3.2) qui permettra de paraphraser *Pierre pousse un soupir* par *Pierre soupire* et vice versa ; *Le chien émet un aboiement* par *Le chien aboie* ; *La voisine se livre à des commérages* par *La voisine*

¹Les différents modules sont présentés à la section 5.3.2 (p. 193).

²Se reporter à Miličević [2007] pour de plus amples détails.

commère, etc. L'équivalence entre les expressions est représentée par le symbole \equiv :

$$\begin{aligned} \text{Oper}_1(L) + L &\equiv V_0(L) \\ \text{pousser [=Oper}_1(L)] \text{ un soupir [= L]} &\equiv \text{soupirer [= V}_0(\text{soupir})] \\ \text{émettre [=Oper}_1(L)] \text{ un aboiement [= L]} &\equiv \text{aboyer [= V}_0(\text{aboiement})] \\ \text{se livrer à [=Oper}_1(L)] \text{ des commérages [= L]} &\equiv \text{commérer [= V}_0(\text{comméragé})] \end{aligned}$$

Figure 3.2 – Exemple de règle de paraphrasage

Les FL et les différentes règles de paraphrasage sont utilisées en TAL pour la génération de paraphrases [Nasr, 1996, Lareau, 2002], la traduction automatique [Apresjan et al., 2003] et l'élaboration automatique de résumés [Bélanger et Kittredge, 2005]. Parmi ces réalisations, certaines ne sont encore qu'à un stade expérimental, alors que d'autres ont donné lieu à des travaux plus aboutis (cf. annexe II).

3.2.2 Caractère graduel des fonctions lexicales

La notion de FL a un caractère graduel : les FL ne se situent pas toutes sur le même plan mais sont réparties selon plusieurs statuts. Il existe des FL standard, semi-standard, localement standard et non standard. Ces statuts sont déterminés par plusieurs conditions. On considère qu'une FL est standard si elle répond aux exigences imposées par les conditions de cardinalité, de diversité et d'universalité. Nous présentons ces conditions ci-dessous puis détaillons chacun des statuts de FL.

3.2.2.1 Les conditions déterminant le caractère standard d'une FL

a) Cardinalité des arguments et valeurs de FL

Premièrement, on entend par cardinalité des arguments de FL, le nombre d'arguments auxquels une FL s'applique. Ce critère impose à une FL de prendre en argument un nombre important d'unités lexicales. Le critère vaut également pour

les valeurs de FL. Une FL doit retourner plusieurs unités lexicales. Ainsi, dans l'exemple cité plus haut (fig. 3.1), *AntiReal*₂ appliquée à l'argument *BARRICADE*, peut prendre les formes suivantes : *ENFONCER*, *FRANCHIR*, *FORCER*, *PASSER*, *SAUTER*, etc.

b) Diversité des arguments et valeurs de FL

La diversité des arguments concerne leur type sémantique. On considère que les arguments d'une FL doivent être diversifiés et non relever d'une classe sémantique spécifique. Selon ce principe, la relation entre *canard* et *d'élevage* ou *domestique*, dont le sens peut être décrit par la formule 'qui ne vit pas à l'état sauvage' ne concerne que certains noms d'animaux et ne pourra donc être représentée une FL standard.

La diversité des valeurs de FL tient à leur différence non plus sur le plan sémantique mais sur le plan formel. Une FL appliquée à plusieurs arguments doit, en principe — mis à part les cas d'héritage — retourner des valeurs différentes. Par exemple, *Magn* correspond à un intensifieur et retourne des valeurs différentes selon l'argument auquel elle est appliquée :

Magn(aboitement#I) = *déchaîné, féroce, fort, furieux* ;

Magn(admiration) = *folle, grande, profonde, 'sans borne', vive, ...* ;

Magn(catastrophe#I) = *énorme, épouvantable, grosse, majeure*.

c) Universalité des FL

Les FL standard sont considérées comme des entités lexicales universelles, leur existence devant être prouvée dans toutes les langues. Toutefois, la notion d'universel linguistique doit être entendue telle que Polguère [2003 : 122] la présente :

Les liens lexicaux tels que la synonymie sont particulièrement intéressants du fait que ce sont des UNIVERSAUX LINGUISTIQUES : toute langue naturelle possède la synonymie, et apprendre une langue naturelle consiste aussi à apprendre (encore une fois, de façon plus ou moins consciente) que la synonymie existe. Il est essentiel de distinguer un lien de synonymie donné, qui lie deux lexies particulières de la langue,

et la synonymie elle-même, en tant qu'entité informationnelle, en tant que type de lien universellement présent dans toutes les langues. [...] Ce que je viens de dire à propos de la synonymie s'applique à d'autres liens bien connus en lexicologie, comme l'antonymie ou la conversivité. Cela s'applique aussi, de façon plus générale, à toutes les FL standard.

3.2.2.2 Les statuts de FL

On accorde un statut aux FL selon qu'elles respectent ou non les conditions présentées ci-dessus. On distingue quatre statuts : les FL standard, les FL semi-standard, les FL localement standard ainsi que les FL non standard.

a) Les FL standard

Les FL standard doivent répondre aux principes de cardinalité, de diversité et d'universalité cités plus haut. Nous distinguons deux types de FL standard : celles entrant en jeu dans la formation de règles de paraphrasage et celles n'y entrant pas.

1. Les FL des règles de paraphrasage

Nous avons mentionné plus haut que l'objectif premier des FL est de former des règles de paraphrasage comme celle de la figure 3.2 (p. 73). Finalement, les FL apparaissant dans les règles de paraphrasage devraient être, par essence, les FL prototypiques, puisqu'elles seules remplissent véritablement ce rôle. Si l'on cherche à poser à tout prix un groupe de FL de référence, nous pensons que c'est précisément cet ensemble qui devrait figurer comme tel. Les FL entrant dans les règles de paraphrasage énoncées dans Mel'čuk et al. [1992] et Miličević [2003] sont les suivantes :

Syn	Real _{i(j)}	V ₀	Labor _{ij}	S _{res}
Conv _{ij(k)(l)}	Anti	Oper _i	S ₂	CausFunc ₀
Adv ₁	S ₀	Func ₀	A ₂	Smed
S ₁	A ₀	Func _i	S _{instr}	CausOper ₁

Gener	Loc _{in}	Pred	Fin	Minus
Figur	S _{loc}	Liqu	Incep	
Sing	Instr	Caus	Cont	
Mult	S _{mod}	Perm	Plus	

Nous n'excluons pas la possibilité que cette liste puisse évoluer. Il est, en effet, probable que de nouvelles règles de paraphrasage puissent être formulées à l'aide d'autres FL.

2. Les FL standard hors règles de paraphrasage

Très semblables aux premières, ces autres FL sont distinguées par le seul fait qu'elles n'entrent pas dans l'élaboration de règles. Elles sont pourtant couramment citées dans les listes de référence des FL standard, et remplissent tous les critères requis pour être considérées comme telles. Ces FL sont très importantes pour l'identification de collocations et de dérivés sémantiques, car elles modélisent des relations les plus récurrentes. Parmi elles, on trouve les FL de modification : Magn, Bon, Ver, Pos, Epit ; les qualificateurs actantiels : Able, Qual ; quelques FL prépositionnelles : Propt, Instr, Loc_{ab/ad} ; un ensemble de FL verbales : Sympt, Manif, Son, Involv, Degrad, Obstr, Excess, Result, Perf, Prox, Prepar ; des FL paradigmatiques : Cf, Contr, Cap, Equip ; etc.

b) Les FL semi-standard

Les FL semi-standard sont, en quelque sorte, des extensions des FL standard permettant d'ajouter une composante de sens qui n'est pas prise en compte dans les FL standard dont elles sont dérivées. Par exemple, la description de la relation (5) ayant le sens 'causer le silence de qqn', se différencie de la relation (6) par la présence de la composante sémantique 'en échange de qqch.' qui apporte un complément au sens de CausOper₁.

(5) CausOper₁(*silence*) = *réduire* [qqn. au ~] ;

(6) en échange de qqch. $\text{CausOper}_1(\textit{silence}) = \textit{acheter}$ [le \sim de qqn.].

c) Les FL localement standard

Les FL localement standard ont été proposées par Polguère [2007] en tant que classe de FL à statut hybride entre le standard et le non standard, pour les relations dont la nature universelle n'est pas démontrée, mais qui peuvent être encodées par des FL standard dans le contexte d'une langue naturelle donnée. Ainsi, les fonctions *De_nouveau* et *Essayer_de*, illustrées en (7), (8) et (9) sont très fréquentes en français sans pour autant être tout à fait standard, car leur universalité reste à prouver. Elles sont, par conséquent, considérées comme des FL standard locales du français.

(7) $\text{De_nouveauFunc}_0(\textit{hostilité}) = \textit{reprendre}$ (\rightarrow commencer à nouveau);

(8) $\text{De_nouveauOper}_{12}(\textit{goût}) = \textit{reprendre}$ [\sim à qqch.] (\rightarrow avoir de nouveau). ;

(9) $\text{Essayer_de Oper}_2(\textit{augmentation\#2}) = \textit{demander, réclamer}$ [ART \sim] (\rightarrow Demander une \sim).

d) Les FL non standard

Les FL non standard (FLNS), quant à elles, ne satisfont aucun critère mentionné. Par exemple, la FLNS suivante :

Que X porte pour se protéger du soleil(*chapeau*) = *de soleil*

ne remplit pas les conditions nécessaires. Elle ne s'applique qu'à deux unités lexicales : CHAPEAU et LUNETTES et retourne la même valeur.

Ces FL n'entrent pas non plus dans la composition de règles de paraphrasage. N'étant pas de vraies FL, on ne devrait pas, en principe, les appeler fonctions lexicales. Cet abus de langage est dû à leur rôle : elles figurent, au même titre que les FL standard, dans les dictionnaires et bases de données lexicales pour recenser les dérivations sémantiques et les collocations.

3.3 Les encodages de fonctions lexicales

À la suite de la présentation de la notion de FL, cette section est destinée à approfondir la question de leur encodage. Premièrement, nous présentons les deux encodages les plus communément utilisés : l'encodage vulgarisé (tel qu'on le trouve, par exemple, dans le *Lexique Actif du Français* (ou LAF) [Mel'čuk et Polguère, 2007]), et l'encodage dit « traditionnel ». Bien que l'encodage vulgarisé ait été créé postérieurement à l'encodage traditionnel dans le but d'en offrir une vulgarisation, il est formellement plus simple et plus intuitif. Il semble donc plus naturel de le présenter en premier. L'encodage traditionnel peut alors être vu comme une formalisation de la description des relations lexicales telle qu'elle est réalisée par l'encodage vulgarisé. Nous proposons une typologie de l'encodage traditionnel pour recenser l'ensemble des formules qu'il recouvre et souligner sa grande hétérogénéité. La visualisation des différentes formes de l'encodage traditionnel nous amène à en dresser un bilan critique. Finalement, nous montrons comment Kahane et Polguère [2001] dépassent les insuffisances de cet encodage en proposant une définition formelle des FL à travers un encodage algébrique et un encodage explicite.

3.3.1 L'encodage vulgarisé

L'encodage des FL dans des formulations claires et accessibles a une très grande importance à la fois dans une optique pédagogique d'enseignement des notions lexicologiques et dans l'utilisation quotidienne d'un dictionnaire offert au grand public. Cet encodage « grand public » des FL figure dans le DiCo en tant que vulgarisation des FL de l'encodage traditionnel et dans le LAF.

Les FL y sont encodées de deux façons. Certaines sont associées à une paraphrase unique, par exemple, la fonction modélisant le nom typique du premier actant d'une unité lexicale (= X) est notée : **Nom pour X**. D'autres FL reçoivent plusieurs formulations selon les lexies auxquelles elles sont appliquées. Ce principe est exposé par Popovic [2004 : 67] comme suit :

La vulgarisation est conditionnée à la fois par la FL que l'on veut encoder et par l'étiquette sémantique de la lexie concernée. Par exemple, si on cherche à faire la paraphrase linguistique de la FL **Magn** [représentant l'intensification], la vulgarisation pourrait être *intense*, pour une lexie de sentiment (par exemple : *peur intense*), mais pas pour une lexie d'événement (**catastrophe intense* vs. *catastrophe importante*).

Selon ce procédé, la FL représentant un adjectif ou un adverbe exprimant la qualification négative d'une lexie (soit **AntiBon** dans l'encodage traditionnel) est vulgarisée par \rightarrow **indésirable** pour la lexie ABOIEMENT, par \rightarrow **mauvais** pour la lexie COMPLIMENT#a, par \rightarrow **Qui a un mauvais comportement** pour la lexie CHIEN, etc. Nous présentons dans la figure 3.3 (p. 80) la correspondance entre une fiche DiCo et un article du LAF à travers la description de la lexie ABOIEMENT.

L'encodage vulgarisé n'est pas propre au DiCo ou au LAF, mais concerne également les autres BLEC. On note par exemple que Alonso Ramos [2006] a établi un certain nombre de critères permettant de guider l'attribution de gloses aux FL lors de la description des collocations du DiCE.

3.3.2 L'encodage traditionnel

L'encodage traditionnel est l'encodage le plus majoritairement utilisé. Il s'agit d'une représentation des FL au moyen de formules destinées au travail lexicographique formel et conçues de façon à être manipulées dans des traitements automatiques, tels que la génération de phrases. Nous souhaitons ici en dresser un bilan. Pour ce faire, nous commençons par expliquer certains aspects formels pour les lecteurs non familiers avec cet encodage. Dans un deuxième temps, nous présentons la façon dont l'encodage traditionnel est constitué d'un ensemble de FL standard simples pouvant former des FL standard combinées. Par la suite, nous proposons une typologie de l'encodage traditionnel, destinée à recenser et organiser l'ensemble des formules qu'il peut recouvrir. La présentation de ces divers types nous amènera, enfin, à une discussion critique sur la validité de cet encodage.

FICHE DICO

ABOIEMENT

QSyn ouah-ouah

Cf grognement#I; hurlement#I.2

Gener cri#I.3

V₀ aboyer

S₁ (→ Nom pour X) chien#I

Magn (→ Intense) fort < furieux; féroce

AntiMagn (→ Peu intense) faible //jappement

AntiBon (→ Indésirables) intempestifs | postpos et A. au pl

Oper₁ (→ [X] émettre ~) émettre, pousser [ART ~]; lancer [ART ~] | A. surtout au sing

Article LAF

ABOIEMENT

☞ ouah-ouah

☞☞ grognement#I; hurlement#I.2

Génér. cri#I.3

Verbe aboyer

Nom pour X chien#I

Intense fort < furieux; féroce

Peu intense faible //jappement

Indésirables intempestifs | postpos et A. au pl

[X] émettre un A. émettre, pousser [ART ~]; lancer [ART ~] | A. surtout au sing

Figure 3.3 – Comparaison de l'article LAF et de la fiche DiCo pour ABOIEMENT

3.3.2.1 Précisions préliminaires pour la compréhension de l'encodage traditionnel des FL

Avant d'entamer la discussion sur l'encodage traditionnel, nous devons préciser quelques éléments formels pour faciliter leur interprétation. Nous nous attardons principalement sur la représentation des actants.

Les actants sont représentés par des indices en chiffres arabes : 1, 2, 3, ..., n. Ces chiffres renvoient aux actants syntaxiques profonds du mot-clé auxquels s'applique

une fonction³.

◇ **Les indices i et j**

– **dans les FL nominales, adjectivales ou adverbiales :**

Le numéro i renvoie à l'actant de la lexie concernée : S_1 représente le nom typique du premier actant syntaxique profond d'une lexie, A_2 le modifieur typique du second actant syntaxique profond d'une lexie, par exemple :

$$S_1(\textit{bicyclette}) = \textit{cycliste} ;$$

$$A_2(\textit{mépris}) = \textit{couvert} [\textit{de} \sim].$$

– **dans les FL verbales :**

Les FL de type $\text{Oper}_{i(j)}$ et $\text{Real}_{i(j)}$ prennent le mot-clé comme complément d'objet direct. L'indice i signifie que le sujet grammatical de la valeur correspond à l'actant syntaxique profond i du mot-clé. Dans les deux exemples suivants, le premier actant de la lexie ABOIEMENT, de type sémantique « animal », correspond au sujet grammatical des verbes ÉMETTRE, POUSSER et LANCER :

$$\text{Oper}_1(\textit{aboiement}) = \textit{émettre} , \textit{pousser} [\textit{ART} \sim];$$

$$\text{Oper}_{12}(\textit{aboiement}) = \textit{lancer} [\textit{ART} \sim_s \ulcorner \textit{en direction de} \urcorner \text{N}=\text{Z}].$$

On note que les FL telles que Oper_{12} sont une invention du DiCo. Elles servent à distinguer les verbes supports bi-actantiels des verbes supports tri-actantiels. Cette question a été traitée notamment par Alonso Ramos [2007].

L'indice j , s'il est présent, indique que la valeur prend un second complément correspondant à l'actant syntaxique profond j du mot-clé. Ainsi, dans l'exemple suivant,

$$\text{Real}_{12}(\textit{colle}\#\text{II} \textit{'question'}) = \textit{poser} , \textit{soumettre} [\textit{ART} \sim \grave{\text{a}} \text{N}=\text{Y}],$$

le premier actant de la lexie COLLE#II est le sujet grammatical des verbes POSER et SOUMETTRE, et son second actant, le second complément de ces verbes.

Les FL de type $\text{Func}_{i(j)}$ et $\text{Fact}_{i(j)}$ prennent le mot-clé comme sujet grammatical. Par conséquent, les indices actantiels i et j renvoient aux premier et deuxième compléments du verbe, par exemple :

$$\text{Fact}_{21}(\textit{piste}\#\text{II}) = \textit{mener} [\text{N}=\text{Y} \grave{\text{a}} \text{N}=\text{X}].$$

³Se reporter à la section 5.3.2 pour une présentation de la structure syntaxique profonde.

Pour les FL de type Labor_{ij} et Labreal_{ij} , l'indice i renvoie toujours au sujet grammatical et l'indice j au premier complément du verbe. Le mot-clé, lui, prend la fonction de second complément.

$$\text{Labreal}_{12}(\text{lait}\#\text{a}) = \text{nourrir} [\text{N}=\text{Y de/avec A-poss}=\text{X} \sim].$$

◇ **L'indice @**

L'indice @ est une autre invention du DiCo. Il ne correspond à aucun actant syntaxique profond du mot-clé, mais représente un participant extérieur à la structure actantielle de ce dernier. Par exemple, dans la relation suivante :

$$\text{Real}_{@}(\text{fleuve}\#\text{I}) = \text{naviguer} [\text{sur ART} \sim],$$

le sujet du verbe n'est pas compris dans la structure actantielle de FLEUVE#I :

$$\sim [\text{QUI COMMENCE AU lieu X, PASSE PAR LES lieux Z ET SE TERMINE DANS L'étendue d'eau Y}]$$

Il existe également un indice zéro. Il apparaît dans les FL dites de dérivation syntaxique : S_0 , V_0 , A_0 et Adv_0 ainsi que dans certaines FL verbales comme Func_0 et Fact_0 . Dans les FL de dérivation syntaxique, il ne représente aucun actant syntaxique profond, mais marque le phénomène de translation qui permet de passer d'une partie du discours à une autre, sans qu'il y ait pour autant de changement de sens, par exemple :

$$\text{S}_0(\text{regretter}\#\text{I.1}) = \text{regret}\#\text{I.1};$$

$$\text{V}_0(\text{assassinat}) = \text{tuer};$$

$$\text{A}_0(\text{chat}) = \text{félin};$$

$$\text{Adv}_0(\text{exagérer}\#\text{1a}) = \text{exagérément}.$$

Dans les FL verbales, il indique le fait que les verbes retournés en valeur n'ont pas de complément, par exemple :

$$\text{Func}_0(\text{applaudissements}) = \text{crépiter, résonner};$$

$$\text{Fact}_0(\text{ancree}) = \text{mordre, tenir}.$$

3.3.2.2 Fonctions lexicales standard simples et fonctions lexicales standard combinées

Nous devons à présent définir ce que l'on entend par FL standard simples et FL standard combinées. Les FL standard simples (= FLSS) forment un ensemble de primitifs sémantiques dégagés à partir de l'observation d'un grand ensemble de FL. De la même façon qu'il est possible d'isoler les composantes sémantiques minimales d'une définition analytique, il est possible d'isoler des FLSS à partir desquelles on peut former d'autres FL standard. Par exemple, dans la relation suivante :

$$S_1 \text{Real}_1(\text{gouvernail}\#\text{I}) = \text{pilote}, \text{timonier},$$

le sens de la FL peut être défini comme 'Individu qui (= 'S₁') actionne ~ (= 'Real₁')'. Les FL standard obtenues par la combinaison de FLSS doivent être interprétées comme des FL à part entière, mais dont le sens est décomposable en termes d'éléments plus simples. Ainsi, la FL *AntiBon* est une FL standard combinée, mais elle constitue une FL unique, au même titre que la FL *Anti* et que la FL *Bon*. Les FL standard combinées ne s'analysent pas toutes de la même façon sur le plan sémantique. Il existe deux types de combinaisons, régies par des règles particulières qui en déterminent l'interprétation. Ces combinaisons sont présentées dans Mel'čuk et al. [1995 : 148-149]. Nous les synthétisons brièvement ci-dessous.

1. Les FL complexes

Une FL complexe est une combinaison de FL simples, syntaxiquement reliées et formant une expression unique. Les FL complexes comportent au moins deux FL simples. D'après Mel'čuk et al. [1995 : 148], une FL complexe n'est pas une composition de FL : la valeur d'une FL complexe ne s'obtient pas par l'application successive de ses éléments : $fg(L) \neq f(g(L))$. Par exemple, la relation

$$\text{IncepOper}_1(\text{feu}) = \text{ouvrir} [\text{ART } \sim],$$

ne peut pas se décomposer de la façon suivante :

$$\text{Oper}_1(\text{feu}) = \text{faire} \text{ puis } *\text{Incep}(\text{faire}) = \text{ouvrir}.$$

De même, la relation

$$\text{IncepOper}_1(\textit{amoureux}) = \textit{tomber} [\sim]$$

n'est nullement le résultat de la composition de ses composantes.

Si $\text{Oper}_1(\textit{amoureux}) = \textit{\hat{e}tre} [\sim]$ est valide, en revanche, la suivante ne l'est pas : $\text{Incep}(\textit{\hat{e}tre}) \neq \textit{tomber}$ (la valeur correspondante serait plutôt *commencer* [à \sim]).

Mel'čuk et al. [1995 : 148] donnent la définition suivante des FL complexes :

Nous appelons *fonction lexicale complexe* un enchaînement de FL simples syntaxiquement liées, cet enchaînement ayant une valeur globale cumulative, qui exprime, de façon indécomposable, le sens de l'enchaînement entier.

2. Les configurations de FL

Une configuration de FL est une chaîne de FL non reliées syntaxiquement et qui ont toutes la même lexie pour mot-clé. Dans une configuration de FL, les différentes FL sont séparées traditionnellement par le symbole « + » et, dans le DiCo, par le symbole « . ». Chaque constituant de FL peut être une FL simple ou une FL complexe. Chacun d'entre eux s'applique directement au mot-clé. Mel'čuk et al. [1995 : 149] en donne la définition et les exemples suivants :

Nous appelons *configuration de fonctions lexicales* une suite de FL simples qui ne sont pas syntaxiquement liées entre elles, mais qui ont le même mot-clé, cette suite ayant une valeur globale cumulative qui exprime de façon indécomposable le sens de la suite entière.

- $\text{Bon}(\textit{joie}) + \text{Magn}(\textit{joie}) = \textit{paradisique}$: « une grande [Magn] joie qui est très bonne [Bon] »
- $\text{Magn}(\textit{maladie}) + \text{A}_1(\textit{maladie}) = \textit{terrassé}$ [par ART \sim] : « qui a [A₁] une maladie grave [Magn] »
- $\text{Magn}(\textit{orage}) + \text{Caus}_2\text{Func}_0(\textit{orage}\#\text{II}) = \textit{déchaîner}$ [ART \sim] : « X cause exister [Caus₂Func₀] un grand [Magn] orage au sujet de X »

Selon Mel'čuk et al. [1995 : 148], seules les FLSS sont dites régulières, dans le sens où elles sont considérées comme étant les plus importantes, « autant quantitativement que qualitativement ». Par ailleurs, les FL complexes et les configurations de FL sont mises sur le même plan que les FL non standard, en tant que FL irrégulières. Nous ne partageons pas cette vision des FL. Les FL complexes et les configurations de FL nous semblent tout autant standard que les FLSS : elles remplissent l'ensemble des conditions nécessaires au statut standard et nombre d'entre elles sont, dans le DiCo, aussi fréquentes que les FLSS (c'est-à-dire appliquées à un nombre de mots-clés similaire). Cette façon d'envisager les FL combinées transparaît dans les diverses classifications de FL, qui ne cherchent à représenter et organiser que les FLSS. Nous traitons plus précisément cette question au chapitre 5.

3.3.2.3 Typologie des formules de FL de l'encodage traditionnel

La typologie de l'encodage traditionnel a pour objectif de recenser l'ensemble des types de formules de FL et de les organiser selon les statuts qui leur sont attribués. Elle permet également de souligner la grande hétérogénéité et quelques-unes des faiblesses de cet encodage. La typologie est résumée dans la figure 3.4 (p. 92).

1. FL standard simples

2. FL standard combinées : FL complexes et configurations de FL

Les formules de FL standard simples et combinées (soit les FL complexes et les configurations de FL) viennent tout juste d'être présentées. Nous renvoyons donc le lecteur à la section précédente.

3. FL standard avec indication de phase

Ce troisième groupe concerne les formules de FL standard (simples ou combinées) comportant une indication de phase. Les phases représentent les différentes étapes

d'une action. Ainsi les relations en (10) constituent les deux phases de l'action consistant, pour un accusé, à « réaliser » son accusation, c'est-à-dire à s'en défendre (10a) puis à s'en blanchir (10b). Ces phases sont représentées par des chiffres romains en exposant.

(10) a. $\text{Real}_2^I(\textit{accusation}\#\text{I.2}) = \textit{se défendre}$ [contre ART \sim];

b. $\text{Real}_2^{II}(\textit{accusation}\#\text{I.2}) = \textit{se blanchir, se disculper, se laver}$ [de ART \sim].

De la même façon, la FL Real_\circ associée à la lexie $\text{CRIME}\#\text{I.A}$ (11) se décompose en trois étapes :

(11) a. $\text{Real}_\circ^I(\textit{crime}\#\text{I.a}) = \textit{instruire}$ [ART \sim], $\textit{enquêter}$ [sur ART \sim];

b. $\text{Real}_\circ^{II}(\textit{crime}\#\text{I.a}) = \textit{élucider}$ [ART \sim], $\textit{trouver l'auteur}$ [de ART \sim];

c. $\text{Real}_\circ^{III}(\textit{crime}\#\text{I.a}) = \textit{punir, châtier, réprimer, venger}$ [ART \sim].

4. FL standard avec pointeurs vers la définition

Le quatrième type de FL standard regroupe un ensemble de FL comportant des pointeurs vers les composantes sémantiques de la définition de leur mot-clé. Ces FL sont révélatrices du lien profond existant entre les définitions des unités lexicales et les relations de dérivation sémantique ou de collocation qui leur sont associées (décrit notamment par Kahane [2003b], Iordanskaja et Polguère [2005]). Nous distinguons deux sous-ensembles parmi ces FL, selon le type de renvoi qui est fait vers la définition.

Premièrement, certaines FL comportent des indications sur la quantité, la durée, le caractère ponctuel ou habituel d'un événement dénoté par une unité lexicale et son dérivé sémantique ou collocatif. Dans les formules de FL, ces indications sont représentées par les notations *quant*, *temp*, *usual* et *actual* en exposant. La première sert à activer la composante de pluralité potentielle de la définition d'une lexie. On voit en (12) que le sens d'intensité de *Magn* ne porte pas sur l'intensité du coup de feu lui-même, mais sur la quantité de coups de feu tirés. En conservant la même FL, tout en remplaçant *quant* par *temp* en (13), on déplace l'intensité sur la longueur de l'action.

(12) $\text{Magn}^{\text{quant}}(\text{coup de feu}\#\text{I}) = \text{nourris, soutenus};$

(13) $\text{Magn}^{\text{temp}}(\text{admirer}\#\text{II}) = \text{longuement, pendant des heures}.$

Les deux autres éléments **usual** et **actual** servent à indiquer respectivement si la valeur est habituelle ou si, au contraire, elle est ponctuelle. La relation en (14) indique que la valeur doit être exprimée pour dénoter un événement habituel et non un événement actuel ou ponctuel. À la question *Où est Pierre ?*, par exemple, on ne pourra répondre : *Il fréquente le cinéma*. En revanche, **actual** dans l'exemple (15) indique que l'action se déroule actuellement et ne doit pas être interprétée comme quelque chose d'habituel, auquel cas, on aurait comme valeurs : *ne pas avoir d'imagination, être terre à terre, etc.*

(14) $\text{Real}_1^{\text{usual}}(\text{cinéma}\#\text{I.3}) = \text{fréquenter} [\text{ART } \sim];$

(15) $\text{NonOper}_1^{\text{actual}}(\text{imagination}\#\text{2}) = \text{être } \ulcorner \text{à court} \urcorner, \text{être } \ulcorner \text{en panne} \urcorner [\text{d}'\sim].$

Le second type de pointeurs dans les formules de FL renvoie à des composantes précises de la définition. On remarque, par exemple, qu'en (16a) et (16b), les notations en indice renvoient respectivement aux composantes 'motif' et 'manière' de la définition de **ACCUSER**#I.1 et apportent une précision supplémentaire à la FL **Ver**. Ces notations forment avec elle deux FL, qui doivent être respectivement interprétées comme \rightarrow De façon justifiée et \rightarrow De façon explicite :

(16) a. $\text{Ver}_{\text{motif}}(\text{accuser}\#\text{I.1}) = \ulcorner \text{à raison} \urcorner, \text{justement};$

b. $\text{Ver}_{\text{manière}}(\text{accuser}\#\text{I.1}) = \text{ouvertement, } \ulcorner \text{tout haut} \urcorner.$

Suivant le même schéma, les indices des exemples (17b) et (17a) fonctionnent comme des pointeurs vers les composantes de sens 'origine' et 'destination' de la lexie **CANALISATION**, dont la structure actantielle est la suivante : \sim POUR LE déplacement X DU liquide OU gaz Y [DE l'endroit Z À l'endroit W].

(17) a. $\text{Involv}_{\text{origine}}(\text{canalisation}) = \text{partir} [\text{de } \text{N}=\text{Z}];$

b. $\text{Involv}_{\text{destination}}(\text{canalisation}) = \text{se déverser} [\text{dans } \text{N}=\text{W}].$

5. FL semi-standard

Les FL semi-standard recouvrent sans doute l'ensemble le plus hétérogène de formules de FL. Dans tous les cas, les éléments ajoutés aux FL standard servent à apporter une précision sémantique qui n'est pas représentée par la FL standard. Ces notations renvoient elles aussi aux composantes de la définition de la lexie auxquelles elles s'appliquent. Cependant, le processus se fait non plus en mentionnant le type de composante comme plus haut, mais en précisant une information sur cette composante. Dans les exemples suivants, on trouve des précisions sur la manière (18), l'aspect (19), la localisation (20), la finalité (21), le nombre (22), la conséquence (23), et la nature des actants (24).

(18) complètement $\text{AntiReal}_1(\textit{armoire}) = \textit{vider}$ [ART \sim de N=Y];

(19) (temporairement) $\text{LiquFunc}_0(\textit{conversation}\#a) = \textit{interrompre}$ [ART \sim];

(20) en cour $\text{Fact}_2(\textit{avocat}(1)\#I) = \textit{plaider}$ [pour N=Y];

(21) pour frapper qqn. $\text{Labreal}_{1@}(\textit{ceinture}\#I.1) = \textit{battre}$ [N avec ART \sim];

(22) plusieurs $\text{Caus}_2\text{Func}_0(\textit{témoignage}\#I.1) = \textit{collecter}$ [ART \sim s];

(23) issue défavorable $\text{Real}_2(\textit{chance}\#2) = \textit{tenter en vain}$ [A-poss \sim];

(24) X et Y étant plutôt des enfants, $\text{Caus}_1\text{Oper}_1(\textit{ami}\#I.1) = \textit{faire, se faire}$ [\sim avec N=Y | sans modificateur].

La ressemblance entre ce type de FL et le précédent entraîne parfois quelques confusions. En (22), par exemple, on pourrait substituer à la FL existante, la formule $\text{Magn}^{\text{quant}}.\text{Caus}_2\text{Func}_0$, la première section de la formule de FL venant justement préciser que l'on cherche à obtenir plusieurs témoignages.

6. FL localement standard

Les FL localement standard sont encodées dans la langue de description selon des formules normalisées. En voici deux exemples :

(25) De_nouveauOper₁₂(goût) = reprendre [~ à qqch.] (→ avoir de nouveau);

(26) Essayer_de Oper₂(augmentation#2) = demander, réclamer [ART ~].

7. FL non standard

Les FL non standard font office de petites définitions des relations lexicales (cf. exemples (27), (28), et (29)). Elles sont traditionnellement encodées par des gloses formulées dans la langue d'étude. Elles sont rédigées dans un métalangage visant à éviter les problèmes de synonymie ou d'ambiguïté. Les lexies choisies doivent être les moins ambiguës et les moins idiomatiques possibles, au risque de produire des formules stylistiquement lourdes. Par exemple, l'expression *destiné à* est retenue au détriment de *pour*, cette dernière étant jugée trop polysémique.

(27) [Qqch./Qqn.] tendre à montrer que ~ est justifiée(*accusation#I.1*) = corroborer, étayer, soutenir [ART ~];

(28) Chair de ~ utilisée comme aliment(*agneau#a*) = agneau#b;

(29) Qui a lieu devant les tribunaux(*bataille#II.1*) = juridique, de procédure.

8. Les lexicalisations de régime

Nous n'avons pas encore mentionné un type particulier de relation qui se situe entre une relation lexicale et la lexicalisation d'un actant du régime d'une unité lexicale. Nous les appelons *lexicalisations de régime*. Il s'agit de relations spécifiant un ou plusieurs actants du mot-clé. Elles expriment, sur le plan conceptuel, des sous-types de l'unité lexicale en mot-clé. La valeur de ces lexicalisations de régime varie en fonction des types d'actants spécifiés. Par exemple, pour les lexies CEINTURE#II (30) et ENCOURAGEMENT#2 (31), dont les structures actantielles sont les suivantes :

CEINTURE#II : ~ OÙ L'ON TROUVE X ET QUI SE TROUVE AUTOUR DE Y

ENCOURAGEMENT#2 : LE fait X EST UN ~ POUR LA personne Y À FAIRE

Z(Y),

on remarque des formules servant à préciser la nature du premier actant X :

- (30) a. X='de la végétation'(ceinture#II) = *de végétation, verte* ;
 b. X='ensemble d'industries'(ceinture#II) = *industrielle* ;
 c. X='une ou plusieurs voies de circulation importantes'(ceinture#II) = *routi-
 tière* ;
 d. X='voies de chemin de fer'(ceinture#II) = *ferroviaire*.
- (31) a. X='soutien moral'(encouragement#2) = *moral* ;
 b. X='soutien financier'(encouragement#2) = *financier, matériel, pécuniaire*.

On note en (32) que plusieurs actants peuvent être spécifiés en même temps :

- (32) X='client de restaurant', Y='cuisinier de restaurant', Z='bonne qualité d'un re-
 pas'
 (félicitations#b) = « Mes ~ au chef! », « Nos ~ au chef! », « ~ au chef! ».

Ces relations sont l'expression du régime d'un mot-clé qui se voit lexicalisé pour former un syntagme avec ce même mot-clé. Bien qu'il s'agisse de relations lexicales, elles ne correspondent pas à proprement parler à des liens de FL.

Toutefois, il arrive que le mot-clé et la spécification d'un actant soit englobés dans une lexie à part entière. La relation devient alors paradigmatique, comme en (33) pour la lexie CRI#I.1 :

- (33) a. X='âne'(cri#I.1) = *braiment, hi-han* ;
 b. X='bovin'(cri#I.1) = *meuglement, mugissement, meuh* ;
 c. X='canard'(cri#I.1) = *nasillement, coin-coin*.

Lorsque la relation est paradigmatique, la valeur est à la fois en relation d'hyponymie avec le mot-clé (un braiment est un type de cri) et l'expression d'un actant du mot-clé (BRAIMENT indique que le premier actant de CRI#I.1 est un âne). Il semble que ces relations devraient recevoir un double encodage pour refléter l'ambivalence de leur statut, du type : X='âne' Hypo(cri#I.1) = *braiment, hi-han*.

Par ailleurs, il nous semble intéressant de distinguer l'hyponymie classique de ce type particulier d'hyponymie. Nous proposons de représenter ces relations par la

FL Hypo_{*i*}, où l'indice *i* précise l'actant de la lexie spécifié dans la relation. De cette façon, les relations suivantes :

(34) a. X='moyen de transport maritime'(gare#b) = *maritime* ;

b. X='moyen de transport fluvial'(gare#b) = *fluviale* ;

(35) a. Y='soupe'(assiette#a) = *creuse, à soupe* ;

b. Y='nourriture solide'(assiette#a) = *plate*,

se verraient remplacées par les formules :

(34') a. X='moyen de transport maritime' Hypo₁(gare#b) = *maritime* ;

b. X='moyen de transport fluvial' Hypo₁(gare#b) = *fluviale* ;

(35') a. Y='soupe' Hypo₂(assiette#a) = *creuse, à soupe* ;

b. Y='nourriture solide' Hypo₂(assiette#a) = *plate*.

3.3.3 Discussion critique sur l'encodage traditionnel

Nous souhaitons discuter quatre aspects de l'encodage traditionnel, dont certains ont été mis en valeur par la typologie qui vient d'être présentée : la question du dénombrement des FLSS, la granularité de l'encodage traditionnel, ses règles de formation, ainsi que l'absence d'un système pour l'organisation des FL.

3.3.3.1 Dénombrement des fonctions lexicales standard simples

Pour aborder la question du dénombrement des FLSS, nous devons examiner trois facteurs : l'existence (ou non existence) d'un comité chargé de valider les propositions de FLSS, les corpus de référence, ainsi que la granularité de description des relations lexicales.

Premièrement, nous insistons sur le fait que le nombre de FLSS n'est pas fixé dans l'absolu. En effet, l'élaboration d'une liste de FLSS ne se fait pas *a priori* mais de

<p>FL STANDARD</p> <p>1. FL standard simples</p> <p style="padding-left: 2em;">$\text{Oper}_1(\text{aboiment}\#\text{I}) = \text{émettre, pousser, lancer}$ [ART ~]</p> <p style="padding-left: 2em;">$\text{Ver}(\text{boisson}\#\text{1}) = \text{désaltérante, rafraichissante}$</p> <p>2. FL standard combinées</p> <p style="padding-left: 2em;">→ FL complexes</p> <p style="padding-left: 2em;">$\text{IncepPredPlus}(\text{fardeau}\#\text{I}) = \text{alourdir}$ [ART ~]</p> <p style="padding-left: 2em;">→ Configurations de FL</p> <p style="padding-left: 2em;">$\text{AntiBon.Oper}_{12}(\text{compliment}\#\text{a}) = \text{débiter, réciter}$ [ART ~ à N=Y]</p> <p>3. FL standard avec indications de phases</p> <p style="padding-left: 2em;">$\text{Real}_2^{\text{I}}(\text{accusation}\#\text{I.2}) = \text{se défendre}$ [contre ART ~]</p> <p style="padding-left: 2em;">$\text{Real}_2^{\text{II}}(\text{accusation}\#\text{I.2}) = \text{se blanchir, se disculper, se laver}$ [de ART ~]</p> <p>4. FL standard avec pointeurs vers la définition</p> <p style="padding-left: 2em;">$\text{Magn}^{\text{quant}}(\text{coup de feu}\#\text{I}) = \text{nourris, soutenus}$</p> <p style="padding-left: 2em;">$\text{Ver}_{\text{manière}}(\text{accuser}\#\text{I.1}) = \text{ouvertement, } \lceil \text{tout haut} \rceil$</p> <p>FL SEMI-STANDARD</p> <p style="padding-left: 2em;">en cour $\text{Fact}_2(\text{avocat}(1)\#\text{I}) = \text{plaider}$ [pour N=Y]</p> <p style="padding-left: 2em;">pour frapper qq. $\text{Labreal}_{1@}(\text{ceinture}\#\text{I.1}) = \text{battre}$ [N avec ART ~]</p> <p>FL LOCALEMENT STANDARD</p> <p style="padding-left: 2em;">$\text{De_nouveauOper}_1(\text{accusation}\#\text{I.1}) = \text{réitérer, renouveler}$ [ART ~]</p> <p style="padding-left: 2em;">essayer de $\text{Oper}_2(\text{augmentation}\#\text{2}) = \text{demander, réclamer}$ [ART ~]</p> <p>FL NON STANDARD</p> <p style="padding-left: 2em;">Qui a lieu devant les tribunaux($\text{bataille}\#\text{II.1}$) = <i>juridique, de procédure</i></p> <p style="padding-left: 2em;">Chair de ~ utilisée comme aliment($\text{agneau}\#\text{a}$) = <i>agneau#b</i></p> <p>LEXICALISATION DE RÉGIME</p> <p style="padding-left: 2em;">X='de la végétation' $\text{Hypo}_1(\text{ceinture}\#\text{II}) = \text{de végétation, verte}$</p> <p style="padding-left: 2em;">X='ensemble d'industries' $\text{Hypo}_1(\text{ceinture}\#\text{II}) = \text{industrielle}$</p>

Figure 3.4 – Typologie de l'encodage traditionnel

façon empirique, au fur et à mesure de la description du lexique. Nous ne voyons aucun principe théorique qui empêche la création de nouvelles FLSS si celles-ci remplissent les critères nécessaires. Le nombre de FLSS donné dans l'affirmation suivante [Mel'čuk et al., 1995 : 129] :

Les fonctions lexicales standard simples sont au nombre de 56 ; elles couvrent, dans toutes les langues, les relations paradigmatiques entre les lexies (synonymes, antonymes, dérivés, ...), ainsi que leurs relations syntagmatiques (intensificateurs, verbes supports, ...).

doit donc être interprété comme étant fixé à un moment donné, « en synchronie », et non de façon irrévocable. Nous proposons (p. 95) une chronologie sommaire de l'évolution des FLSS afin de mettre en valeur les propositions, ajouts et suppressions de FL depuis leur création. Nous posons comme l'un des premiers inventaires des FLSS, le texte de Zholkovsky et Mel'čuk [1970], qui recense les FL présentées dans la figure 3.5.

Syn	Liqu	Loc	FL candidates :
Conv	Oper	Magn	Mult
Der (dérivés syn- taxiques)	Func	Anti	Sing
Copul	Labor	Gener	Bon
Incep	Fact	Ver	Degrad
Fin	Real	Son	
Caus	S_i	Prepar	
	A_i		

Figure 3.5 – Liste des FL présentées dans Zholkovsky et Mel'čuk [1970]

Notons que nous n'indiquons ni ne détaillons ici l'ensemble des propositions, la plupart d'entre elles sont détaillées dans le chapitre suivant (section 4.2). Le recensement des différentes propositions de FLSS invite à se demander qui statue sur les FL et détermine si une proposition va mener à un ajout. Les listes de référence sont habituellement considérées comme les FL appartenant au DEC ou au DiCo, ou présentes dans les publications de Mel'čuk et Polguère. Cependant, il n'existe pas

de comité de validation des FL. Aussi certaines propositions sont mises en usage dans certaines bases de données lexicales sans recevoir de réelle « validation ». Dans la chronologie que nous présentons, nous distinguons donc s'il s'agit d'une simple proposition ou si la proposition est intégrée à un système et est opératoire. Nous utilisons les termes *proposition*, *ajout*, *usage*, et *suppression*. Notons également que certaines propositions ont été suggérées très longtemps avant qu'elles ne soient intégrées dans les « listes de référence ». C'est le cas, par exemple, de la FL Spec, proposée par Fontenelle [1997b] pour rendre compte de la relation d'hyponymie qui n'a été intégrée au DiCo qu'en 2007 sous le nom de Hypo.

CHRONOLOGIE SOMMAIRE DE L'ÉVOLUTION DE L'ENSEMBLE DES
FONCTIONS LEXICALES STANDARD SIMPLES

– **1966-67**

Premiers inventaires [Zholkovsky et Mel'čuk, 1966, 1967, 1970]

– **1969**

- **PROPOSITIONS** : Erastov [1969]

Matur (→ Commencer à être prêt), Cum (→ Appartenant au même groupe, avec un statut équivalent), Pel (→ Quelque chose qui couvre (et protège L)), Mobil (→ Être en mouvement de façon normale/prototypique), Prim (→ premier/plus important parmi un ensemble de L), Vacant (→ libre/vacant/qui peut être utilisé), Metr (→ paramètres quantifiés typiques), Fas (→ Côté/Partie avant de L) et LocLim (→ Dans les limites de L).

– **1988**

- **AJOUT** : Mel'čuk et Zholkovsky [1988] proposent les fonctions Dimin (→ Diminutif) et Augm (→ Augmentatif) :

Dimin(*dom* 'house.') = *domik* 'a small [and pleasant] house';

Dimin(*ozero* 'lake') = *ozerko* 'a small and pleasant lake'.

Augm(*dom* 'house') = *domisce, domina* 'a huge house';

Augm(*ruka* 'hand') = *rucisca* 'a huge hand'.

Ces FL sont abandonnées dans les publications ultérieures sur les FL.

– **1990**

-**PROPOSITIONS** : Grimes [1990] propose la notion de FL inverses (qui sera détaillée à la section 4.2.2 du chapitre suivant).

– **1993**

- **PROPOSITION** : Gener_p et Gener [Alonso Ramos, 1993] pour la distinction entre les relations d'hyponymie pouvant s'exprimer syntagmatiquement (par exemple, Gener(*amour*) = *sentiment* [de ~]) et celles s'exprimant seulement dans une relation paradigmatique (par exemple, Gener_p(*chien*) = *animal*).

– 1997

- **PROPOSITION** : Spec et Part [Fontenelle, 1997b] pour représenter l'hyponymie et la relation de méronymie fonctionnelle.

– 1999

- **SUPPRESSION** : La FL Nocer (\rightarrow Nuire à \sim) illustrée par les deux exemples ci-dessous, est supprimée dans Mel'čuk et al. [1999] :

Nocer(*moustique*) = *piquer*, Nocer(*voiture*) = *faucher, happer, renverser,*

– 2002

- **PROPOSITIONS** : À l'examen de relations lexicales fréquentes dans un corpus terminologique [Jousse, 2002] :

Méta(*donnée*) = *métadonnée*

Avant(*compiler*) = *précompiler*

Réciprocité(*connexion*) = *interconnexion*

Un seul(*utilisateur*) = *mono-utilisateur*

– 2003

- **AJOUT** : en tant que « localement standard » De_nouveau [Polguère, 2007]

- **PROPOSITIONS** : Loc_{per}(\rightarrow à travers) et Lim (\rightarrow Limite/bord de) [Jousse, 2003]

Loc_{per}(*mer*) = *par* [\sim]

Loc_{per}(*champ*) = *à travers* [\sim]

Lim(*territoire*) = *frontière*

Lim(*forêt*) = *bordure, lisière, orée*

Lim(*tuyau*) = *ouverture*

- **PROPOSITION** : Trait caractéristique [Grizolle, 2003] (cf. section 4.2.3 du chapitre suivant pour plus de détail sur cette proposition)

- **PROPOSITION ET USAGE** : S-INIT (\rightarrow État précédent une action) [Apresjan et al., 2003] :

S-INIT(*wake up*) = *sleep*, S-INIT (*calm down*) = *anxiety*, S-INIT(*stop*)= *movement*

- **PROPOSITION ET USAGE** : ANTI1 et ANTI2 [Apresjan et al., 2003] (ces FL sont présentées p. 99).

- **2005**

- **PROPOSITION** : Loc_{between} [Bélangier et Kittredge, 2005]

Loc_{between}(*bourgeons*) : *entre* [~], // *intergemmaire*

- **2007**

- **AJOUT** : Dans le DiCo, apparition des FL Hypo (→ Hyponyme) et Mero (→ Méronyme) (cf Jousse [2007a])

L'ensemble des FL est resté relativement stable au cours des années. Cependant, il est à noter que certaines propositions intéressantes n'ont pas été prises en compte ni intégrées à la liste, certainement par manque d'un « comité de validation ». Il paraît donc difficile, dans ces conditions, de statuer sur un nombre précis de FLSS.

Nous devons à présent aborder la question du corpus de référence. Les FL proposées dans les listes de référence⁴ n'ont été dégagées que d'après des échantillons de lexique de plusieurs langues. Or, nous pensons que l'importance du corpus de référence, c'est-à-dire à partir duquel on souhaite comptabiliser un nombre de relations (ou de FL), détermine le nombre de FLSS à découvrir. En effet, la recherche de FL évolue au fur et à mesure que le corpus s'accroît : plus on observe de relations et plus il est probable de voir apparaître des régularités parmi ces relations et, éventuellement, de découvrir de nouvelles FL standard.

À ce propos, nous devons citer les travaux de Fontenelle [1997b], consistant en la conversion du dictionnaire bilingue Robert & Collins [Atkins et Duval, 1978] vers une base de données lexico-sémantique. À l'instar de la majorité des dictionnaires bilingues, le Robert & Collins met l'accent sur la combinatoire des unités lexicales

⁴Se reporter à la page 156 pour un relevé des différentes présentations des FL.

et fournit un grand nombre de collocations. Fontenelle constate que beaucoup de dictionnaires se révèlent très peu performants pour mettre en valeur et différencier les collocations, et choisit d'utiliser les FL pour les encoder et les organiser. Ce faisant, il effectue une première application de masse des FL. Tester les FL sur l'ensemble du Robert & Collins représente une tâche considérable jusqu'alors jamais réalisée, qui a permis de confronter ce mode de représentation à un ensemble de données et d'en vérifier la validité pratique.

The general conclusions I draw concerning the imprecise nature of some functions and the evaluation of the material would never have been possible, however, if only a fragment of the dictionary had been analyzed in detail. Exploiting the whole dictionary and focussing on such a large section of the lexicon had never been attempted and the potential applications, whether purely linguistic or pedagogical, could only have been hinted at if I had decided to focus on a few dozen items and their semantic environment instead of taking into account the whole metalinguistic apparatus of the dictionary. [Fontenelle, 1997b : 102]

Dans la base de données du Robert & Collins [Fontenelle, 1997b], près de 70 000 liens ont pu être encodés par des fonctions standard et plus de 18 000 liens, dont certains sont pourtant très fréquents, sont restés sans description. D'après l'analyse d'un nombre si conséquent de relations, Fontenelle suggère la création de nouvelles FL⁵. De notre point de vue, ce n'est qu'en étendant la description des relations lexicales à l'échelle du lexique d'une langue, que l'on pourra véritablement déterminer un nombre de FLSS.

On ne peut non plus traiter la question du nombre de FL sans aborder la notion de la granularité dans la description des relations lexicales. La difficulté principale, lorsque l'on cherche à décrire le sens d'une relation lexicale, est d'opter pour un degré de précision ou de granularité idéal. Pour que l'on puisse parler de FL standard, la granularité d'une description ne doit pas être trop restreinte, c'est-à-dire qu'elle doit être suffisamment large pour concerner un grand ensemble de lexies. Ceci est aussi valable pour le cas contraire, à savoir, pour éviter la description trop générale

⁵Certaines de ces propositions sont présentées au chapitre 4, portant sur les FL non standard.

d'une relation. Le défi est donc de trouver le bon équilibre entre un degré trop fin ou trop large de granularité. Considérons les exemples en (36). Ils représentent tous une relation entre une entité et un produit qui en est dérivé. On pourrait choisir de regrouper les liens entre les noms d'animaux et les plats préparés avec leur chair. Dans ce cas, la FL choisie reflèterait une granularité fine qui exclurait les relations entre ÉLÉPHANT et IVOIRE, ou entre PAIN et TARTINE, etc. Au contraire, choisir une description plus large comme *Produit dérivé*, par exemple, permet de les regrouper toutes sous une seule et même formule.

- (36) a. Foie de \sim produit par gavage(*canard*) = *foie gras* ;
 b. Tranche de \sim enduite de N(*pain*) = *beurrée, tartine* ;
 c. Friandise contenant beaucoup de \sim (*sucré*) = *sucrerie* ;
 d. Plat à base de \sim (*agneau*) = *blanquette, rôti, méchoui* ;
 e. Boisson alcoolisée faite avec \sim (*cerise*) = *cherry, kirsch, guignolet* ;
 f. Plat à base de viande de \sim (*coq*) = *au vin* ;
 g. Matériau que l'on obtient à partir des défenses de \sim (*éléphant*) = *ivoire* ;
 h. Aliment préparé dérivé de \sim (*lait*) = *fromage, laitage, beurre*.

La description des relations lexicales doit faire face à cette question de granularité. En effet, si l'on considère les FL standard et semi-standard, on remarque qu'elles sont situées sur un continuum, du moins granulaire vers le plus granulaire. Cependant, cette granularité ne semble pas être traitée de façon homogène sur l'ensemble des relations lexicales. La cause en est assez évidente puisque, dès lors qu'on s'intéresse de près à une relation, on va vouloir en accentuer la granularité. Ainsi, dans le cadre du projet ETAP-3 (cf. annexe II, p. xlix), Apresjan et al. [2003 : 10] ont dégagé deux types d'antonymes qui diffèrent par la position de la négation dans leur représentation sémantique.

The first type of antonyms is based on inner negation. Consider the well-known pair *begin* and *stop*. *To stop doing P* can be explicated as 'to begin not to do P'. This type of antonymy is called ANTI1. The second type of antonyms is based on outer negation. Consider the pair

observe (the regulations) and *violate* (the regulations); *to violate* means 'not to observe'. This type of antonymy is called ANTI2. [...]

À partir de ces deux types de négation, on peut formuler deux règles de paraphrasage distinctes :

1. $X \Leftrightarrow \text{Neg} + \text{ANTI2}(X)$:

He violated the regulations - He did not observe the regulations ;

2. $X+Y \Leftrightarrow \text{ANTI1}(X) + \text{ANTI2}(X)$:

He began to observe the regulations - He stopped violating the regulations.

Il semble que les problèmes de gestion de la granularité dans les descriptions sont responsables d'une démultiplication de FL, qui contribue à disperser des relations similaires et à complexifier leur encodage. Pour remettre à niveau l'ensemble des relations, une solution serait de déterminer plusieurs niveaux de granularité. Un niveau plus large permettrait d'englober les relations similaires, et un niveau plus précis, de les différencier. Le modèle de FL que nous proposons d'élaborer dans le chapitre 5 réalisera une description à plusieurs niveaux en répartissant les FL dans des grandes classes sémantiques, et permettra de regrouper les différents types de FL sous le même plan.

La question du nombre de FLSS ne peut donc être résolue sans prendre en compte les trois facteurs qui viennent d'être présentés. Le nombre de FLSS dépend tout d'abord des critères qu'utilise la personne qui le détermine. Par ailleurs, tant que les relations lexicales du lexique entier d'une langue n'auront pas été couvertes, il semble difficile de prétendre à un nombre définitif de FLSS. Enfin, le choix de la granularité a une grande incidence sur la démultiplication des FL et la scission d'une FL potentiellement standard en plusieurs FL semi-standard ou non standard.

3.3.3.2 Granularité de l’encodage traditionnel

Devant l’hétérogénéité de l’encodage traditionnel (et notamment celui des FL semi-standard), il est légitime de questionner la pertinence du degré de granularité choisi pour décrire les relations lexicales. On observe plusieurs situations :

1. Premièrement, le degré de granularité n’est pas systématiquement respecté dans la description des différents liens lexicaux relevant d’une relation donnée. Par exemple, dans les quatre exemples suivants, on observe que (37) et (38) sont distinguées par la présence de l’élément semi-standard **rapidement** en (38), alors que la valeur *manger* en (39) n’est pas différenciée de *avaler* et *s’enfiler*, qui auraient pu être décrites par la FL **rapidement** Real_1 . Le lien lexical en (40) aurait également pu être encodé par la FL **rapidement** Real_1 .

(37) $\text{Real}_1(\text{bière}(1)\#b) = \textit{boire, consommer, prendre}$ [ART \sim];

(38) **rapidement** $\text{Real}_1(\text{bière}(1)\#b) = \textit{s’enfiler, siffler, caler}$ [ART \sim];

(39) $\text{Real}_1(\text{sandwich}\#\text{II}) = \textit{avaler, s’enfiler, grignoter, manger}$ [ART \sim];

(40) [Individu X] **abaisser et relever rapidement** $\sim(\textit{paupière}) = \textit{battre, ciller, cligner}$ [des \sim s].

2. On observe des cas où la granularité choisie pour décrire une relation lexicale est trop lâche, c’est-à-dire que la relation encodée par la FL est si vague que les valeurs retournées sont trop nombreuses et trop variées. Dans ce cas, il est nécessaire d’apporter un élément de sens supplémentaire pour distinguer les relations. Par exemple, il semble difficile de regrouper les actions typiques réalisées par un chat sous un seul encodage. On utilise alors les notations semi-standard pour les différencier :

$\text{Fact}_{0\text{mouvement}}(\textit{chat}) = \ulcorner \textit{faire le gros dos} \urcorner; \textit{s’étirer};$

$\text{Fact}_{0\text{nettoyage}}(\textit{chat}) = \textit{faire sa toilette};$

$\text{Fact}_{0\text{gratter}}(\textit{chat}) = \textit{faire ses griffes};$

$\text{Fact}_{0\text{chasser}}(\textit{chat}) = \textit{chasser}$ [N];

$\text{Fact}_{0\text{boire}}(\textit{chat}) = \textit{laper}.$

Cependant, ceci est gênant, dans la mesure où dans ce contexte, la FL standard Fact_0 , est tellement générale qu'elle s'applique à des valeurs n'ayant plus vraiment de parenté sémantique. Or, si les valeurs sont si différentes sur le plan sémantique, la validité de la fonction doit être réévaluée. Il semble que la sous-spécification de certaines fonctions soit un indice d'une insuffisance dans leur définition sémantique profonde ou d'une sur-utilisation des FL. Il semble que Fact_0 soit ici employée en dehors de son domaine de définition.

3. À l'inverse, on remarque que la volonté de précision amène souvent les rédacteurs du DiCo à sur-spécifier la description des liens pour souligner des différences sémantiques entre les valeurs d'une FL, ou pour préciser certaines composantes de sens apportées à une FL. Cela se traduit par l'existence de FL semi-standard. Or, si les éléments semi-standard apportent des précisions sémantiques non prises en charge par la fonction standard, ils contribuent également à donner un caractère brouillon à l'encodage : les fonctions paraissent différentes, alors qu'elles partagent toutes la même FL standard, tel qu'on peut le voir en (41) avec les différents types de Real_1 . Ceci rend l'extraction et la manipulation des relations difficiles.

- (41) a. pour éclairer $\text{Real}_1(\textit{allumette}) = \textit{brûler}$ [ART \sim];
 b. rapidement $\text{Real}_1(\textit{bière}(1)\#b) = \textit{siffler}$ [ART \sim];
 c. destinée à fermer un passage Y $\text{Real}_1(\textit{chaîne}\#I) = \textit{tendre}$ [ART \sim « en travers de » N=Y];
 d. \sim étant parure $\text{Real}_1(\textit{couronne}\#I.2) = \textit{porter}$ [ART \sim];
 e. Qui est un symbole de respect pour le mort N $\text{Real}_1(\textit{couronne}\#I.2) = \textit{déposer}$ [ART \sim];
 f. pour subsistance $\text{Real}_1(\textit{plume}\#III.a) = \textit{vivre}$ [de A-poss=X \sim];
 g. En tant que carrière $\text{Real}_1(\textit{théâtre}\#I.1) = \textit{faire}$ [du \sim].

S'il est nécessaire d'ajouter aussi fréquemment des notations aux FL, c'est peut-être que l'encodage n'a pas été suffisamment bien conçu en amont et n'a

pas, dans tous les cas, prévu la gestion de différents niveaux de granularité.

3.3.3.3 Règles de formation de l'encodage traditionnel

Le troisième aspect que nous souhaitons aborder est l'absence de véritables règles de formation de l'encodage traditionnel. Nous avons pu le constater à travers les nombreux exemples de FL semi et non standard cités dans le chapitre.

1. La structuration des formules de FL semi-standard

Les FL semi-standard forment un ensemble très hétérogène. Les notations accompagnant les FL standard ne sont pas rédigées selon un patron défini et ne respectent aucune norme d'écriture. Nous pensons qu'il serait utile d'établir une grammaire de formation des FL semi-standard, qui, d'une part, identifie le type de composantes sur lequel on apporte une précision (manière, conséquence, finalité, etc.), et, d'autre part, permette de régulariser la syntaxe des notations.

2. La structuration des formules de FL non standard

Les fonctions non standard, quant à elles, sont encodées selon un métalangage naturel qui met l'accent sur la lisibilité et la clarté de la formulation. Ceci a pour conséquence l'existence de variantes synonymiques, comme on peut le constater dans les trois exemples suivants :

Discipline qui étudie $\sim(\textit{médicament}) = \textit{pharmacologie}$;

Science qui étudie $\sim(\textit{poisson}) = \textit{ichtyologie}$;

Étude de $\sim(\textit{signe}) = \textit{astrologie}$.

Tel quel, cet encodage est très hétérogène et inexploitable pour la maintenance des BLEC et les applications en TAL : il n'est ni repérable, ni analysable de façon automatique. Nous traitons en détail le problème des FL non standard dans le chapitre 4, où nous cherchons à déterminer des stratégies de normalisation de l'encodage des FL non standard, dans le but de les homogénéiser et de faire apparaître certaines régularités sémantiques.

3.3.3.4 Absence d'un système pour l'organisation des FL

Le quatrième problème que nous souhaitons mettre en avant est l'absence d'un système d'organisation pour les FL. Celles-ci formalisent plusieurs aspects d'une relation lexicale : le sens de la relation (notée **sens** dans la figure 3.6), la partie du discours de la valeur (**pos**), le régime syntaxique de la valeur (**reg.**), la nature paradigmaticque ou syntagmaticque de la relation (**rel.**). À partir de cet encodage, il est donc possible de dégager des sortes de fiches d'identité des relations lexicales, comme l'illustrent les quatre exemples de la figure 3.6 extraites de Jousse [2007b].

<p>S_1 ex : $S_1(\text{vendre}\#1) = \text{vendeur}$ sens : 'premier actant du mot-clé' pos : nom rel : paradigmaticque reg. : même régime que le mot-clé moins le premier actant</p>	<p>S_{loc} ex : $S_{loc}(\text{mouton}\#I.a) = \text{bergerie}$ sens : 'endroit typique où se trouve \sim' pos : nom rel : paradigmaticque reg. : –</p>
<p>$IncepReal_1$ ex : $IncepReal_1(\text{gare}\#a) = \text{entrer} [\text{en } \sim]$ sens : 'commencer à réaliser l'action inhérente à \sim' pos : verbe rel. : syntagmaticque reg. : sujet : premier actant de \sim / objet : \sim</p>	<p>V_0 ex : $V_0(\text{haine}) = \text{haïr}$ sens : 'équivalent verbal de \sim' pos : verbe rel. : paradigmaticque reg. : mêmes actants (X et Y) que \sim</p>

Figure 3.6 – Fiches d'identité des FL S_1 , S_{loc} , $IncepReal_1$ et V_0

Or, cette encapsulation pose problème dès lors que l'on vise l'extraction des éléments encodés de façon indépendante. Le regroupement d'informations de niveaux différents dans un seul encodage contribue à séparer des FL sémantiquement équivalentes, mais relevant de patrons syntaxiques différents, ou syntaxiquement équivalentes, mais avec des sens différents. Le DiCo et les autres bases de données construites avec les FL ne procèdent pas à des regroupements de FL en fonction de ces aspects. Les FL y sont généralement présentées en une simple liste organisée en fonction des différentes parties du discours et du statut des fonctions. Nous

consacrons le chapitre 5 à l'élaboration d'un système de FL permettant plusieurs types d'organisation des FL. Ce système permettra d'effectuer des recherches sur les relations lexicales selon différents critères.

Nous venons de présenter un certain nombre de lacunes de l'encodage traditionnel. La plupart de ces lacunes résultent du caractère relativement flou de cet encodage, tant du point de vue conceptuel que formel. Cette observation a amené Kahane et Polguère [2001] à poser la définition et les fondements d'un encodage formel des FL standard à travers la proposition de deux nouveaux encodages : l'encodage explicite et l'encodage algébrique. Nous les présentons dans la section suivante.

3.3.4 L'encodage explicite et l'encodage algébrique

Afin de pallier certaines lacunes de l'encodage traditionnel, Kahane et Polguère [2001] cherchent à donner une définition formelle des FL standard à travers deux encodages, dont la vocation est de décrire le sens et la combinatoire de chaque FLSS, et de mettre en valeur les propriétés formelles des FLSS, c'est-à-dire les opérations de combinaison naturelles auxquelles elles peuvent se soumettre.

L'encodage explicite

L'encodage explicite expose les valeurs et rôles profonds des FL dans le langage naturel, et facilite la compréhension et l'utilisation des FL aux non-initiés de la théorie Sens-Texte. Il décrit les FL au moyen de deux formules distinctes : l'encodage de son contenu sémantique et l'encodage de sa structure syntaxique.

Le contenu sémantique est représenté par une configuration de la relation prédicat-argument entre des primitifs sémantiques, qui correspondent à des FL déjà identifiées par la théorie Sens-Texte. Celles-ci sont accompagnées de notations formelles, qui réfèrent à des sens spécifiques : « # » représente le sens du mot-clé, « 1, 2, 3, ... » renvoient aux premier, deuxième, et troisième actants sémantiques du mot-clé, et « Ω », à d'autres participants sémantiques non-spécifiés. Par exemple, pour la lexie COLÈRE :

Caus[2/3,#] (\rightarrow Y/Z cause la colère de X) : Y/Z *met* X en colère ;
 Symp[#, 'poings de X'] (\rightarrow la colère de X est symptomatisée par les
 poings de X) : X *sert* les poings de colère.

La structure syntaxique est exprimée au moyen de deux types d'information sur les valeurs de la fonction : leur partie du discours et leur diathèse, c'est-à-dire la liste de leurs dépendants syntaxiques, par ordre croissant d'oblicité. Les trois exemples de la figure 3.7 donnent une illustration de l'encodage explicite complet pour les FL Func_1 , Oper_1 et Oper_2 . Le premier rang représente le contenu sémantique, le second, le contenu syntaxique. Les trois verbes étant des verbes support, soit sémantiquement « vides », le contenu sémantique de la FL est donc équivalent à celui du mot clé : « # » (cf. figure 3.7).

$$\begin{aligned} \text{Func}_1 &: \left[\begin{array}{c} \# \\ V[\#, 1] \end{array} \right] (\text{colère}) = \text{habiter} [\text{N} = \text{X}] \\ \text{Oper}_1 &: \left[\begin{array}{c} \# \\ V[1, \#] \end{array} \right] (\text{colère}) = \text{éprouver} [\text{ART} \sim] \\ \text{Oper}_2 &: \left[\begin{array}{c} \# \\ V[2, \#] \end{array} \right] (\text{colère}) = \text{encourir} [\text{ART} \sim] \end{aligned}$$

Figure 3.7 – Illustration de l'encodage explicite

L'encodage algébrique

L'encodage algébrique est, quant à lui, très proche de l'encodage traditionnel. Il est fondé sur des formules linéaires unidimensionnelles, qui synthétisent à la fois le contenu sémantique et la structure syntaxique des FL. L'encodage algébrique est « un métalangage avec lequel le lexicographe et le linguiste peuvent penser » [Kahane et Polguère, 2001 : 14], c'est-à-dire tout à fait adapté pour les lexicographes, puisque les formules sont très proches de l'expression en langage naturel. Par exemple, la formule $\text{Caus.Non.Manif}(L)$ doit être comprise comme : 'causer que L ne se manifeste pas'.

Ce métalangage doit être défini en amont de l'encodage explicite et non l'inverse,

le passage du premier au second étant trivial : il consiste à remplacer les simples FL par leurs définitions explicites, et à calculer les opérations.

Opérations sur les FL

Il existe des relations entre les FL, que l'on peut modéliser par deux opérations : le produit et la fusion. Notons que ces deux opérations ne sont pas représentées de façon explicite dans l'encodage traditionnel. Le produit est le mode de combinaison de FL le plus productif. Kahane et Polguère [2001 : 15] le présentent comme suit.

Le produit h de deux FL syntagmatiques f et g a les propriétés suivantes :

1. $h(L)$ est un collocatif de L ,
2. $h(L)$ est une paraphrase de $g \circ f(L) + f(L)$, où $+$ est le symbole d'union linguistique et \circ le symbole de composition.

Par exemple, la FL IncepOper_1 appliquée à l'unité lexicale AMOUREUX retourne la valeur *tomber*, qui est un collocatif de *amoureux*.

$g \circ f(L)$ représente la composition de g et de f , c'est-à-dire l'application de g à f . Si l'application de Oper_1 à *amoureux* retourne la valeur *être* [\sim], l'application de Incep à $\text{Oper}_1(\text{amoureux})$ (soit *être*) retourne la valeur *commencer* : *commencer à être amoureux*.

La combinaison de $\text{Incep} \circ \text{Oper}_1(\text{amoureux})$ et $\text{Oper}_1(\text{amoureux})$: *commencer à être amoureux* est synonyme de *tomber amoureux*. Par conséquent, $\text{IncepOper}_1(\text{amoureux})$ est une paraphrase de $\text{Incep} \circ \text{Oper}_1(\text{amoureux}) + \text{Oper}_1(\text{amoureux})$.

Dans l'encodage algébrique, le produit de deux fonctions est noté par un point, soit, par exemple : $\text{Incep} \cdot \text{Oper}_1$. Dans l'encodage explicite, c'est le résultat du produit qui est représenté dans une structure similaire à celles de la figure 3.7. Pour l'interpréter, notons que $c(f)$ est le contenu sémantique de f , $d(f)$ la diathèse de f , et $\text{pos}(f)$, la partie du discours de f . Ainsi, le produit de la FL Incep et d'une FL f , telle que Oper_1 , est défini par la formule de l'encodage explicite suivante :

$$\text{Incep.f} : \begin{bmatrix} \text{Incep}[c(f)] \\ \text{pos}(f)[d(f)] \end{bmatrix}$$

Prenons un exemple pour déchiffrer cet encodage :

$$\text{IncepOper}_1(\text{mutisme}\#a) = s'enfoncer, tomber \text{ [dans ART } \sim].$$

Remplaçons f par Oper_1 , dont l'encodage explicite est présenté plus haut dans la figure 3.7. On peut maintenant interpréter l'encodage explicite de IncepOper_1 de la façon suivante : le contenu sémantique de IncepOper_1 est égal à $\text{Incep}(\#)$ (soit ' $\#$ commence'), étant donné que le contenu sémantique de Oper_1 est égal à celui du mot-clé : $\#$. La partie du discours est égale à celle de Oper_1 , soit un verbe, et la diathèse à celle de Oper_1 , soit : $[1, \#]$. On peut donc attribuer à IncepOper_1 l'encodage explicite suivant :

$$\text{IncepOper}_1 : \begin{bmatrix} \text{Incep}[\#] \\ V[1, \#] \end{bmatrix}$$

La seconde opération effectuée sur les FL est la fusion. Elle vise à expliciter le lien qui peut exister entre des relations paradigmatiques et syntagmatiques. Nous avons déjà évoqué ce point au chapitre précédent (p. 31). La fusion permet d'associer à chaque FL syntagmatique f , une relation paradigmatique notée $//f$, où $//$ représente la fusion et $//f(L)$ est une paraphrase de $L+f(L)$. La FL Magn , par exemple, est une FL syntagmatique :

$$\text{Magn}(\text{pluie}) = \text{forte, diluvienne, torrentielle.}$$

On peut lui associer la relation $//\text{Magn}$ qui retourne des valeurs en relation paradigmatique avec le mot-clé L , et qui contiennent à la fois le sens de L et celui de $\text{Magn}(L)$. On dira, en effet, qu'un déluge est une très forte (= $\text{Magn}(L)$) pluie (= L) :

$$//\text{Magn}(\text{pluie}) = \text{déluge.}$$

La fusion entraîne la disparition du mot-clé dans la structure syntaxique de f . Ainsi, pour la lexie ZIGZAG , la valeur de Oper_1 est : *faire un zigzag*, alors que la valeur de $//\text{Oper}_1$ est *zigzaguer*, qui comprend le sens de ZIGZAG :

$$\begin{aligned}\text{Oper}_1(\text{zigzag}) &= \text{faire} [\text{ART-indéf } \sim], \\ //\text{Oper}_1(\text{zigzag}) &= \text{zigzaguer}.\end{aligned}$$

Dans l'encodage explicite, Oper_1 et $//\text{Oper}_1$ sont représentées comme suit :

$$\text{Oper}_1 : \begin{bmatrix} [\#] \\ V[1, \#] \end{bmatrix}, //\text{Oper}_1 : \begin{bmatrix} [\#] \\ V[1] \end{bmatrix}$$

L'objectif de ces deux encodages est de fournir une définition formelle satisfaisante de l'encodage des FL tout en proposant un formalisme facilement calculable et pouvant entrer en compte dans les applications comme la traduction automatique et la génération de phrases. L'encodage explicite est tout à fait approprié pour ces tâches, dans la mesure où il est entièrement défini sur les plans sémantique et syntaxique, et permet la définition des opérations formelles de produit et de fusion. Notons également que ces différents types d'encodage sont reliés entre eux. Comme l'ont montré Kahane et Polguère [2001], il est possible, à partir de l'encodage algébrique, de générer les gloses de l'encodage vulgarisé.

3.4 Conclusion

Ce troisième chapitre était destiné à dresser un bilan sur les FL. Nous avons, dans un premier temps, présenté la notion de FL dépouillée de son encodage, en formulant les critères qui permettent de la définir. Par la suite, nous avons montré la façon dont la notion de FL se matérialise à travers différents encodages, établis pour répondre à des objectifs particuliers : lexicographie formelle, traitement automatique et vulgarisation. Nous avons porté une attention particulière à l'encodage traditionnel, le plus majoritairement utilisé, et tenté d'en établir une typologie, afin de visualiser et d'organiser l'ensemble des formules qu'il peut revêtir. Ceci nous a permis de mettre en avant un certain nombre de problèmes dans l'encodage traditionnel : la détermination d'un nombre de FLSS, la gestion de la granularité dans

la description des liens lexicaux, l'absence de règles pour la formation de l'encodage, ainsi que le manque d'organisation des FL en fonction des différents types d'information qu'elles encapsulent.

L'absence de définition des formules complexes de l'encodage traditionnel a conduit Kahane et Polguère [2001] à proposer deux encodages formels : l'encodage explicite et l'encodage algébrique. Ces encodages sont entièrement définis au niveau de leur sémantique et de leur syntaxe. Notons que nous utilisons l'encodage explicite au chapitre 5 pour établir la liste des différents patrons syntaxiques associés aux FL. Plusieurs problèmes liés aux FL ont été évoqués dans ce chapitre, dont l'hétérogénéité de l'encodage des FL non standard. Nous consacrons le chapitre suivant à la normalisation de leur encodage et à la définition de stratégies pour repérer les analogies parmi les relations lexicales. La normalisation des FL non standard va leur permettre de s'intégrer au système plus global de FL, que nous présentons dans le chapitre 5.

Chapitre 4

Normalisation de l’encodage des fonctions lexicales non standard

4.1 Introduction

À la suite de la présentation des FL, le travail exposé dans ce chapitre vise la recherche de régularités sémantiques parmi un ensemble de relations lexicales représentées par des FL dites non standard et le traitement homogénéisé de leur encodage. Commençons par un bref rappel de ce qui a été présenté au chapitre précédent. Les FL ont un caractère graduel et s’organisent autour de quatre statuts :

1. Les **FL standard** sont construites à partir d’un noyau d’une soixantaine de FL simples [Mel’čuk et al., 1995 : 125]. Elles doivent répondre aux principes de cardinalité, de diversité et d’universalité cités à la section 3.2.2.1. Elles peuvent être composées d’un seul élément (42) ou combinées entre elles (43).

(42) $\text{Anti}(\textit{joyeux}) = \textit{triste}$ (\rightarrow antonyme) ;

(43) $\text{IncepPredPlus}(\textit{abcès}) = \textit{mûrir}$ (\rightarrow commencer à devenir plus important).

2. Les **FL semi-standard** sont constituées d’une FL standard et d’une composante sémantique non prise en charge par la FL standard. Par exemple, la description de la relation (44) ayant le sens ‘causer le silence de qqn’, se différencie de la relation (45) par la présence de la composante sémantique ‘en échange de qqch.’, qui vient compléter la FL standard.

(44) $\text{CausOper}_1(\textit{silence}) = \textit{réduire}$ [qqn. au \sim] ;

(45) en échange de qqch. $\text{CausOper}_1(\textit{silence}) = \textit{acheter}$ [le \sim de qqn.].

3. Les **FL localement standard** sont des FL dont l'universalité n'a pas encore été prouvée dans toutes les langues, mais qui remplissent les autres conditions nécessaires pour être considérées comme de véritables FL. Par exemple, en français, les FL *De_nouveau* (46) et *Essayer de* (47).

(46) *De_nouveau* $Oper_{12}(goût) = reprendre$ [\sim à qqch.] (\rightarrow avoir de nouveau);

(47) *Essayer de* $Oper_2(augmentation\#2) = demander, réclamer$ [ART \sim].

4. Les **FL non standard** (FLNS), quant à elles, encodent des dérivations sémantiques et des collocations, cf. (48) et (49), mais ne satisfont aucun critère requis pour être standard.

(48) *Qui est liée à des irrégularités commises par X* (*faillite*) = *frauduleuse* ;

(49) *Produite à l'étranger*(*bière*) = *d'importation*.

Ce sont les FL non standard (désormais *FLNS*) qui font l'objet de ce chapitre, nous exposons ci-dessous l'enjeu de leur normalisation.

4.1.1 Enjeu d'une normalisation de l'encodage des FLNS

Comme nous l'avons mentionné au chapitre précédent et présenté en détail à l'annexe II, les FL standard ont déjà été largement utilisées dans les applications en traitement automatique [Apresjan et al., 2000], que ce soit pour des programmes de paraphrasage et génération de texte [Nasr, 1996, Lareau, 2002], de résumés automatiques [Bélanger et Kittredge, 2005], de traduction automatique [Apresjan et al., 2003], ou encore pour développer des outils d'apprentissage des langues [Diachenko, 2006, Boguslavsky et al., 2006]. Cependant, les FL utilisées dans ces projets ne couvrent qu'une partie des dérivations sémantiques et des collocations recensées. Les FLNS ne peuvent pas entrer en ligne de compte dans les traitements automatiques en raison de l'hétérogénéité de leur encodage. Elles sont, en effet, encodées selon un métalangage naturel qui met l'accent sur la lisibilité et la clarté de la formulation. Par conséquent, cet encodage est très hétérogène et difficilement exploitable pour des applications en TAL : on ne peut en repérer les similitudes

ni aisément l'analyser de façon automatique. La distinction de plusieurs statuts pour les FL et la différence d'encodage qui en résulte conduit à mettre à l'écart les FLNS, qui constituent pourtant une part non négligeable des relations représentées dans le DiCo, soit environ 16%.

La position de Mel'čuk et al. [1995 : 150] à l'égard des FLNS est assez radicale :

Elles ne sont pas prévisibles et ne peuvent donc pas être dégagées et recensées de façon méthodique. La seule manière de les inventorier, c'est de les retrouver dans les textes. En plus, pour les mêmes raisons, on ne peut pas les décrire à l'aide d'un formalisme prédéterminé ; au contraire, le sens de chacune doit être formulé dans le DEC au coup par coup.

Notre position vis-à-vis des FLNS va à l'encontre de cette affirmation et postule qu'il existe, au contraire, un grand nombre de relations lexicales non encodées par des FL standard, que ces relations sont, du moins en partie, prévisibles et qu'elles mériteraient, non pas d'être décrites au coup par coup, mais d'être formulées selon un encodage normalisé qui mette en valeur leurs régularités.

Nous avons également questionné, dans le chapitre précédent, le fondement de la distinction entre les FL standard et non standard. Mise à part la condition d'universalité, Polguère [2007 : 46] propose d'assigner le statut standard aux FL satisfaisant les deux conditions suivantes :

- La **condition de cardinalité** stipule qu'une FL doit s'appliquer à un nombre important d'unités lexicales et non être limitée aux unités lexicales d'un seul champ sémantique. La relation lexicale suivante, par exemple : **Qui ne vit pas à l'état sauvage**(*canard*) = *d'élevage, domestique*, ne concerne que certains noms d'animaux et ne pourra donc pas être représentée par une FL standard.
- La **condition de diversité** impose une diversité parmi les valeurs retournées par les FL. Ainsi, les valeurs de la FL **Bon** (\rightarrow **qualificatif positif**) sont diversifiées selon les mots-clé auxquels elle s'applique, par exemple :

Bon(*aspiration*) = *noble* ;

Bon(*bijou*) = *somptueux* ;
 Bon(*déjeuner*) = *succulent*.

Cette FL remplit donc ce critère pour être considérée comme une FL standard.

Ces conditions offrent des critères opératoires pour la description des relations, mais ne se révèlent pas toujours suffisantes pour déterminer si une relation doit être encodée par une FL standard ou une FLNS. Selon Polguère [2007 : 48-50], la distinction entre standard et non standard est une affaire de choix :

Standardness¹ is not presented as being an inherent property of some LFs but the result of an estimation made by “people in charge”. [...] The notion of standard LF can be precisely defined and it is only the criteria for standardness that are vague and depend on subjective estimates. This leads Mel’čuk [1996 : 91] to say :
 “(...) it is immaterial — in both practice and theory — whether we decide to treat an LF as Standard or Non-standard.”

Ces citations viennent appuyer notre propos sur le manque de rigueur et de systématisme dans le choix de l’un ou l’autre des deux statuts. Le caractère flou de ces critères tient également au choix de la granularité choisie pour le regroupement des liens lexicaux. Si l’on opte pour une granularité large, on aura tendance à regrouper sous un même encodage un grand nombre de liens lexicaux. En revanche, si l’on ressert la granularité, on verra apparaître des distinctions sémantiques entre ces liens, ce qui conduira à l’attribution d’encodages distincts pour représenter les liens étudiés. Nous ne souhaitons pas remettre en question la distinction entre les statuts standard et non standard. Nous tenterons cependant, tout au long du chapitre, de confronter certaines FLNS régulières aux critères de standardité afin de proposer d’éventuelles candidates au statut standard.

Nous devons préciser à présent deux objectifs tout à fait distincts de notre travail : la standardisation et la normalisation des FLNS.

¹Par commodité, nous créons le néologisme *standardité* pour traduire le terme anglais *standardness*.

- La **standardisation** consiste à créer de nouvelles FL standard, ce qui nécessite de confronter les candidates aux critères de standardité et de les valider ou, au contraire, de les rejeter. La standardisation conduit également à effectuer des regroupements de liens lexicaux auparavant isolés.
- La **normalisation** consiste à la fois à modifier la granularité de l’encodage des FLNS pour regrouper des liens lexicaux similaires et à proposer des règles pour la formalisation de l’encodage des FLNS, sans pour autant les faire passer au statut standard.

Gardons bien cette distinction à l’esprit car, tout au long du chapitre, il sera question tantôt de normaliser les FLNS et tantôt de les standardiser. À ces deux objectifs vient s’ajouter un troisième :

- il s’agit de définir des stratégies, d’une part, pour la recherche d’analogies sémantico-syntaxiques parmi les liens lexicaux et, d’autre part, pour gérer l’évolution de la description de ces liens lexicaux. Le corpus que nous utilisons ne couvre pas l’ensemble du lexique. Nous devons, par conséquent, prévoir de nouveaux liens lexicaux qui ne sont pas encore décrits actuellement.

4.1.2 Présentation du corpus de travail

Le travail de normalisation porte sur une partie des FLNS du DiCo : ses 820 FLNS nominales et ses 727 FLNS adjectivales. Les relations à valeurs nominales sont presque exclusivement des relations de dérivation sémantique, par exemple, *poisson* → *ichtyologie* ; *gâteau* → *farine*, etc.

Les relations adjectivales, quant à elles, peuvent être réparties selon deux types : d’une part, un petit nombre de dérivés sémantiques (qui très souvent prennent la forme de dérivés morphologiques), du type *barbe* → *barbu* ; *bagarre* → *bagarreur*_{Adj} ; *soupçon* → *méfiant* ; *sévérité* → *sévère*, *strict*, etc., d’autre part, de façon majoritaire, des relations collocationnelles du type *barbe* → \lceil *poivre et sel* \rceil ; *bagarre* → *générale* ; *soupçon* → *léger*, etc. Nous traiterons ces types de relation de façon différente.

On remarque que si l’hétérogénéité parmi les formules des FLNS nominales réside

principalement dans le choix du lexique employé et les problèmes de synonymie et d’ambiguïté qui en découlent, l’hétérogénéité des FLNS adjectivales tient davantage à la syntaxe de l’encodage. Il n’existe pas de patrons d’encodage préétablis. Ainsi, tandis que la FLNS en (50a) est amorcée par un pronom relatif, celle en (50b) débute directement par un participe passé ; de la même façon, les trois FLNS adjectivales (50c), (50d) et (50e) sont formulées selon trois structures syntaxiques différentes.

- (50) a. Qui est commis par des soldats pendant une guerre(*crime*) = *de guerre* ;
 b. Commis par Adam et Eve(*péché*) = *originel* ;
 c. Provoqué volontairement à des fins thérapeutiques(*abcès*) = *de fixation* ;
 d. ~ des pirates(*drapeau*) = *noir* ;
 e. Tel que Z est la France(*drapeau*) = *tricolore*.

4.1.3 Organisation du chapitre

La recherche de nouvelles FL et la volonté de normaliser l’encodage des FLNS n’est pas une problématique nouvelle. Certaines méthodes de recherche ont déjà été proposées, nous les présentons à la section 4.2. Nous consacrons la section 4.3 à la présentation de regroupements de FLNS effectués de façon empirique, sans stratégie particulière. Certains de ces regroupements de FLNS peuvent donner lieu à la création de FL uniques pouvant être considérées comme de bonnes candidates au statut standard, d’autres ne seront que des FLNS normalisées. Toutefois, même si la simple analyse des données du corpus permet de dégager des similarités entre relations, nous pensons qu’il est également nécessaire d’élaborer des stratégies, à la fois pour le repérage des régularités sémantiques, et pour la régularisation de leur encodage. Nous avons élaboré deux méthodologies qui permettront, d’une part, de définir plus précisément la nature et l’origine des dérivations sémantiques et des collocations, d’autre part, de fournir un cadre et une méthode de travail pour l’encodage de ces relations et leur normalisation. Premièrement, on remarque une forte corrélation entre les FL et le phénomène de dérivation morphologique. Nous exploitons ce lien entre FL et dérivés morphologiques dans la section 4.4. La grande

majorité des FL correspondent à certains types de dérivés morphologiques des langues de plusieurs familles linguistiques. Notons toutefois que la piste de la morphologie dérivationnelle se révèle davantage valable pour les dérivés sémantiques. En ce qui concerne plus particulièrement les collocations (mais non exclusivement), nous proposons une seconde méthodologie. Nous exposons, dans la section 4.5, une stratégie permettant de mieux déterminer la nature des relations lexicales et de les organiser selon des attributs typiques liés aux unités lexicales. Par exemple, à la lexie HORLOGE sont associés des attributs de forme et de fonctionnement, qui peuvent être exprimés au moyen de collocations : par exemple, *murale*, *à carillon*, *à l'heure*, etc. Normaliser les FLNS en fonction de ces attributs va permettre de les identifier et de les organiser de façon plus pertinente dans les bases de données lexicales.

4.2 Travaux antérieurs

Plusieurs recherches ont été menées pour proposer de nouvelles FL standard et tenter de normaliser les FLNS. Erastov [1969], Grimes [1990] et Fontenelle [1997b] ont proposé quelques nouvelles FL d'après des régularités observées parmi un ensemble de relations lexicales. Frawley [1998], L'Homme [2002], Jousse [2002], Jousse et Bouveret [2003] et Bouveret [2006] ont postulé quelques adaptations des FL aux relations terminologiques. Grizolle [2003] et Jousse [2003] ont réfléchi à des moyens d'homogénéiser l'encodage de certaines FLNS, Popovic [2004] a travaillé sur l'homogénéisation des gloses de vulgarisation des FL standard. Cependant, la question de l'encodage des FLNS dans sa globalité n'a pas encore été abordée. Sans détailler l'ensemble de ces travaux, nous en présentons quelques-uns pour illustrer les trois grandes approches qui ont conduit à la recherche de régularités : une approche empirique, une approche logique ainsi qu'une approche par champs sémantiques.

4.2.1 Approche empirique

L'approche empirique consiste à proposer de nouvelles FL à partir de l'observation des liens lexicaux. Il ne s'agit pas d'une méthode particulière ni d'une stratégie à suivre, mais plutôt d'une démarche visant à dégager des analogies d'après la simple observation d'un ensemble de liens. Cette méthode est la plus communément utilisée pour proposer de nouvelles FL, c'est également cette approche que nous suivrons dans la section 4.3.1. Nous illustrons cette démarche à l'aide des propositions de Erastov [1969], dégagées au fur et à mesure de l'encodage de relations lexicales récurrentes dans la langue, mais non couvertes par les fonctions existantes. Nous en reproduisons quelques-unes ci-dessous :

- Matur (\rightarrow Commencer à être prêt) :
 $Matur(\textit{fruit}) = \textit{to ripen}$; $Matur(\textit{cheese}) = \textit{to age, to mature}$;
- Cum (\rightarrow Appartenant au même groupe, avec un statut équivalent) :
 $Cum(\textit{citizen}) = \textit{compatriot, fellow countryman}$; $Cum([\textit{team}] \textit{member}) = \textit{teammate}$;
- Pel (\rightarrow Quelque chose qui couvre (et protège L)) :
 $Pel(\textit{bean}) = \textit{pod}$; $Pel(\textit{egg}) = \textit{shell}$; $Pel(\textit{book}) = \textit{cover}$; $\textit{binding}$;
- Mobil (\rightarrow Être en mouvement de façon normale/prototypique) :
 $Mobil(\textit{water}) = \textit{to flow, to pour, to run}$; $Mobil(\textit{bird}) = \textit{to fly}$;
- Prim (\rightarrow premier/plus important parmi un ensemble de L) :
 $Prim(\textit{entrance}) = \textit{main}$; $Prim(\textit{door}) = \textit{front}$; $Prim(\textit{city}) = \textit{//capital}$;
- Vacant (\rightarrow Libre/vacant/qui peut être utilisé) :
 $Vacant(\textit{ressources}) = \textit{available}$; $Vacant(\textit{seat}) = \textit{free, vacant}$;

- Fas (\rightarrow Partie avant de L) :
 Fas(*house*) = *front, frontage* [of ART \sim]; Fas(*ship*) = *bow, prow* [of ART \sim];
- LocLim (\rightarrow Aux limites de L) :
 LocLim(*country*) = *within the borders* [of ART \sim].

Polguère [2007] émet quelques réticences à l'égard de certaines de ces fonctions : **Matur**, premièrement, ne semble pas satisfaire la condition de cardinalité puisqu'elle ne concerne *grosso modo* que les lexies qui dénotent des entités comestibles. **Mobil** peut être considérée comme un cas de **Fact₀** (\rightarrow Action typique de \sim). **Vacant** est un cas spécifique de **A₂AntiReal₁** (\rightarrow Dont la fonction n'est pas remplie). Les autres propositions semblent intéressantes, mais n'ont pas été retenues dans le DEC ni dans le DiCo.

Un désavantage de l'approche empirique est qu'elle considère des nouvelles FL de façon isolée. Aussi, même si elle offre la possibilité de découvrir de nouvelles fonctions, elle ne se révèle pas toujours pertinente si les FL ne sont pas créées avec une volonté d'intégration au sein d'un système plus général de FL. Selon Polguère [2007 : 57] :

There should be, at the same time, a strong commitment to continuing to consider the set of standard LFs as a system, with strict theoretical foundations, and therefore to carefully examine how LF candidates can find their place in this system. Without such discipline, the modeling tool of standard LFs will soon become a huge mess. It will lose its elegance and, eventually, its usefulness.

Ce principe d'intégration à un système de FL nous semble très important. Nous l'érigions en critère pour la validation des FL que nous proposons en 4.3.1.

4.2.2 Approche logique

L'approche logique consiste à partir d'une observation précise du fonctionnement des FL afin d'en inférer logiquement de nouvelles. Voyons l'exemple des FL in-

verses : suivant une approche logique ou théorique, Grimes [1990] propose la notion de fonction inverse en faisant remarquer qu'un sous-ensemble de fonctions pourrait logiquement se voir associer des fonctions inverses. Par exemple, à la fonction *Result* (\rightarrow Action résultant de \sim) (cf. *Result(to buy) = to have, to possess*), il est possible d'associer une fonction inversant la relation. Ainsi, les valeurs de cette FL *Result*⁻¹ appliquée à la lexie TO HAVE seraient l'ensemble des lexies pour lesquelles la fonction *Result* produit *to have* comme valeur, soit :

$$\text{Result}^{-1}(\textit{to have/to possess}) = \textit{to get, to obtain, to steal, to take, ...}$$

De même, à partir de la FL *Liqu* (\rightarrow Causer que \sim n'existe plus) on pourrait logiquement inférer une FL *State* pour inverser la relation :

$$(51) \textit{Liquidative (mistake) = erase} (\rightarrow \textit{Causer la fin de } \sim);$$

$$(52) \textit{State (erase) = mistake, error} (\rightarrow \textit{Ce dont on cause la fin}).$$

Cette réflexion conduit à la création de nouvelles paires de FL fonctionnant selon le principe d'inversion, comme les FL *Antecedent* et *Consequent* :

$$\textit{Antecedent(shut[event]) = open [state]} (\rightarrow \textit{Action antécédente});$$

$$\textit{Consequent (open[state]) = shut [event]} (\rightarrow \textit{Action conséquente}).$$

Un des avantages de l'approche logique est qu'elle permet de visualiser les FL comme formant un système à partir duquel il est possible d'inférer de nouvelles FL qui sembleraient y occuper légitimement une place. Nous reprenons le concept de FL inverse dans le chapitre 6 et montrons comment il est possible de s'appuyer sur certaines paires de FL pour inférer l'encodage de certaines relations lexicales.

4.2.3 Approche par champ sémantique

La troisième approche consiste à envisager des régularités en traitant les relations associées à des lexies de même champ sémantique, par exemple, les relations des lexies d'animaux, de sentiment, etc. Grizolle [2003] a effectué une recherche dans le but d'établir une classification de FLNS et FL semi-standard en effectuant des regroupements de lexies partageant les étiquettes sémantiques **animal** et **artefact**. L'objectif de ce travail est de dégager, parmi toutes les FL associées à ces lexies,

des classes sémantiques de FL et d'en déduire une normalisation en comparant les différents encodages. Pour les FL des lexies subordonnées à l'étiquette **animal**, vingt-deux classes ont été créées (cf. fig. 4.1).

Élément de la lignée de ~	Connotation liée à ~
[~] d'une espèce particulière	Adjectif tiré d'une connotation nominale
Espèce particulière de ~	Équipement pour ~
Particularité physique de ~	Nom propre de ~
Trait distinctif de ~	Action ou activité d'un individu impliquant ~
Élément du corps de ~	Activité professionnelle liée à ~
Nourriture de ~	Personne dont l'activité professionnelle est liée à ~
Particularité comportementale/du mode de vie de ~	Artefact conçu par X pour attraper/tuer/se protéger de ~
Quelque chose produit par ~	Habitat de ~
[~] utilisé/conditionné pour (faire) quelque chose	Caractérisation du lieu où l'on trouve ~
Produit dérivé de ~	
Excrément de ~	

Figure 4.1 – Classes sémantiques de FL associées aux lexies étiquetées **animal**

À partir de ces classes sémantiques, les FLNS et FL semi-standard de toutes les lexies d'animaux ont été comparées et homogénéisées. Par exemple, les FL de la classe sémantique « trait distinctif de ~ » ont été régularisées comme suit (cf. tab. 4.I).

Lexie	Ancienne FLNS	Nouvelle FLNS	Valeur
AGNEAU	Pelage de l'A.	Trait distinctif de l'A. : pelage	toison ; laine
CHAT	Caractéristique de la langue des C.	Trait distinctif du C. : langue	râpeuse
ÉLÉPHANT	Caractéristique physique de la peau des É.	Trait distinctif de l'É. : peau	peau rugueuse
LION	Poils couvrants la tête d'un L.	Trait distinctif du L. : pelage de la tête	crinière
TIGRE	Trait distinctif du pelage d'un T.	Trait distinctif du T. : pelage	pelage rayé

Tableau 4.I – FLNS liées aux lexies d'animaux avant et après normalisation

Cette approche offre une méthodologie cadrée et organisée et se révèle notamment très efficace pour la révision des encodages de FLNS. Cependant, elle ne permet pas de tenir compte des régularités de relations entre champs sémantiques. En effet,

les relations similaires sont observées à l'intérieur d'un champ sémantique donné. Or, comme nous l'avons vu à la page 113, les FL sont soumises à la condition de cardinalité, c'est-à-dire qu'elles doivent être suffisamment larges pour s'appliquer aux lexies relevant de plusieurs champs sémantiques. Ainsi, la FLNS *Élément du corps de* \sim que l'on retrouve dans les articles des lexies d'animaux pourrait être fondue avec la FLNS *Partie de* \sim très fréquente dans les articles des lexies d'artefact. Il est donc nécessaire de confronter les FLNS associées à un champ sémantique donné à celles d'autres champs sémantiques pour effectuer d'éventuelles fusions.

4.3 Propositions d'après l'observation des FLNS du DiCo

La première vague de propositions que nous présentons dans ce chapitre suit l'approche empirique. Nous avons étudié les FLNS nominales et adjectivales du DiCo et relevé un certain nombre de similitudes entre des liens lexicaux encodés par des formules de FLNS. Toutefois, ces analogies entre liens lexicaux ne nous ont pas semblé équivalentes au niveau de leur fréquence et de la diversité de leurs valeurs. Par conséquent, nous les avons réparties en deux ensembles. Le premier ensemble regroupe des relations paraissant de bonnes candidates pour être encodées par des FL standard, elles sont traitées dans la section 4.3.1. Le deuxième ensemble, présenté dans la section 4.3.2, comporte des propositions de normalisation de l'encodage de FLNS présentant de très fortes similitudes sans toutefois être candidates au statut standard.

4.3.1 FLNS candidates au statut standard

Nous envisageons dans cette section la création de cinq FL standard : *Hypo*, *Mero*, *Hollo*, *Compos* et *Produit dérivé*. Chacune d'entre elles sera confrontée aux critères requis pour le statut standard, soit le critère de cardinalité, selon lequel une FL doit pouvoir s'appliquer à un grand ensemble de lexies ; le critère de diversité, qui impose une diversité parmi les valeurs des FL, ainsi que le critère d'intégrabilité de

la FL au système général de FL, qui préconise qu'une FL nouvelle ne doit pas être isolée mais se trouver en relation logique avec d'autres FL existantes.

4.3.1.1 La relation d'hyponymie : Hypo

Il existe une FL modélisant une relation sémantique très proche de l'hyponymie. Il s'agit de la FL **Gener**. Elle détermine, pour une lexie L_1 , un mot générique pouvant apparaître dans la construction L_1, L_2 et autres **Gener**(L), soit par exemple : *poire, pomme et autres fruits*. La relation inverse de **Gener** n'existe pas. Grimes [1990] puis Fontenelle [1997b] ont proposé la création de la fonction **Spec** sans que celle-ci soit pourtant retenue. Récemment, nous avons suggéré la modélisation de l'hyponymie dans le DiCo par la fonction **Hypo** pour remplacer des FLNS comportant de très fortes similarités, comme illustré par les exemples (53) ci-dessous :

- (53) a. Type de $\sim(\text{gâteau}) = \text{clafoutis}$;
 b. Espèce de $\sim(\text{canard}) = \text{colvert}$;
 c. Variété de $\sim(\text{cerise}) = \text{griotte}$.

Notons que la FL **Hypo** ne doit pas être considérée comme la fonction inverse de **Gener** dans la mesure où cette dernière est plus spécifique que la relation lexicale d'hyponymie.

Observons les conditions pour la standardité :

- La **condition de cardinalité** est tout à fait remplie car la FL potentielle concerne l'ensemble des lexies nominales et verbales.
- La **condition de diversité** semble également satisfaite dans la mesure où les valeurs sont très différentes selon les lexies auxquelles la FL est appliquée.
- Concernant enfin l'**intégration au système de FL**, **Hypo** s'oppose à la FL **Gener** sans en être pour autant l'exacte inverse.

4.3.1.2 Les relations de méronymie et de holonymie : Mero et Holo

Bien qu'elle soit connue et représentée dans de nombreuses bases de données lexicales, la relation de méronymie n'était pas normalisée ni standardisée dans le DiCo.

Cette fonction a déjà été proposée par Fontenelle [1997b] sous la forme **Part** mais rejetée par Mel'čuk sous prétexte de sa nature davantage encyclopédique que linguistique. Nous avons suggéré l'intégration dans le DiCo de la fonction notée **Mero**, modélisant la relation partie-tout (plus précisément partie constituante-tout). Toutefois, les relations méronymiques encodées, c'est-à-dire entre un nom d'entité et les noms de ses parties constituantes, doivent conserver un statut linguistique. En d'autres termes, **Mero(L)** doit se définir par 'partie de L ...'. Par exemple, les lexies **BRAS**, **JAMBE** et **TÊTE** sont des méronymes de **CORPS**, dans la mesure où elles incluent **CORPS** dans leur définition : *bras de X / jambe de X / tête de X* = 'partie du corps de X ...'. En revanche, la lexie **NEZ** doit être considérée comme une partie du visage plutôt que comme une partie du corps. Selon ce principe, la lexie **ROUE** n'est pas considérée comme un méronyme de **AVION**, car elle ne se définit pas par rapport à **AVION**, mais plus largement par rapport à un véhicule. En revanche, la lexie \lceil TRAIN D'ATTERRISAGE \rceil inclut nécessairement **AVION** dans sa définition, elle est donc considérée comme un de ses méronymes.

La relation de méronymie peut couvrir, selon l'approche envisagée, un certain nombre de sous-types. Nous considérons ici la méronymie comme une relation partie constituante-tout. Les relations ingrédient-entité et matériau-entité, seront représentées par une autre FL proposée à la section suivante. La FLNS **Mero** permet de remplacer un certain nombre de FLNS très similaires, comme illustré en (54) ci-dessous :

- (54) a. Partie de \sim (*épingle*) = *tête, tige, pointe* ;
 b. Partie de \sim que l'on remplit d'eau(*évier*) = *bac, cuvette* ;
 c. Élément du corps de \sim (*abeille*) = *dard, ...*

La relation inverse, l'holonymie (illustrée en (55)) n'est pas non plus représentée par une FL standard dans le DiCo. Nous suggérons la création de la FL **Holo**.

- (55) Entité géologique dont fait partie \sim (*plaque#I.3*) = *terre#III*, \lceil *écorce terrestre* \rceil ,
lithosphère.

Confrontons maintenant ces FL potentielles aux conditions requises pour le statut standard.

- La **condition de cardinalité** est remplie. Ces deux relations sont considérées dans la plupart des approches ontologiques comme des relations fondamentales au même titre que l’hyponymie et l’hyponymie. Elles sont très fréquentes et touchent une partie très importante du lexique.
- La **condition de diversité** est, elle aussi, largement satisfaite dans la mesure où les valeurs sont presque aussi variées que les lexies auxquelles s’appliquent les deux FL.
- Au niveau de l’**intégration au système de FL**, ces deux FL fonctionnent comme une paire de FL inverses, ce qui motive d’autant plus leur création.

Ces deux FL nous semblent donc de très bonnes candidates au statut standard.

4.3.1.3 Les relations ‘matériau — artefact’ et ‘ingrédient — artefact’ et la relation ‘X — produit réalisé à partir de X

Les relations ‘matériau typique-artefact’ et ‘ingrédient typique-artefact’ sont parfois considérées comme de la méronymie. On retrouve dans WordNet [Fellbaum, 1998], par exemple, trois types de méronymie : member meronymy (*association* → *associate*), substance meronymy (*steel* → *iron*) et part meronymy (*table* → *leg*). La FL Mero, que nous venons de présenter, représente le troisième type. Nous proposons la création de la fonction Compos pour encoder les relations ‘matériau typique-artefact’ (56) et ‘ingrédient typique-artefact’(57), qui correspondent au deuxième type de méronymie proposé dans WordNet.

- (56) a. Matériau typique dont sont faites les $\sim(\textit{assiette}) = \textit{porcelaine}$;
 b. Matériau dont on fait des $\sim(\textit{bijou}) = \textit{pierre, argent, or}$;
 c. Matériau pour $\sim(\textit{corde}) = \textit{lin, coton, nylon}$;
 d. Matériau dont sont faits les $\sim(\textit{pneu}) = \textit{caoutchouc}$;
 e. Matériau précieux dont peut être fait la $\sim(\textit{vaisselle}) = \textit{argent, porcelaine}$;
- (57) a. Ingrédient utilisé pour faire $\sim(\textit{bière}) = \textit{blé, houblon, malt}$;

- b. Substance alimentaire avec laquelle on fait $\sim(\textit{pain}) = \textit{farine}$;
- c. Ingrédient avec lequel on fait $\sim(\textit{yaourt}) = \textit{lait}$.

Nous pensons que ces deux relations sont suffisamment proches sémantiquement pour être décrites par un seul et même encodage.

Il existe une relation inverse aux deux qui viennent d'être présentées : 'X - produit réalisé à partir de X'. Celle-ci représente ce que l'on fait à partir d'une entité, que cette entité soit concrète ou abstraite. On peut en observer quelques exemples en (58).

- (58) a. Foie de \sim produit par gavage($\textit{canard}\#\text{I.b}$) = $\lceil \textit{foie gras} \rceil$;
- b. Gâteau fait avec $\sim(\textit{cerise}) = \textit{clafoutis}$;
 - c. Boisson alcoolisée faite avec $\sim(\textit{cerise}) = \textit{cherry, kirsch}$;
 - d. Friandise contenant beaucoup de $\sim(\textit{sucre}\#\text{a}) = \textit{sucrerie}$;
 - e. Plat à base de $\sim(\textit{agneau}\#\text{b}) = \textit{blanquette, méchoui}$;
 - f. Aliment préparé dérivé de $\sim(\textit{lait}\#\text{b}) = \textit{fromage, beurre}$.

Nous proposons de créer la fonction **Produit dérivé** pour rendre compte de cette relation.

Observons à présent les conditions déterminant si les deux fonctions **Compos** et **Produit dérivé** que nous proposons peuvent être considérées comme standard :

- La **condition de cardinalité** est remplie dans la mesure où ces relations concernent au moins un grand ensemble de lexies dénotant des entités physiques.
- La **condition de diversité** au vu des exemples semble également satisfaite.
- Concernant l'**intégration au système de FL**, ces deux FL forment une paire de FL inverses, ce qui motive leur création et en fait de bonnes candidates au statut standard.

4.3.2 FLNS normalisées

Les FLNS que nous présentons ci-dessous n'ont pas pour ambition d'être standard, soit que le critère de cardinalité n'est pas rempli, c'est-à-dire que ces relations ne concernent pas un nombre suffisamment important d'unités lexicales, soit que le critère de diversité a échoué. De plus, les formules normalisées que nous proposons ne s'intègrent pas forcément au système de FL. Nous présentons, à titre d'illustration de la démarche empirique, cinq formules d'encodage normalisées : **Metr**, **Marque**, **Document associé**, **Axe** et **Vêtement typique**. Notons bien que les formules proposées ont pour objectif de regrouper des relations similaires. Il s'ensuit donc inévitablement une généralisation et, par conséquent, une perte de précision dans la définition du sens de la relation. Par conséquent, les anciennes FLNS seront conservées à titre de gloses pour conserver une explicitation maximale des relations encodées.

4.3.2.1 Les relations de mesure

D'après l'observation des FLNS, nous avons dégagé un ensemble de relations partageant toutes une idée de mesure. Ces relations associent un nom d'entité (ou de fait) et le nom que porte un paramètre de mesure lorsqu'il est associé à cette entité. Ainsi, le paramètre *taille* qui se mesure en mètres, pieds, etc. devient **POINTURE** s'il s'applique à des chaussures ou des gants, **FORMAT** si l'on parle d'une feuille de papier, **HAUTEUR** s'il s'agit d'un immeuble, **ALTITUDE** pour une montagne, etc. Nous souhaitons créer une famille de FLNS pour regrouper et encoder l'ensemble de ces relations de mesure de façon homogène. Nous nous inspirons de la FL **Metr** proposée par Eratsov [1969]. Il s'agit d'une fonction nominale associant un nom à un ensemble de noms signifiant 'paramètres quantifiés typiques', par exemple :

$\text{Metr}(\textit{room}) = \textit{dimension, size}$ [of ART \sim];

$\text{Metr}(\textit{attack}) = \textit{scale}$ [of ART \sim];

$\text{Metr}(\textit{inflation}) = \textit{rate}$ [of \sim].

Nous proposons d'ajouter à la formule **Metr** une notation précisant quel paramètre mesuré est en cause, par exemple, la mesure de qualité : **Metr**_{qualité} ou de capacité :

$\text{Metr}_{\text{capacit }} \cdot$ Cet encodage permet une indexation coh rente des diff rentes relations de mesure. Nous l'illustrons ci-dessous   l'aide d'un ensemble d'exemples. Dans ces exemples, les FL $\text{Metr}_{\text{param tre}}$ sont destin es   remplacer les FLNS cit es.

$\text{Metr}_{\text{dimension}}$: Nom de la mesure de dimension

Dimension des balles que l'on utilise avec $\sim(\text{fusil}(1)\#a) = \text{calibre}$;

Mesure de la taille de $\sim(\text{bateau}) = \text{calaison, d placement, } \lceil \text{tirant d'eau} \rceil$;

Taille de $\sim(\text{enveloppe}\#II) = \text{dimension, format, taille}$.

$\text{Metr}_{\text{capacit }}$: Nom de la mesure de capacit 

Unit  de mesure de la capacit  en Y de $\sim(\text{clinique}\#1) = \text{lit}$;

Unit  de mesure de la capacit  en Y de $\sim(\text{restaurant}) = \text{couvert}$.

$\text{Metr}_{\text{dur e}}$: Nom de la mesure de dur e

Longueur moyenne de \sim dans une population de X($\text{vie}\#I.1$) = *esp rance*.

$\text{Metr}_{\text{qualit }}$: Nom de la mesure de qualit 

Qualit  de \sim de X($\text{conversation}\#b$) = *niveau* ;

Param tres selon lesquels on  value \sim dans la soci t ($\text{vie}\#I.2$) = *conditions, niveau, qualit *.

$\text{Metr}_{\text{r sultat}}$: Nom de la mesure de r sultat

Points obtenus par X/Y dans $\sim(\text{partie}(2)) = \text{score, r sultat}$.

$\text{Metr}_{\text{vitesse}}$: Nom de la mesure de vitesse

Unit  de mesure pour la vitesse de $\sim(\text{bateau}) = \text{n ud}$;

Unit  de mesure pour la vitesse de $\sim(\text{navire}) = \text{n ud}$.

$\text{Metr}_{\text{courant}}$: Nom de la mesure de courant/d bit

Mesure du courant de $\sim(\text{fleuve}\#I) = \text{d bit}$;

Vitesse d' coulement de Y par $\sim(\text{robinet}) = \text{d bit}$.

Cette liste n'est pas exhaustive, d'autres types de mesure pourraient s'ajouter en respectant le mod le propos .

4.3.2.2 Marque

Nous proposons la FLNS *Marque* pour regrouper les relations entre un nom d'entité et le nom de la marque laissée par cette entité. Les exemples ci-dessous (59) montrent que cette relation s'exprime au moyen de substantifs diversifiés et propres aux unités lexicales en mot-clé. Cependant, si la condition de diversité semble satisfaite, ce n'est pas le cas pour les deux autres conditions. Le nombre de lexies concernées est relativement restreint et la FL ne s'intègre pas avec d'autres FL standard. Nous proposons donc un encodage normalisé : la formule *Marque*. Rappelons que les anciennes formules de FLNS serviront de gloses pour préciser le sens de la relation.

- (59) a. Dépôt qui se forme au fond d'un contenant de $\sim(\text{bière}(1)\#a) = \text{lie}$;
 b. Trace laissée à la surface de l'eau par \sim en déplacement(*navire*) = *sillage* ;
 c. Séquelle de $\sim(\text{plaie}\#I) = \text{cicatrice, balafre}$;
 d. Marque laissée dans le sol par un $\sim(\text{pneu}) = \text{empreinte, trace}$.

4.3.2.3 Document associé

La relation que nous souhaitons encoder ici unit un fait et un document témoignant de ce fait, comme illustré par les exemples en (60). Pour cette relation, on observe une fois de plus que les valeurs sont diversifiées et très dépendantes des lexies auxquelles la relation s'applique. La condition de cardinalité semble être plus difficilement satisfaite. Enfin, pour ce qui est du critère d'intégration, cette FL reste relativement isolée des autres FL existantes. Cependant, il nous paraît utile de mettre en valeur cette relation lexicale et d'homogénéiser son encodage à l'aide d'une FLNS commune : *Document associé*.

- (60) a. Publication résultant de $\sim(\text{congrès}) = \text{actes, comptes-rendus}$ [de ART \sim] ;
 b. Transcription de ce qui est dit durant $\sim(\text{interrogatoire}) = \text{procès-verbal}$ [de ART \sim] ;
 c. Document officiel où les informations sur la \sim de X sont enregistrées(*nai-*
sance $\#I) = \text{acte, certificat, extrait}$ [de \sim] ;

- d. Rapport rédigé par X pour rendre compte de $\sim(\textit{autopsie}) = \textit{rapport}$ [d'~].

4.3.2.4 Axe

Cette relation concerne une entité (qu'elle soit physique ou non) ou un fait et le nom exprimant la trajectoire, l'axe ou encore la direction suivie par cette entité ou ce fait. Outre les exemples extraits du DiCo ci-dessous en (61), cette relation peut unir des paires de lexies comme HISTOIRE et FIL, ROMAN et TRAME, etc.

- (61) a. Trajectoire de $\sim(\textit{comète}\#I) = \textit{course}, \textit{route}$ [de ART ~];
 b. Parcours suivi par $\sim(\textit{fleuve}\#I) = \textit{cours}$ [de ART ~];
 c. Parcours suivi par $\sim(\textit{rivière}\#I) = \textit{cours}$ [de ART ~];
 d. Parcours suivi par $\sim(\textit{ruisseau}) = \textit{cours}$ [de ART ~];
 e. Trajectoire de $\sim(\textit{tir}\#3) = \textit{axe}, \textit{trajectoire}$ [de ART ~];
 f. Stratégie mise au point pour $\sim(\textit{bataille}\#I.1) = \textit{plan}(2)\#III.1$ [de ~].

4.3.2.5 Vêtement typique

Nous proposons la création de la FLNS **Vêtement typique** pour formaliser le lien entre un individu occupant un certain métier ou une certaine fonction et l'uniforme qu'il revêt à cet effet. En complément des exemples (62) extraits du DiCo :

- (62) a. Pièce d'habillement de ~ à la cour(*avocat*\#I) = *épitoge, toge, toque, robe* ;
 b. Vêtements de $\sim(\textit{facteur}(1)) = \textit{uniforme}$;

nous pouvons ajouter les quelques liens lexicaux suivants pour illustrer l'intérêt de cette formule :

- (63) a. Vêtement typique(*policier*) = *uniforme, képi ; gilet pare-balle* ;
 b. Vêtement typique(*soldat*) = *uniforme, treillis* ;
 c. Vêtement typique(*curé*) = *soutane*.

Nous concluons l'exposé des analogies entre relations lexicales encodées par des FLNS différentes, repérées d'après la simple observation de notre corpus. Les propositions qui viennent d'être faites ont pour but d'illustrer cette démarche empirique. Bien d'autres similitudes pourraient être dégagées suivant ce procédé. Nous

allons à présent proposer deux cadres méthodologiques pour guider la recherche d'analogies : la dérivation morphologique et les attributs de description.

4.4 Recherche de régularités via la dérivation morphologique

4.4.1 Corrélation entre les fonctions lexicales et la dérivation morphologique

La définition de la notion de dérivation sémantique donnée par Polguère [2000] (et présentée en détail à la section 2.2.1 p. 28) ne met pas suffisamment en évidence le rapport qu'elle entretient avec la dérivation morphologique. Dans cette section, nous souhaitons développer l'étroit lien existant entre les FL et le système morphologique des langues. En privilégiant la synthèse et la sémantique en tout premier lieu, Mel'čuk et al. [1995 : 133] minimisent le lien entre les FL de dérivation (S_0 , V_0 , A_0 , Adv_0 , et par ailleurs, S_i , S_{instr} , S_{med} , etc.), qui modélisent avant tout des dérivés sémantiques, et les valeurs qu'elles retournent. Ils présentent les trois exemples suivants :

(64) $V_0(\textit{serment}) = \textit{jurer}$;

(65) $Adv_0(\textit{durer}) = \textit{pendant}$;

(66) $Adv_0(\textit{absolu}) = \textit{tout à fait}$;

et les commentent comme suit :

Comme les exemples le montrent, il ne s'agit nullement de dérivation morphologique. Les FL « dérivées » ne reflètent que des relations lexicales pures, où les liens morphologiques sont fortuits.

Il est vrai qu'il n'existe pas toujours de concordance entre dérivation morphologique et dérivation sémantique. Certains dérivés morphologiques s'affranchissent de leur base pour développer un sens particulier. Ainsi, comme le montre Corbin [1987 : 89], alors que les unités lexicales FAC-SIMILÉ, FACTEUR, FACILE, FAÇON, FACTOTUM et

FACULTÉ sont fondées sur une étymologie commune, elles ne partagent plus de lien sémantique en synchronie. Cependant, nous pensons qu'un lien morphologique apporte un indice de récurrence non négligeable. De plus, la forte corrélation entre dérivés morphologiques et dérivés sémantiques démontre que la notion de dérivation sémantique s'est construite autour de celle de dérivation morphologique. S'il est vrai que les dérivés sémantiques ne présentent pas toujours de parenté morphologique avec leur base, comme on le voit dans les exemples (64), (65) et (66), on ne doit pas occulter le fait que ces mêmes valeurs sont potentiellement synonymes de dérivés morphologiques comme on le voit en (64') (65') et (66') :

(64') $V_0(\textit{serment}) = \text{"disparu" } \textit{sermenter, jurer}$;

(65') $Adv_0(\textit{durer}) = \textit{durant, pendant}$;

(66') $Adv_0(\textit{absolu}) = \textit{absolument, tout à fait}$.

Toute dérivation sémantique peut potentiellement s'exprimer par un dérivé morphologique. Lorsque la dérivation morphologique n'est pas attestée ou qu'elle est vieillie et trop éloignée de sa base sur le plan sémantique, les dérivés sémantiques non morphologiquement marqués permettent de remplacer le dérivé morphologique attendu. Par exemple, le verbe SERMENTER (XII^e siècle), défini dans [Hécart, 1834] par : « faire prêter serment, le prêter soi-même » a disparu en français contemporain et se voit remplacé par la lexie JURER ou l'expression collocative *prêter serment*. À l'observation des FLNS, nous avons remarqué un certain nombre de valeurs morphologiquement reliées à leur mot-clé, par exemple :

Appareil servant à mesurer l'intensité avec laquelle $X \sim(\textit{applaudir}\#I) =$
applaudimètre ;

[Qqch.] qui a la consistance <le goût> de $\sim(\textit{farine}) = \textit{farineux}$.

Or, selon nous, l'existence même d'une parenté morphologique est la preuve d'une relation privilégiée entre les deux unités lexicales, dans le sens où la dérivation morphologique procède par analogie et non par cas isolés. La parenté morphologique invite à détecter les autres relations construites analogiquement. Ainsi, on trouve sur le modèle de FARINE → FARINEUX les paires : CIRE → CIREUX ; PÂTE → PÂ-

TEUX ; LAIT → LAITEUX, etc.

Afin de mettre en valeur le lien existant entre les FL représentant des relations lexicales récurrentes et les dérivés morphologiques, nous introduisons la notion de dérivatème développée par Mel'čuk [1994a] dans le *Cours de Morphologie Générale*. Les dérivatèmes (ou affixes dérivationnels) sont des significations lexicales servant à former de nouvelles unités lexicales qu'on appelle dérivés morphologiques. Par exemple, la lexie ARROSOIR est dérivée de ARROSER, par le dérivatème 'nom d'instrument', marqué par le suffixe *-oir* ; le verbe anglais (TO) MODERNIZE est dérivé de MODERN par le dérivatème 'transformatif' marqué par le suffixe *-ize*, etc. Selon les langues, les sens portés par les dérivatèmes (par exemple, transformatif, ou causatif, etc.) ne s'expriment pas tous au moyen d'affixes dérivationnels. Certaines langues nécessitent l'emploi d'une unité lexicale supplémentaire. Par exemple, alors qu'en abkhase (langue de la famille nord-caucasienne) il existe un infix *-r-* exprimant la causation, le français et l'anglais utilisent les verbes FAIRE et MAKE pour exprimer ce sens :

Abkhase : *abara* → *arbara* ;

Français : mourir → *faire mourir* ;

Anglais : die → *make die*.

En d'autres termes, ces différences font en sorte que certaines relations sémantiquement équivalentes sont paradigmatiques dans certaines langues, où elles s'expriment au moyen de dérivés sémantiques, et syntagmatiques dans d'autres.

La notion de dérivatème à présent introduite, nous souhaitons montrer, dans le tableau 4.II, la correspondance entre les FL standard et différents dérivatèmes extraits de plusieurs langues. Nous nous appuyons sur le recensement des dérivatèmes effectué par Mel'čuk [1994a], sur la liste de dérivés morphologiques du français proposée par Bonnard [1983] ainsi que sur nos propres collectes d'exemples.

FL	dérivatème	exemple
S ₁	Nom d'agent	Basque : eros → eros+ dun : acheter → celui qui achète
S ₂	Nom de patient	Anglais : examine → examin+ ee : examiner → celui qui est examiné
S _{loc}	Nom de lieu	Évène : tavat → tavat+ čak : forger → forge
S _{instr/med}	Nom d'instrument	Mansi : kons → kons+ up : peigner → peigne
S _{mod}	Nom de moyen	Tcherkesse : txe → txe+ k'e : écrire → façon d'écrire
S _{res}	Nom de résultat	Mari : pun → pun+ em : tresser → tresse
Caus	Causatif	Abkhase : abara → arbara : mourir → faire mourir
Perm	Permissif	Batsbi : Dajvten žagnõ dagĩ → As Dajvtego žagnõ dag+ it + ẽ : David a vu le livre → J'ai laissé David voir le livre
CausFunc ₀	Productif	Évène : xačik → xačik+ alta : ouverture → pratiquer une ouverture
Real	Usitatif	Anglais : hammer → (to) hammer : marteau → marteler
Fact ₁	Usitatif	Français : buveur → boire
Labreal ₁₂₊₃	Usitatif	Français : colle#I → coller #I.1a
Réciproque	Réciproque	Kayardilt : bala → bala+ tu' : frapper → se frapper réciproquement
Anti	Oppositif	Esperanto : faldi → mal +faldi : plier → déplier
Able/Qual	Possibilitif	Français : lire → lis+ ible
Adv ₀	Adverbialisation	Anglais : mad → mad+ ly : folle → folle+ ment
A ₀	Adjectivisation	Français : aliment → aliment+ aire
S ₀	Nominalisation	Français : élever → élev+ age
V ₀	Verbalisation	Français : danse → dans+ er
Sing	Singulatif	Espéranto : neĝo → neĝ+ ero : neige → flocon de neige
Mult	Collectif	Français : feuille → feuil+ age
Cap	Chef de ~	Espéranto : urbo → urb+ estro : ville → maire
A ₁	Qui est ~	Français : mensonge → mensong+ er
A ₂	Tel qu'on le ~	Français : bague → bagué
Adv ₁	En étant ~	Français : gai → gai+ ement
Magn	Intensificatif	Français : sensible → ultra +sensible
AntiMagn	Diminutif	Français : botte → bott+ ine
AntiBon	Mauvais	Espéranto : homo → fi +homo : homme → méchant homme
Pejor	Dépréciatif	Français : vin → vin+ asse

FL	dérivatème	exemple
Loc _{ad}	Se déplaçant vers ~	Français : mener → a +mener
Loc _{ab}	Se déplaçant à partir de ~	Français : porter → em +porter
PredLoc _{in}	Se trouver à/dans	Inuktitut : snaq → sna+ mi : rive → être sur la rive
Instr	Instrumentatif	Hongrois : bicikli → bicikli+ vel : vélo → à vélo
Propt	Consécutif	Hongrois : harag → harag+ ból : colère → par colère
CausPred	Factitif	Français : égal → égal + iser
Func, Oper, Labor, Involv	Verbalisateur	Français : bourgeon → bourgeonn+ er
Incep	Inchoatif	Espéranto : ami → ek +ami : aimer → tomber amoureux
Cont	Continuatif	Allemand : machen → weiter +machen : suivre (un cours) → poursuivre
Fin	Terminatif	Hongrois : enni → meg +enni : manger → finir de manger
Liqu	Liquidatif	Français : honorer → dés +honorer
Excess	Fonctionnement excessif	Français : charger → sur +charger
De_nouveau	Réitératif	Français : tenter → re +tenter
Centr	Le milieu de ~	Français : carême → mi +carême
Syn	Synonymie	Français : ciné → cin+ oche

Tableau 4.II: Correspondance entre FL et dérivatèmes

Le tableau 4.II met en évidence une très forte corrélation entre les FL standard et les dérivatèmes de plusieurs langues du monde. La majeure partie des relations encodées par les FL standard peuvent être exprimées au moyen de la dérivation morphologique. Voyons à présent s'il est possible d'effectuer une corrélation entre la dérivation morphologique et les FLNS. Dans le *Cours de Morphologie Générale*, à la suite des dérivatèmes les plus courants que nous venons de mettre en correspondance avec les FL standard, Mel'čuk [1994a : 377] expose un certain nombre

de dérivatèmes non classifiés, formant un ensemble ouvert et se prêtant mal à une systématisation :

- a. 'X est une personne/dispositif qui fabrique/utilise Y' ;
- b. 'X est une personne qui provient de Y' ;
- c. 'X est une personne qui aime les Ys' ;
- d. 'X est la femelle (moins souvent, le mâle) de Y' ;
- e. 'X est le matériau destiné à fabriquer Y' ;
- f. 'X est un récipient/ un local destiné à Y' ;
- g. 'X est la langue parlée par les Ys' ;
- h. 'X est la peau/la fourrure de Y' ;
- i. 'X est la viande de Y' ;
- j. 'X est le renne destiné à transporter Y', etc.

Nous avons mis ces dérivatèmes (mis à part g. et j.) en correspondance avec des FLNS observées dans le DiCo :

- a'. Individu qui fabrique $\sim(\text{charrette}\#\text{a}) = \text{charron}$;
- b'. Habitant de $\sim(\text{terre}\#\text{III}) = \text{terrien-n}$;
- c'. Individu qui s'intéresse à $\sim(\text{curiosité}\#\text{II}) = \text{amateur}$ [de $\sim\text{s}$] ;
- d'. Femelle du $\sim(\text{canard}) = \text{cane}$;
- e'. Matériau dont sont faits $\sim(\text{pneu}) = \text{caoutchouc}$;
- f'. Endroit où l'on range $\sim(\text{balai}) = \text{placard}$ [à $\sim\text{s}$] ;
- h'. Fourrure/peau de \sim utilisée comme matériau $(\text{mouton}\#\text{I.a}) = \text{mouton}\#\text{I.c}$;
- i'. Chair de \sim utilisée comme aliment $(\text{mouton}\#\text{I.a}) = \text{mouton}\#\text{I.b}$.

Il serait bon de chercher à vérifier si cet ensemble de dérivatèmes est véritablement ouvert et dans quelle mesure il l'est. Contrairement à ceux du tableau 4.II, les dérivatèmes ci-dessus (de a. à j.) sont moins fréquents dans la langue, ce qui explique qu'ils soient mis en correspondance avec des FLNS et non des FL standard. Cependant, rien n'indique qu'il ne soit pas possible de les énumérer. De nombreuses études en dérivation morphologique ont déjà entrepris des inventaires plus ou moins systématiques des dérivatèmes du français [Bonnard, 1983, Corbin, 1987, Dubois

et Dubois Charlier, 1999]. Si l'on peut dégager des listes exhaustives d'affixes dérivationnels, c'est que l'ensemble des dérivatèmes n'est pas ouvert, du moins, à l'intérieur d'une langue donnée. Dans cette optique, partir d'un inventaire de dérivatèmes peut constituer une méthodologie tout à fait pertinente pour découvrir des analogies entre relations lexicales. Nous illustrons cette approche dans la section suivante.

4.4.2 Recherche d'analogies entre liens lexicaux à partir de marqueurs morphologiques

Comme nous venons de le présenter, la présence d'un marqueur morphologique pour une relation lexicale est généralement un révélateur de la fréquence de cette relation. Afin d'illustrer la méthodologie fondée sur la piste de la morphologie dérivationnelle, nous exposons, dans cette section, la façon dont il est possible de repérer, à partir de marqueurs morphologiques, des analogies entre relations lexicales. Nous proposons de nouvelles FLNS pour normaliser sept relations lexicales encodées de façon hétérogène. Rappelons une fois de plus que ces FLNS sont destinées à remplacer les anciennes FLNS, celles-ci servant dorénavant de gloses pour expliciter plus en détail le sens de la relation.

4.4.2.1 Les équivalences masculine et féminine et les noms de petits

Nous avons relevé dans le corpus des relations entre unités lexicales morphologiquement liées (cf exemples (68), (69a) et (69b) plus bas). Les suffixes permettant de construire les unités lexicales dérivées correspondent respectivement aux dérivatèmes 'équivalent féminin' et 'diminutif'. Ces deux dérivatèmes permettent d'inférer une certaine récurrence de ces relations lexicales. Par analogie à la relation d'équivalence féminine, nous avons postulé la relation d'équivalence masculine, celle-ci étant parfois, mais non systématiquement, morphologiquement marquée.

Les trois relations suivantes peuvent difficilement prétendre à la standardité dans la mesure où elles ne peuvent s'appliquer qu'à un nombre restreint d'unités lexicales :

les noms d'individus, de métiers et d'animaux. Toutefois, leur encodage étant inutilement hétérogène, nous proposons de les regrouper sous trois formules uniques de FLNS : *Masc* pour les équivalents masculins ou mâles (cf. exemples (67)), *Fem* pour les équivalents femelles ou féminins (cf. exemples 68)) et *Petit de* \sim pour les noms de petits d'animaux (cf. (69)) ou, éventuellement, de jeunes individus.

(67) a. *Mâle(abeille)* = \lceil faux bourdon \rceil ;

b. *Mâle de* \sim (poule) = *coq* ;

(68) a. *De sexe féminin(avocat)* = *avocate* ;

b. *Femelle de* \sim (canard) = *cane* ;

c. *Femelle(chien)* = *chienne* ;

d. \sim *femelle(mouton)* = *brebis*.

(69) a. *Petit de* \sim (chat) = *chaton* ;

b. *Jeune* \sim (coq) = *coquelet* ;

c. *Larve de* \sim (grenouille) = *têtard*.

4.4.2.2 La relation ‘fait/entité X — discipline étudiant X’

Dans bien des cas, la relation entre un fait ou une entité et son étude scientifique est morphologiquement marquée par les suffixes *-logie* ou *-tique*, ce qui témoigne d'une récurrence notable de la relation. Cependant, dans le DiCo, cette relation est encodée de façon hétérogène (cf. exemples (70) ci-dessous). Nous proposons d'homogénéiser les différents encodages de cette relation à l'aide d'une FLNS unique : *Étude de* \sim .

(70) a. *Science qui étudie ce qui à rapport à* \sim (crime#I.a) = *criminologie* ;

b. *Discipline qui étudie* \sim (médicament) = *pharmacie, pharmacologie* ;

c. *Science qui étudie* \sim (poisson#I.a) = *ichtyologie* ;

d. *Science qui étudie* \sim (signe#I.1) = *sémiologie, sémiotique* ;

- e. Étude des $\sim(\text{signe}\#\text{III}) = \text{horoscope, astrologie}$;
- f. Discipline scientifique qui étudie les phénomènes liés à $\sim(\text{vie}\#\text{I.1}) = \text{sciences de la } \sim$;
- g. Étude des $\sim(\text{astre}) = \text{astronomie}$;
- h. Discipline qui étudie $\sim(\text{langue}) = \text{linguistique}$.

Cette relation concerne un très grand nombre d'unités lexicales, la condition de cardinalité serait donc remplie si l'on souhaitait faire de cette FL une candidate au statut standard. Toutefois, la diversité des valeurs est relativement limitée dans la mesure où celles-ci sont souvent dérivées morphologiquement de leur base par le dérivatème *-logie*. La FLNS Étude de \sim semble bien s'intégrer au système de FL : elle peut se combiner à la FL S_1 , par exemple, pour encoder la relation 'Personne qui étudie \sim ', cf. exemples (71), ce qui motive d'autant plus sa normalisation.

- (71) a. S_1 Étude de $\sim(\text{crime}\#\text{I.a}) = \text{criminologue}$;
- b. S_1 Étude de $\sim(\text{poisson}\#\text{I.a}) = \text{ichtyologue}$;
- c. S_1 Étude de $\sim(\text{signe}\#\text{III}) = \text{astrologue}$.

4.4.2.3 La relation 'fait/entité X — qui a l'aspect/qui évoque X'

La plupart des relations dérivationnelles adjectivales sont encodées au moyen de FL standard : A_0 ('équivalent adjectival'), A_1 ('tel qu'il est \sim '), A_2 ('tel qu'on le \sim '), $Able_1$ ('tel qu'il est susceptible de \sim '), $Qual_2$ ('tel qu'il est potentiellement objet de \sim ') ainsi que leurs modulations avec les différents actants concernés, comme le montrent les exemples (72) :

- (72) a. $A_2(\text{déception}) = \text{décevant}$;
- b. $Able_1(\text{crainte}) = \text{craintif}$;
- c. $Qual_2(\text{haine}) = \text{odieux}$;
- d. $A_1(\text{abandon}\#\text{I.2b}) = \text{abandonné}$;
- e. $A_0(\text{catastrophe}\#\text{I}) = \text{catastrophique}$.

Nous avons trouvé parmi les FLNS des relations adjectivales dérivationnelles correspondant au sens ‘qui a l’aspect de \sim ’ ou ‘qui évoque \sim ’, dont certaines figurent dans les exemples (73). Ces relations sémantiquement très similaires sont, la plupart du temps, morphologiquement marquées et nous semblent relativement fréquentes. Nous pensons qu’elles pourraient être encodées à l’aide d’une FLNS unique et générale, comme *Qui évoque \sim* .

- (73) a. [Qqch.] qui a l’aspect ou la consistance de $\sim(\textit{lait}\#\textit{b}) = \textit{laiteux}$;
 b. [Qqch.] qui a la consistance <le goût> de $\sim(\textit{farine}) = \textit{farineux}$;
 c. Qui a l’aspect <la consistance> de $\sim(\textit{cire}) = \textit{cireux}$;
 d. [Qqch.] qui évoque le pelage de $\sim(\textit{tigre}) = \textit{tigré}$;
 e. Qui fait penser à $\sim(\textit{éléphant}\#\textit{I}) = \textit{pachydermique, éléphantique}$.

En complément de l’observation des dérivatèmes parmi des FLNS du DiCo, nous sommes également partie de certaines listes de référence de dérivés ou composés savants et avons vérifié si les relations dont ils sont les marqueurs étaient présentes et fréquentes dans le DiCo sans pour autant être marquées morphologiquement. Par exemple, les deux relations ‘X — qui se nourrit de X’ et ‘X — altération de X’ présentées dans la suite de cette section ont été dégagées d’après les éléments *-vore* et *-ite*, extraits d’une liste de dérivés et composés savants du français [Bonnard, 1983]. Une liste de référence peut donc servir de guide pour découvrir et prévoir des analogies entre relations lexicales.

4.4.2.4 La relation ‘X — qui se nourrit de X’

Cette relation ne s’applique qu’aux noms de plantes ou d’animaux pour caractériser ceux qui s’en nourrissent. Ses moyens d’expression sont limités, il s’agit principalement des éléments de composition savante *-vore-* et *-phage-*. Nous proposons la FLNS *Qui se nourrit de \sim* pour représenter cette relation. Notons que cette FLNS est à la fois nominale et adjectivale, elle peut produire en valeur aussi bien des noms que des adjectifs, par exemple : CARNIVORE_N et CARNIVORE_{Adj} . Nous n’avons relevé qu’une occurrence de cette FLNS dans notre corpus (74a). Toutefois,

un bon nombre d'exemples pourraient s'ajouter à cette occurrence unique comme les relations (74b), (74c) et (74d).

- (74) a. Animal qui se nourrit de $\sim(\textit{grain}(1)\#\text{I.1}) = \textit{granivore}$;
 b. Qui se nourrit de $\sim(\textit{viande}) = \textit{carnivore}$;
 c. Qui se nourrit de $\sim(\textit{homme}) = \textit{anthropophage}$;
 d. Qui se nourrit de $\sim(\textit{fruit}) = \textit{carpophage}, \textit{fructivore}$.

4.4.2.5 La relation 'X — altération de X'

La relation entre une partie du corps ou une entité plus générale et le nom de l'affection ou de l'altération qui lui est associée est bien souvent marquée par les éléments *-algie* ou *-ite*, qui désignent respectivement une douleur ou une inflammation. Nous proposons la FLNS plus générale *Altération de \sim* pour représenter cette relation. Le sens de *Altération de \sim* se veut relativement large, il couvre le nom d'une affection comme en (75a) ou, plus largement, une chose affectant l'entité dénotée par le mot-clé comme en (75b).

- (75) a. Affections de $\sim(\textit{estomac}\#\text{I}) = \textit{crampe}, \textit{colique}, \textit{gastrite}, \textit{mal d'estomac}, \textit{ulcère}$;
 b. Bouton/furoncle sur $\sim(\textit{paupière}) = \textit{compère-loriot}, \textit{orgelet}$.

La méthodologie fondée sur les marqueurs morphologiques peut ainsi se poursuivre pour découvrir des relations lexicales récurrentes. Elle permet également de prévoir ou d'anticiper les liens lexicaux à décrire au fur et à mesure que le corpus de travail s'agrandit. Enfin, elle s'avère tout à fait utile pour expliquer l'origine des relations lexicales recensées dans le DiCo — et notamment des dérivés sémantiques, intimement liés au phénomène de dérivation morphologique. Nous allons à présent explorer une autre méthodologie très différente destinée à regrouper, organiser et normaliser les FLNS syntagmatiques.

4.5 Organisation des FLNS selon des attributs de description

4.5.1 Présentation

À la suite de la méthodologie de repérage des similitudes entre relations lexicales fondée sur la dérivation morphologique, nous présentons dans cette section, une autre méthodologie consistant à décrire les relations à l'aide d'attributs descriptifs. Il s'agit d'une démarche tout à fait différente qui vise non plus à créer de nouvelles FL standard ou non standard *ad hoc*, mais à organiser, présenter et normaliser les FLNS associées à une lexie dans leur ensemble. Alors que la section précédente concernait principalement les dérivés sémantiques encodés par des FLNS nominales, celle-ci sera davantage tournée vers les collocations encodées par les FLNS adjectivales.

Nous partons de l'hypothèse que tout fait ou entité dénoté par une lexie possède des attributs très généraux de type : 'temps', 'fonction', 'fonctionnement', 'aspect', 'participants', etc., susceptibles d'être exprimés au moyen de collocatifs ou de dérivés sémantiques. Ces attributs peuvent être mis en correspondance avec les composantes sémantiques des définitions des unités lexicales, mais sont conceptuellement plus larges et ne sont pas hiérarchisés (contrairement aux composantes sémantiques, qui sont considérées centrales ou périphériques). Nous cherchons donc, dans cette section, à déterminer quels sont les attributs modélisés par les FL. Nous pensons que raisonner en termes d'attributs va permettre d'organiser globalement les FL standard et non standard. Plus précisément, cette approche vise deux objectifs :

1. identifier et normaliser les FLNS en mettant en valeur certaines régularités ;
2. organiser les FLNS dans les bases de données lexicales.

4.5.1.1 Une perspective nouvelle sur les FL

En raisonnant en termes d'attributs, nous pouvons considérer les FL standard sous un nouvel angle. Il est possible, par exemple, de percevoir les FL adjectivales

comme modélisant les pôles positifs et négatifs des attributs d'importance (Magn et AntiMagn), d'appréciation (Bon et AntiBon) et de finalité (Ver et AntiVer) (cf. figure 4.2).

Importance :	+	Magn(<i>autopsie</i>) = <i>méticuleuse</i>
	-	AntiMagn(<i>aboitement</i>) = <i>faible</i>
Appréciation	+	Bon(<i>aspiration</i>) = <i>noble</i>
	-	AntiBon(<i>aboitement</i>) = <i>intempestifs</i>
Finalité	+	Ver(<i>ami</i>) = <i>fidèle</i>
	-	AntiVer(<i>applaudir</i>) = <i>mollement</i>

Figure 4.2 – FL standard adjectivales sous l'angle d'attributs polarisés

D'une certaine façon, ces attributs et d'autres de même type sont déjà présents dans certaines FL semi-standard et non standard, mais de façon non explicite. On observe, par exemple, pour la lexie CONVERSATION#a, les collocations adjectivales suivantes :

- (76) Magn_{temp} (→ Qui dure longtemps) : *longue, interminable, « sans fin »* ;
- (77) AntiMagn_{temp} (→ Qui dure peu de temps) : *brève, courte, rapide* ;
- (78) Magn_{échange} (→ Durant laquelle X et Y sont très actifs) : *enflammée, grande, animée, nourrie* ;
- (79) AntiMagn_{échange} (→ Durant laquelle X et Y sont peu actifs) : *languissante, morne* ;
- (80) Magn_{plaisir} (→ Plaisante) : *amicale, détendue, enjouée, gaie, agréable, charmante* ;
- (81) AntiMagn_{plaisir} (→ Peu plaisante) : *fatigante, difficile, hostile, houleuse, orangeuse, tendue* ;
- (82) AntiBon_{clarté} (→ Peu claire) : *décousue, embrouillée, incohérente, pagailleuse* ;
- (83) Bon_{contenu} (→ Intéressante) : *intelligente, intéressante* ;
- (84) AntiBon_{contenu} (→ Peu intéressante) : *fade, insipide, plate, terne, inintéressante* ;
- (85) Ver_{but} (→ Productive) : *constructive, productive, utile, fructueuse* ;
- (86) AntiVer_{but} (→ Non productive) : *inutile, stérile, vaine* ;
- (87) Ver_{contenu} (→ Durant laquelle X et Y disent ce qu'ils pensent) : *franche* ;

- (88) Durant laquelle X et Y traitent en détail un sujet donné : *suivie* ;
- (89) Durant laquelle X et Y sont seuls : \lceil *en tête-à-tête* \rceil , *privée* ;
- (90) Dont le contenu Z est important : *grave, sérieuse* ;
- (91) Dont le contenu Z est peu <n'est pas> important : *légère, frivole, futile, insignifiante* ;
- (92) Durant laquelle X et Y interviennent comme ils veulent sans que \sim ait une structure définie : \lceil *à bâtons rompus* \rceil .

En décomposant le sens de ces différentes relations, on peut dégager des composantes de sens récurrentes comme : ‘intense’, ‘peu intense’, ‘long’, ‘court’, ‘agréable’, ‘désagréable’, ‘non structuré’, ‘intéressant’, ‘inintéressant’, ‘sans but’, ‘vide’, etc. On observe donc que certaines de ces composantes correspondent à des FL standard (cf. figure 4.2) alors que d’autres, toutes aussi importantes pour définir la relation, ne sont représentées que par des notations associées aux FL standard, du type *contenu*, *plaisir*, *but*, ou simplement décrites dans des FLNS comme dans les exemples (88) à (92). Face à cette hétérogénéité, il semble légitime de questionner la pertinence de ce choix d’encodage. Les fonctions semi-standard (dans les exemples (76) à (87)) donnent l’impression d’une hiérarchisation parmi les composantes de sens des relations, les composantes les plus saillantes semblant être représentées par des FL standard, et les autres, ajoutées sous forme de complément au moyen de notations. Or, dans le processus d’encodage, rien n’indique qu’il faille privilégier les composantes représentables par des FL standard aux dépens d’autres, éventuellement plus saillantes. Si l’on considère les exemples (80), (81), (82), (83) et (84), il semble que les composantes ‘contenu’, ‘plaisir’ et ‘clarté’ constituent les éléments majeurs de la relation que les éléments **Bon**, **AntiBon**, **Magn**, et **AntiMagn** polarisent en leur conférant une valeur positive ou négative.

Cette façon d’encoder les relations contribue à renforcer l’hétérogénéité parmi les formules de FL. Certaines sont standard, d’autres semi-standard, d’autres encore non standard alors que toutes ces valeurs sont très semblables les unes aux autres dans leur rapport à la lexie CONVERSATION#a, dont elles donnent une description. La solution pour palier ces problèmes d’encodage serait de chercher à définir pour les lexies des propriétés ou attributs du type ‘forme’, ‘taille’, ‘fonction’, ‘contenu’,

et d'encoder l'ensemble des FL qui leur sont associées au moyen de ces attributs. Toutefois, avant d'entreprendre une analyse des FL en termes d'attributs, nous souhaitons en premier lieu mettre notre travail en perspective avec l'approche du projet FrameNet. Nous présentons donc les différents éléments de cadres proposés dans FrameNet qui s'avèrent relativement similaires, du moins dans leur fonction d'identification, aux attributs que nous cherchons à définir.

4.5.1.2 Les éléments de cadre dans FrameNet

Nous avons présenté FrameNet en détail à la section 2.2.2. Dans cette ressource lexicale, chaque unité lexicale est reliée à un ou plusieurs cadres conceptuels qu'elle évoque. Ces cadres conceptuels sont définis au moyen d'éléments de cadre (= *Frame Elements* ou *FE*). Ceux-ci permettent de décrire la valence sémantique d'une unité lexicale. Chaque cadre contient des éléments de cadre centraux (= *core FE*) et des éléments de cadre secondaires (= *non core FE*). Par exemple, le cadre <TRANSFER> a pour éléments centraux : DONOR, RECIPIENT, THEME et TRANSFEROR, et pour éléments secondaires : MANNER, MEANS, PLACE, PURPOSE, REASON et TIME.

Chacun des éléments de cadre est susceptible d'être exprimé au moyen d'unités lexicales de divers types. Notons qu'il s'agit avant tout de cooccurrences libres exprimées dans une phrase et participant à la mise en scène d'un cadre conceptuel. Aussi, n'importe quelle unité lexicale employée dans une phrase mettant en scène un cadre est susceptible de correspondre à un élément de cadre. À titre d'illustration, nous avons reproduit, à la figure 4.3 (p. 146), le cadre <ACCOUTREMENTS>. On peut ainsi voir, à travers les exemples, que l'élément de cadre WEARER peut se réaliser par le pronom *She*, l'élément TIME OF CREATION, par le syntagme *17th century* et que ni l'un ni l'autre ne sont des collocations ou des dérivés sémantiques des lexies évoquant le cadre <ACCOUTREMENTS>.

Beaucoup d'éléments sont communs à plusieurs cadres, par exemple, TIME est commun aux cadres <ARREST>, <ARRIVING>, <BEING_BORN>, <KILLING>, etc. Cependant, certains d'entre eux sont très précisément liés à certains cadres : l'élément COPARTICIPANT, par exemple, n'existe que pour le cadre <ACCOMPANIMENT> ;

Accoutrements

Definition : A **Wearer** wears accessories, which are made of some **Material** and may have a **Style**. Most of the accessories are non-essential (clothing) items.

Ex : *She*_(Wear.) wore a *black*_(Desc.) *velvet*_(Mat.) *HAT on her head*_(Body_loc.).

→ **Core Fe :**

– ACCOUTREMENT [Acout] : This FE identifies the accoutrement. Ex : *She wore a tiara*_(Acout).

→ **Non-Core Fes :**

– BODY_LOCATION [Body_loc] : This FE identifies the place on the body where a particular Accoutrement is located. Ex : *Tess wore a single RING in her left nostril*_(Body_loc).

– CREATOR [cre] : The individual or group who created the Accoutrement. Ex : *De Beers*_(cre) *JEWELRY is justly famous*.

– DESCRIPTOR [Desc] : This FE identifies a characteristic or description of the Accoutrement. Ex : *Kim wore an exquisite*_(Desc) *TIARA*.

– MATERIAL [Mat] : This FE identifies the Material from which the Accoutrement is made. Ex : *He wore a gold*_(Mat) *WATCH*.

– NAME [nam] : A name given to an individual Accoutrement to distinguish it from all other objects. Ex : *The RING Narya*_(nam) *was passed to Gandalf*.

– PART [Part] : This FE identifies a part of the Accoutrement. Ex : *He bought a leather CAP with sheepskin flaps*_(Part).

– STYLE [St1] : This FE identifies the Style of the Accoutrement. Ex : *He wore a Jacobean*_(St1) *WIG*.

– TIME_OF_CREATION [tim] : The time at which the Accoutrement was created. Ex : *What was a 17th century*_(tim) *BROOCH doing in her attic?*

– USE [Use] : This FE identifies the purpose for which an Accoutrement is worn or made. Ex : *She was wearing reading*_(Use) *GLASSES*.

– WEARER [Wear] : The Wearer wears the Accoutrement. Ex : *Why must you always wear your*_(Wear) *HELMET when I'm driving?*

Figure 4.3 – Le cadre <ACCOUTREMENTS> et ses différents éléments de cadre

MENTAL CONTENT n'apparaît que sous <REMEMBERING_INFORMATION>, UNCONFIRMED CONTENT seulement sous <VERIFICATION>, etc. Le nombre total d'éléments s'élève à plus de 550 pour environ 700 cadres décrits. Nous avons reproduit ci-dessous (figure 4.4) les vingt éléments les plus fréquents.

MANNER	REASON	INSTRUMENT	TOPIC
TIME	DEPICTIVE	THEME	SPEAKER
PLACE	AGENT	SOURCE	
MEANS	DURATION	PATH	
DEGREE	GOAL	CAUSE	
PURPOSE	RESULT	ENTITY	

Figure 4.4 – Les vingt éléments de cadre les plus fréquents dans FrameNet

Les éléments de cadre dans FrameNet sont très nombreux. Comme nous l'avons brièvement évoqué, les lexies pouvant exprimer les éléments ne sont pas forcément en relation privilégiée avec le mot-clé. L'analyse de FrameNet constitue un point de repère et de comparaison intéressant avec notre approche, cependant, nous souhaitons définir des attributs seulement à partir des collocatifs et dérivés sémantiques. Pour permettre une maintenance et un traitement raisonnable des FLNS, nous tenterons également de limiter le nombre d'attributs en choisissant des noms suffisamment généraux pour englober plusieurs sens proches.

4.5.2 Organisation des FLNS selon des attributs de description

4.5.2.1 Présentation des différents attributs

La recherche d'attributs permet de faire émerger un grand nombre de régularités parmi les FLNS et de proposer une normalisation de leur encodage. Rappelons que nous appliquons principalement cette méthode aux FLNS adjectivales et plus particulièrement à celles représentant des collocations. Nous avons déterminé les attributs de façon empirique, à partir de l'observation des FLNS du DiCo. Nous avons tenté de respecter au maximum le principe d'économie dans la détermination

des attributs afin de contrôler et limiter leur nombre. Toutefois, lorsqu'un attribut potentiel recouvre un certain nombre de sous-types différents, nous le scindons en différents attributs. Par exemple, les relations ayant trait à l'aspect d'une entité sont réparties entre les attributs COULEUR, FORME, TAILLE, MATÉRIAU, FABRICATION et ORIGINE, comme on peut le voir dans les exemples ci-dessous. L'attribut ASPECT a tout de même été conservé pour représenter les relations qui n'entrent pas dans un de ses sous-types :

– ASPECT

Qui est orné de plumes(*chapeau*) = à plumes, à plumet ;

Durci, renforcé par une armature métallique(*béton*) = armé ;

– COULEUR

De couleur blanche(*barbe*) = blanche ;

De couleur grise(*barbe*) = grisonnante, « poivre et sel » ;

– FORME

Qui a une forme évasée(*barbe*) = « en éventail » ;

Qui comporte deux niveaux pour les passagers(*wagon*) = à impériale ;

– TAILLE

Fait pour être utilisé simultanément par deux personnes(*drap*) = double ;

Fait pour être utilisé par une seule personne(*drap*) = simple ;

– MATÉRIAU

Qui est fait de feutre(*chapeau*) = //feutre ;

Qui est fait de paille(*chapeau*) = de paille ;

– FABRICATION

Qui est fabriqué par cuisson(*jambon*) = braisé ;

Qui est fait, pour ce qui est du mode de fabrication, de façon traditionnelle ou qui en a l'apparence(*pain#I*) = « à l'ancienne » ;

– ORIGINE

Qui provient d'un mélange de races(*chien#I*) = //bâtard ;

Qui provient d'un tonneau, et non d'une bouteille(*bière(1)#b*) = pression.

Nous avons recensé un certain nombre de valeurs ayant trait au contenu, ou plus spécifiquement, aux ingrédients utilisés et créé les attributs **CONTENU** et **INGRÉDIENT** :

– **CONTENU**

Qui a trait à la religion(*juron#1*) = *blasphématoire* ;

Dont le contenu Z est important(*conversation#a*) = *grave* ;

– **INGRÉDIENT**

Qui est fait, pour ce qui est des ingrédients, sans sel(*pain#I*) = *sans sel* ;

Qui est fait, pour ce qui est des ingrédients, à base de farine brute(*pain#I*) = *complet*.

Nous avons, par la suite, dégagé des relations adjectivales concernant les éléments d'une situation dénotée par le mot-clé, c'est-à-dire : la position, l'état, la fonction, le fonctionnement, le statut, la situation/localisation, le temps, la cause et la conséquence.

– **POSITION**

Qui est le dernier dans un ensemble(*wagon*) = *dernier, de queue* ;

Qui se trouve en-dessous d'où l'on est(*étage#1*) = *du dessous* ;

– **ÉTAT**

Qui a été utilisée(*assiette#a*) = *sale* ;

Qui a la rage(*chien#I*) = *enragé* ;

– **FONCTION**

Équipé pour que les passagers y dorment(*wagon*) = *~lit* ;

Que X porte pour chasser(*chapeau*) = *de chasse* ;

– **FONCTIONNEMENT**

Qui possède un carillon(*horloge*) = *à carillon* ;

Qui fonctionne grâce à un mécanisme(*montre*) = *mécanique* ;

– **STATUT**

Dans lequel il est interdit de fumer(*wagon*) = *non-fumeur* ;

Telle qu'elle n'est pas professionnelle(*compétition#2*) = *amicale* ;

– **SITUATION/LOCALISATION**

Telle qu'elle est sous la terre(*canalisation*) = *souterraine* ;

Qui vit à l'état sauvage(*canard#I.a*) = *sauvage* ;

– TEMPS

Qui a lieu le matin(*concert*) = *matinée* ;

Valable pour un an(*enveloppe#III*) = *annuelle* ;

– CAUSE

Qui est causé par un incendie(*dégât*) = [~] *du feu* ;

Qui a lieu malgré la volonté de X(*célibat*) = *forcé, obligé* ;

– CONSÉQUENCE

Qui tue l'être Y(*coup de feu#I*) = *mortel* ;

Qui nuit à la réputation de Y(*peine#II*) = *infamante*.

Enfin, deux attributs permettent de représenter les collocations portant sur les participants d'un événement ou sur les actants d'un mot-clé :

– PARTICIPANT

Qui a lieu entre les chefs de deux partis(*bataille#II.1*) = *de chefs* ;

Qui est plutôt fréquenté par des familles(*restaurant*) = *familial* ;

– CIBLE

Qui vise le grand public(*cinéma#I.1*) = *commercial* ;

Pour les troupes d'une armée en guerre(*théâtre#I.1*) = *aux armées*.

Nous sommes en présence d'une vingtaine d'attributs englobant l'ensemble des FLNS adjectivales collocationnelles. Cette liste n'est toutefois pas définitive : certains attributs pourraient être amenés à se préciser, à se scinder en sous-attributs ou, au contraire, à se généraliser. La recherche d'attributs s'inscrit dans un cadre plus large que la simple normalisation des FLNS. Les attributs sont intimement liés aux classes sémantiques de FL que nous définirons dans le module d'organisation sémantique au chapitre suivant et au sein desquelles ils prendront place. Ainsi, sous la classe sémantique LOCALISATION, on pourra trouver aussi bien des FL standard comme S_{loc} et Loc_{in} que des FLNS du type Qui est le dernier dans

un ensemble ou encore Telle qu'elle est sous la terre qui auront, entre temps, subi une normalisation de leur encodage selon le principe que nous présentons dans le paragraphe suivant.

4.5.2.2 Encodage normalisé des attributs dans les FLNS

Les FLNS adjectivales collocatives exprimant un attribut particulier peuvent maintenant être encodées selon un patron bien défini. L'encodage que nous proposons revêt la forme d'une fonction à deux arguments : un paramètre venant préciser ce qui est dit de l'attribut ainsi que le mot-clé traditionnel des FL. Le modèle ainsi obtenu est le suivant :

Attribut : paramètre(*mot-clé*) = *valeurs*, par exemple :

Forme : évasée(*barbe*) = \ulcorner *en éventail* \urcorner ;

Cible : grand public(*cinéma#I.1*) = *commercial*.

Ce nouvel encodage permet d'organiser les FLNS dans les bases de données lexicales. Plutôt qu'une liste de FLNS non structurée, un utilisateur peut repérer les relations qu'il recherche selon ces attributs. Voyons un exemple avec les FLNS adjectivales liées aux unités lexicales CHAPEAU et WAGON dans les figures 4.5 et 4.6. Notons que les figures doivent être lues selon le modèle suivant : ancienne FL \rightarrow FL normalisée.

CHAPEAU

ASPECT

Qui est fait de paille \rightarrow Matériau : paille : *de paille*

Qui est fait de feutre \rightarrow Matériau : feutre : *de feutre, //feutre*

Qui est orné de plumes \rightarrow Aspect : orné de plumes : *à plumes, à plumet*

Qui a une forme ronde \rightarrow Forme : ronde : *rond*

Qui a une forme pointue \rightarrow Forme : pointue : *pointu*

FONCTION

Que X porte pour chasser \rightarrow Fonction : chasser : *de chasse*

Que X porte pour se protéger du soleil \rightarrow Fonction : se protéger du soleil : *de soleil*

Figure 4.5 – FLNS adjectivales normalisées de la lexie CHAPEAU

WAGON

ASPECT

Qui comporte deux niveaux pour les passagers → Forme : à deux niveaux : *à impériale*

Qui a des parois et un toit pour abriter Y → Forme : avec paroi et toit : *couvert*

Qui n'a ni paroi ni toit → Forme : sans paroi ni toit : *à plateforme, découvert, plat*

UTILISATION

Équipé pour que les passagers y dorment → Fonction : dormir : *~couchette(s), ~lit(s)*

Équipé pour que les passagers y mangent → Fonction : manger : *~restaurant*

Équipé d'un bar → Fonction : boire ou manger : *~restaurant*

Qui est le premier dans un ensemble de W. → Position : premier : *premier, de tête*

Qui est le dernier dans un ensemble de W. → Position : dernier : *dernier, de queue*

Qui est à la Num-e position dans ensemble de W. → Position : n-ième : *Num-e*

Qui suit ou précède le W. en question → Position : qui suit ou précède le ~ : *voisin-adj*

Dans lequel il est permis de fumer → Statut : permis de fumer : *fumeurs*

Dans lequel il est interdit de fumer → Statut : interdit de fumer : *non-fumeurs*

Catégorisé comme offrant le meilleur confort aux passagers

→ Statut : meilleur confort aux passagers : *de première (classe)*

Catégorisé comme étant le plus économique pour les passagers

→ Statut : plus économique pour les passagers : *de seconde (classe)*

Figure 4.6 – FLNS adjectivales normalisées de la lexie WAGON

4.6 Conclusion

Dans ce quatrième chapitre, nous avons cherché à traiter les problèmes posés par les FLNS. En raison de leur encodage hétérogène, sans structure définie ni langage véritablement contrôlé, elles constituent un obstacle majeur à tout traitement automatisé. Par ailleurs, le très fin degré de granularité choisi pour l'identification des relations lexicales conduit à une démultiplication des formules de FLNS et complexifie la recherche d'analogies entre liens lexicaux. Les aspects formels des FLNS les écartent radicalement des autres FL et contribuent à former deux ensembles disjoints de FL. Nous avons questionné la pertinence de l'établissement d'une frontière aussi étanche entre les FL standard et non standard, tant du point de vue théorique que du point de vue de leur encodage et posé la nécessité d'une normalisation des FLNS.

Notre objectif dans ce chapitre était de proposer deux méthodologies pour le repérage des analogies entre liens lexicaux et la régularisation de leur encodage en harmonie avec le système plus global de FL. La première méthode consiste à exploi-

ter le lien étroit existant entre les FL et le phénomène de dérivation morphologique pour repérer et même prévoir des similitudes entre liens lexicaux. Elle concerne plus particulièrement les dérivés sémantiques. La seconde méthodologie, très différente de la première, porte sur les FL syntagmatiques et consiste à dégager les différents attributs concernés par les relations encodées. Ces deux méthodologies permettent de prévoir les liens lexicaux à décrire au fur et à mesure de l'évolution de la base de données. Par exemple, les relations : *Étude de* \sim (\rightarrow *Discipline étudiant* \sim) et *Qui évoque* \sim , dégagées à l'observation des suffixes *-logie* et *-eux*, ont de fortes chances de s'appliquer à un certain nombre de lexies non décrites actuellement dans la base DiCo. Enfin, les encodages proposés — principalement ceux de la seconde méthodologie — vont permettre une organisation explicite des FLNS dans le DiCo et les autres bases de données du même type. Leur formalisation permettra de les repérer et de les manipuler automatiquement. Les attributs viendront s'intégrer à l'intérieur d'une organisation sémantique commune à toutes les FL, peu importe leur statut. Nous présentons cette organisation sémantique, ainsi que, plus globalement, une organisation des FL selon plusieurs perspectives, dans le chapitre suivant.

Chapitre 5

Élaboration d'un système de fonctions lexicales

5.1 Introduction

Le quatrième chapitre de cette thèse a été consacré à définir des stratégies de mise en valeur des analogies entre les fonctions lexicales non standard (FLNS) ainsi qu'à proposer un cadre pour la normalisation de leur encodage. Grâce à cette normalisation, un grand nombre de FLNS peuvent maintenant être identifiées et manipulées au même titre que les FL standard et figurer à leurs côtés dans une organisation globale des FL.

Ce cinquième chapitre porte sur la question de l'organisation des FL. Dans les bases de données et dictionnaires élaborés selon les principes de la LEC, les FL sont listées les unes à la suite des autres ou ordonnées selon des critères n'étant pas toujours très intuitifs. Cette façon de présenter les FL ne met pas en évidence les différents types de liens qui les unissent et rend l'organisation des données peu naturelle et figée dans une organisation linéaire. Nous souhaitons offrir une organisation non plus linéaire, mais modulaire, c'est-à-dire épousant plusieurs axes de structuration indépendants les uns des autres et intégrés au sein d'un même **système**. Avant de poser l'avantage d'un système de FL sur une classification, nous devons faire la distinction entre ces deux termes. Alors qu'une classification organise et présente des éléments dans des classes déterminées selon un ou plusieurs critères (forme, sens, etc.), un système vise à expliciter les liens entre ses éléments selon plusieurs perspectives, c'est-à-dire à refléter l'ensemble de leurs connexions possibles. Notons que le terme *système de FL* est couramment employé dans la littérature portant sur

la LEC sans qu'il fasse pour autant référence à un véritable système explicité. Nous tâchons donc d'explicitier le système que forment naturellement les FL, en mettant en valeur un certain nombre de leurs connexions. Nous mettons en avant les relations entre les FL selon quatre perspectives, détaillées dans les sections 5.2, 5.3, 5.4 et 5.5 : une perspective sémantique, une perspective liée aux parties du discours, une perspective combinatoire et, finalement, une perspective communicationnelle. Ces quatre perspectives constituent un modèle d'organisation des FL à partir duquel nous pourrions, dans de futurs travaux, créer des interfaces de consultation pour une BLEC adaptées à des besoins particuliers : recherche onomasiologique, recherche par structure syntaxique, etc.

Le système est réalisé d'après les FL du DiCo, plus particulièrement, l'ensemble des FL standard et semi-standard ayant deux ou plus de deux arguments distincts (= 670), les FLNS adjectivales (= 727) et les FLNS nominales (= 820).

Avant de détailler les différentes facettes du système, nous allons à présent analyser les classifications des FL proposées antérieurement. Ces classifications portent sur les FL standard simples. Rappelons que notre objectif, dans ce chapitre, est d'étendre la problématique de la classification des FL à toutes les FL, c'est-à-dire, des FL standard (simples et complexes) aux FLNS.

5.1.1 Regard critique sur les classifications des fonctions lexicales

Depuis la fin des années 60, on recense un grand nombre de textes dans lesquels les FL sont présentées, dont Zholkovsky et Mel'čuk [1966, 1967, 1970], Mel'čuk [1982], Mel'čuk et Zholkovsky [1984, 1988], Mel'čuk et al. [1984, 1988], Steele [1986], Steele et Meyer [1990], Grimes [1990], Alonso Ramos [1993], Mel'čuk et al. [1995], Alonso Ramos et Tutin [1996], Fontenelle [1997b], etc. Certains de ces textes proposent une classification ou typologie afin de regrouper les FL dans des classes aisément identifiables. Notons bien que ces classifications ne portent que sur les FL standard simples. Après la parution de l'ILEC [Mel'čuk et al., 1995], la présentation des FL s'est standardisée en une classification qui est majoritairement

reprise dans les articles et mémoires portant sur les FL. Nous reproduisons trois classifications déjà présentées et critiquées par Alonso Ramos [1993] dans sa thèse de doctorat, ainsi que celle qu'elle propose alors pour pallier les insuffisances des précédentes. Il nous semble que ces quatre tentatives, entreprises par quatre auteurs à des époques différentes illustrent l'évolution de la perception des FL. Nous avons choisi ces quatre classifications, car, dans une certaine mesure, elles se répondent et se corrigent pour tenter d'aboutir à une meilleure représentation. Nous souhaitons apporter ici une perspective diachronique et épistémologique sur la conception et la catégorisation des FL. Nous reproduisons une première classification de Zholkovsky et Mel'čuk [1970], puis vingt ans plus tard et à l'intérieur d'un même ouvrage, deux classifications reposant sur des principes divergents : celle de Steele et Meyer [1990], qui cherche à donner un caractère plus intuitif à l'organisation des FL, et celle de Grimes [1990], qui sort des sentiers battus et propose une nouvelle voie pour la représentation des liens lexicaux fondée sur des paires de FL inverses. Nous exposons par la suite la classification de Alonso Ramos [1993], raffinée dans Alonso Ramos et Tutin [1996], qui repose sur une critique des trois premières. Nous concluons cette section par un bilan critique de l'ensemble des classifications proposées.

5.1.1.1 Classification de Zholkovsky et Mel'čuk [1970]

Dans une perspective diachronique, cette classification des FL est intéressante puisqu'il s'agit d'une des toutes premières à être présentée. La classification repose sur la distinction entre les *substitutions* et les *paramètres sémantiques*. Comme nous l'avons mentionné auparavant dans la thèse, l'objectif principal des FL est de participer aux règles de paraphrasage. Les deux classes de FL sont distinguées ici selon le rôle que jouent les FL vis-à-vis du mot-clé dans les paraphrases. Voyons les définitions données par Zholkovsky et Mel'čuk [1970] :

Les substitutions équivalentes sont des mots ou des groupes de mots qui dans le texte peuvent être utilisés à la place du mot-clé (ou se combiner avec lui), la phrase conservant exactement le même sens sans qu'on ait besoin de changements auxiliaires autres que les raccords syntaxiques. Les substitutions équivalentes peuvent appartenir, soit à la même partie

du discours que celle du mot-clé (synonymes ou conversifs), soit à une autre partie (dérivés). [Zholkovsky et Mel’čuk, 1970 : 35]

Les paramètres sémantiques sont des fonctions lexicales dont les valeurs sont des moyens d’expression idiomatique pour un mot donné de certains “sens” réguliers qui peuvent accompagner de nombreux mots en présence d’un mot i_0 . Ces “sens” ont chacun plusieurs expressions dans la langue en question ; de plus, le choix de l’expression dépend entièrement du mot-clé i_0 . [Zholkovsky et Mel’čuk, 1970 : 40]

Les substitutions regroupent les FL *Syn*, *Conv* et les quatre les dérivés syntaxiques (A_0 , S_0 , V_0 , et Adv_0). Les FL correspondant aux paramètres sémantiques sont : *Copul*, *Incep*, *Fin*, *Caus*, *Liqu*, *Oper*, *Func*, *Labor*, *Fact*, *Real*, S_i , A_i , *Loc*, S_{loc} , S_{mod} , S_{instr} , *Magn*, *Anti*, *Gener*, *Ver*, *Son*, *Prepar*, *Mult*, *Sing*, *Bon*, *Degrad*, etc.

Le classement proposé est relativement sommaire dans la mesure où, mise à part la distinction initiale opposant les substitutions aux paramètres sémantiques, aucune sous-classe n’est proposée. Par ailleurs, les paramètres se résument à une longue liste désordonnée de FL. Dans les classifications qui ont suivi, on remarque que la distinction entre paramètres sémantiques et substitutions a évolué vers une opposition entre FL syntagmatiques et FL paradigmatiques. Toutefois, nous devons bien garder à l’esprit que ces deux paires de notions ne sont pas équivalentes. Les substitutions ne constituent qu’un sous-ensemble des FL paradigmatiques et peuvent, en outre, s’exprimer de façon syntagmatique avec le mot-clé. Les paramètres peuvent se répartir à la fois parmi les FL syntagmatiques (par exemple : *Real*, *Magn*, etc.) et parmi les FL paradigmatiques (par exemple : S_{loc} , *Gener*, *Anti*, etc.).

5.1.1.2 Classification de Steele et Meyer [1990]

La classification proposée par Steele et Meyer [1990], vingt ans plus tard, repose sur l’opposition entre les FL syntagmatiques et les FL paradigmatiques, qui, comme nous venons de le mentionner, a relayé, entre temps, la précédente opposition entre paramètres sémantiques et substitutions. Ils définissent ces deux grandes classes de la façon suivante [Steele et Meyer, 1990 : 42] :

The paradigmatic class consists of functions that are based essentially on meanings regularly associated with a keyword as an element in a language's system of semantic relations. [...] The syntagmatic class consists of functions that are based primarily on relations arising from the collocational properties of their keyword as an element in a language's speech system.

L'objectif recherché dans cette typologie est le regroupement des FL dans des catégories intuitives, aisées à retenir [Steele et Meyer, 1990 : 41], c'est pourquoi les auteurs ont introduit un certain nombre de sous-classes. Nous reproduisons dans la figure 5.1 (p. 163) leur classement tel qu'il est présenté dans l'article.

Alonso Ramos [1993 : 228-229] émet un certain nombre de critiques vis-à-vis de cette proposition. Elle remet en cause le fait de considérer les FL des classes (4), (9), (11), (12) (cf. figure 5.1), ainsi que la plupart de celles des classes (13) et (14) comme des relations paradigmatiques, dans la mesure où ces FL se combinent toujours avec le mot-clé. Sous l'étiquette « *qualifiers of a keyword* » sont regroupées des FL qui ne qualifient nullement le mot-clé, comme *Cap* et *Equip*, *Mult* et *Sing*, etc. Elle remarque que la FL *Pos* modifie le mot-clé, non pas un de ses actants, et ne figure donc pas dans la bonne sous-classe. La distinction paradigmatique *vs.* syntagmatique ne semble donc pas respectée dans la typologie elle-même, ce qui, par conséquent, la rend invalide. Elle critique également une terminologie imprécise qui tantôt utilise le terme *actant*, et tantôt le terme *participant*, et emploie indifféremment les termes *specific*, *typical* et *standard*.

5.1.1.3 Classification de Grimes [1990]

Dans le même ouvrage édité par Steele, Grimes propose une nouvelle voie pour l'organisation des FL. Il rejette la distinction traditionnelle entre paradigmatique et syntagmatique jugée non pertinente :

Like many other things in language, lexical functions do divide according to the well-established opposition between paradigmatic and syntagmatic phenomena.[...]

It fits well with the history of linguistic thought to view the functions in this way, but I have not found so far that it does any more for us than that. Instead, I believe there are more productive ways to relate the functions to each other than by appealing to the opposition between paradigmatic and syntagmatic. [Grimes, 1990 : 350-351]

Grimes propose alors de mettre en valeur les liens d'inférence existants entre les FL. Cette façon de procéder conduit à séparer deux types de FL : celles qui s'associent naturellement en paires (appelées « INVERSE LF ») et celles qui décrivent une caractérisation ou une action liée au mot-clé sans se grouper à d'autres FL (appelées « DETAIL LF »). Dans la figure 5.2 (p. 164), nous avons reproduit sa classification avec la terminologie anglaise originale telle qu'elle est présentée dans Grimes [1990 : 360]. Selon son point de vue, les FL ne sont plus considérées comme un simple moyen de formaliser des associations sémantiques régulières parmi les paires de mots. Il s'agit plus profondément de questionner la façon dont la structure des associations entre FL se couple avec les patrons de cohérence textuelle et de mettre en rapport les FL avec les aptitudes humaines à faire des inférences. Cette façon de raisonner a été proposée par Martha Evens et résumée dans Evens et al. [1985]. Elle est en partie motivée par les tentatives de simuler les chaînes de raisonnement qui doivent être activées lorsque l'on cherche des éléments dans une base de données complexe ou lorsqu'on essaie de montrer la façon dont un mot contribue au développement logique d'un texte.

La recherche de fonctions inverses contribue à la création d'un grand nombre de FL comme *State*, qui joue le rôle de l'inverse de *Liquidative*, voire de nouvelles paires de FL en opposition, comme les FL *Antecedent* et *Consequent* illustrées ci-dessous :

Liquidative (*mistake*) = *erase* (→ Causer la fin de ~) ;
State (*erase*) = *mistake, error* (→ Ce dont on cause la fin) ;
Antecedent(*shut*[event]) = *open* [state] (→ Action antécédente) ;
Consequent (*open*[state]) = *shut* [event] (→ Action conséquent).

La proposition de Grimes est intéressante, car elle bouleverse le schéma habituel de classification des FL. L'idée de mettre en valeur les connexions entre FL est une

des raisons qui nous poussent à constituer un système explicite de FL, et la notion de fonction inverse s'inscrit dans cette perspective. Cependant, on doit y apporter quelques critiques. Tout comme Alonso Ramos [1993 : 203-231], nous pensons que créer de nouvelles FL selon cette méthode augmente inutilement le nombre de FL. De plus, l'idée de créer des fonctions inverses va dans une certaine mesure à contre courant du principe d'orientation du lien¹. Par exemple, la relation entre ERASE et MISTAKE est orientée : aussi le sens de Liquidative dans le contexte de MISTAKE se lexicalise par ERASE, cependant ceci ne vaut pas forcément pour la relation inverse : le sens State dans le contexte de ERASE ne se lexicalise pas nécessairement par MISTAKE. En d'autres termes, Grimes, en cherchant une nouvelle voie, semble s'écarter quelque peu des fondements théoriques inhérents aux FL pour établir une représentation des FL qui cherche à tout prix à être logique. Par ailleurs, si la première partie de sa classification semble apporter une perspective nouvelle, en revanche, la seconde partie n'apporte pas d'éléments novateurs.

5.1.1.4 Classification de Alonso Ramos [1993] et Alonso Ramos et Tutin [1996]

Dans sa thèse de doctorat, Alonso Ramos [1993] fait le point sur les classifications que nous venons de présenter et en propose une nouvelle. Cette classification est reprise et raffinée dans Alonso Ramos et Tutin [1996]; nous la reproduisons avec notre traduction dans la figure 5.3 (p. 165).

Cette classification, bien qu'elle soit construite en regard des insuffisances formulées sur les autres classifications, comprend elle-même quelques petites imperfections. Nous pensons qu'une classe du type (15) « Verbes divers » devrait être évitée dans une classification, car elle sert à récupérer les FL qui n'ont pas trouvé de classes adéquates. Par ailleurs, plusieurs critères pour déterminer des sous-classes sont utilisés sur le même plan : parties du discours (prépositions, verbes), fonctions syntaxiques (modification, qualificatif), sens (verbes de réalisation), si bien que

¹Cette notion est définie dans le deuxième chapitre, p. 30.

divers phénomènes linguistiques se retrouvent mélangés, car confinés dans un seul classement.

I PARADIGMATIC FUNCTIONS

A. Substitutions

1. Basic logical relations : Syn, Anti, Conv
2. Derivations : S₀, V₀, A₀, Adv₀
3. Contrastive terms : Contr

B. Qualifiers of a keyword

4. Evaluative qualifiers : Magn, Ver, Bon
5. Categorical qualifiers : Gener
6. Distributional qualifiers : Mult, Sing
7. Organization qualifiers : Cap, Equip
8. Quantitative qualifiers : Plus, Minus
9. Figurative qualifiers : Figur
10. Pleonastic qualifiers : Epit

C. Aspects of a keyword-situation

11. Process features : Caus, Germ, Culm, Degrad, Excess, Obstr, Pejor, Liqu, Perf, Result
12. Phasal features : Incep, Cont, Fin
13. Attendant features : Invov, Instr, Manif, Nocer, Perm, Prepar, Propt, Prox, Sympt, Son, S_{instr}, S_{med}, S_{mod}, S_{loc}, S_{res}
14. Spatio-temporal features : Loc_{in}, Loc_{ab}, Loc_{ad}, Loc_{in}^{temps}, Centr
15. Standard names for participants : S₁, S₂, S₃
16. Standard command to participants : Imper
17. Copulas : Copul

D. Qualifiers of actants

18. Typical qualifiers of actants : A₁, A₂, A₃, ..., Adv₁, Adv₂, Adv₃, ...
19. Specific qualifiers of actants : Able₁, Able₂, ... ; Qual₁, Qual₂, ... Pos₁, Pos₂, ...

II SYNTAGMATIC FUNCTIONS

A. Verbal operators

20. Verbal operators that are 'semantically empty' : Oper_i, Func_i, Labor_{ij}
21. Verbal operators that have a semantic component : Real_i, Fact_i, Labreal_{ij}

B. Predicators

22. Predicators : Pred

Figure 5.1 – Classification de Steele et Meyer [1990]

INVERSE LEXICAL FUNCTIONS

Category functions : regroupent les FL N_0 ($=S_0$), V_0 , A_0 , Adv_0 , Copula et Pred

Simple inverse functions : sont constituées de Syn_{\subset} et Syn_{\supset} , Antecedent et Consequent, Liqu et State, Sympt et Manifest, Leader et Organization (= Cap et son inverse), Group et Focus (= Equip et son inverse), Unit et Mass (Sing et son inverse), Multiple et Individual (Mult et son inverse).

Symmetric inverses functions : incluent Syn (synonymie exacte), Syn_{\cap} , Conv, et trois variétés d'antonymes : Complementarity, Antonym et Reversal.

Clustered inverse functions : il s'agit de deux FL différentes qui ont la même inverse : les FL causatives Caus et Perm ont toutes deux Event pour FL inverse.

Hierarchical inverse functions : Gener et Spec, Whole et Part, Sequence et Phase.

DETAIL FUNCTIONS

Actant functions : N_1 , N_2 , N_3 ($=S_1$, ...), A_1 , $Able_1$, Adv_1 , Pos_1 , ...

Peripheral functions : Location, Location_{from}, Location_{to}, Location_{temporal}, $N_{locative}$ (= Loc, Loc_{ab}, Loc_{ad}, Loc_{temp} et S_{loc}), Instr et N_{instr} (= S_{instr}), N_{mod} , N_{means} et $N_{consequent}$ (= S_{mod}, S_{med}, S_{res})

Aspect functions : Incep, Cont, Centr, Degrad, Fin et Perf, Prep, Prox, Antecedent et Consequent.

Degree functions : Magn, Bon, Ver, Increase et Decrease (= Plus et Minus)

Factives functions : Subject Factive_i, Object Factive_i, and Adjunct Factive_{ij}.

Miscellaneous functions : Imper, Figur, Involv, Manif, Sympt, Son.

Figure 5.2 – Classification de Grimes [1990]

A) FL paradigmatiques	B) FL syntagmatiques
(1) Relations de substitution Syn, Anti, Anti _{comp} , Contr, Conv _{ijk} , Gener _p	(7) Verbes supports Oper _i , Func _i , Labor _{ij}
(2) Dérivations syntaxiques S ₀ , A ₀ , V ₀ , Adv ₀	(8) Verbes de réalisation Real _i , Fact _i , Labreal _{ij}
(3) Dérivations actantielles S _i , A _i , Adv _i	(9) Prédicateur Pred
(4) Dérivations circonstancielles S _{instr} , S _{med} , S _{res} , S _{mod}	(10) Verbes phasaux Incep, Cont, Fin
(5) Qualificatifs d'actants potentiels Able _i , Qual _i	(11) Verbes causatifs Caus, Liqu, Perm
(6) Ordre typique Imper	(12) Verbes aspectuels Prox, Perf, Result
	(13) Verbes introducteurs d'une nouvelle situation Involv, Manif, Sympt _{ijk}
	(14) Verbes ayant un sens négatif Degrad, Excess, Obstr, Nocer
	(15) Verbes divers Prepar, Son
	(16) Modificateurs adjectivaux et adverbiaux Magn, Bon, Ver, Pos _i , Epit, Plus, Minus, Pejor
	(17) Prépositions liées au mot-clé Instr, Loc _{in/ab/ad} , Propt
	(18) Noms quasi-synonymes Figur, Gener
	(19) Noms quasi-méronymes Sing, Mult, Cap, Equip, Centr, Culm, Germ

Figure 5.3 – Classification de Alonso Ramos et Tutin [1996]

5.1.1.5 Bilan sur les classifications proposées

Voyons ce qu'on peut retenir de ces différentes classifications : premièrement, il apparaît que même avec un principe de base identique pour certaines propositions — la distinction entre relations paradigmatiques et syntagmatiques —, on observe une forte divergence d'une classification à une autre. Ces divergences ne manquent pas d'attirer l'attention sur la validité de cette distinction. On peut en déduire, soit que les concepts de relations paradigmatiques et syntagmatiques ne sont pas perçus de la même manière par les différents auteurs, soit que les critères posés pour la répartition des FL ne sont pas respectés, ce qui invalide alors les typologies proposées. Si l'on veut être plus radical encore, on dira, tout comme Grimes, que la distinction n'est peut-être pas pertinente pour classer les FL.

En ce qui concerne la structure des typologies, on remarque que certaines d'entre elles ne sont pas toujours organisées en sous-classes [Zholkovsky et Mel'čuk, 1970], que l'on trouve parfois un grand nombre de sous-classes composées d'une seule FL [Steele et Meyer, 1990], ou encore qu'une case « fourre-tout » (cf. « miscellaneous functions » [Grimes, 1990]) est utilisée pour combler une difficulté de classification, ce qui ne peut être considéré comme une sous-classe satisfaisante.

Une autre critique que l'on peut formuler à l'encontre de ces classements est qu'ils ne prennent en compte que les FL standard simples (composées d'un seul élément standard). Or, ce que nous cherchons à organiser, ce sont les FL qui figurent dans les DEC et les BLEC, c'est-à-dire les FL standard simples et complexes (pouvant combiner jusqu'à cinq éléments différents), les FL semi-standard, localement standard ainsi que les FLNS.

Il semble que les difficultés rencontrées sont la preuve qu'on ne peut parvenir à un classement satisfaisant s'il doit être cantonné dans une ou deux dimensions. Pour mettre à profit les différents types d'information fournies par les FL sans les entremêler, on doit en proposer non pas une classification uni- ou bidimensionnelle, mais multidimensionnelle. Cette représentation doit pouvoir fournir un mode d'organisation pour chaque type d'information encodé. Il existe, en effet, plusieurs façons de regrouper un ensemble de données. Prenons l'exemple des relations ci-dessous :

- (93) Gener(*bonne humeur*) = *ambiance, atmosphère* ;
 (94) Magn(*abattement*#I) = *profond* ;
 (95) IncepPredMinus(*désir*#I) = *s'émousser* ;
 (96) AntiBon(*aboiments*) = *intempestifs* ;
 (97) S₁(*aboiments*) = *chien*#I ;
 (98) Degrad(*humeur*#Ia) = *s'altérer* ;
 (99) Real₁(*bateau*) = *naviguer* [sur ART ~], *piloter* [ART ~].

Deux FL peuvent être rapprochées :

- en fonction du sens de la relation qu'elles modélisent : les FL en (96) et (98) ont toutes deux une composante de sens négative ;
- en fonction du patron syntaxique de la relation : les FL en (95) et (98) prennent toutes deux le mot-clé en sujet et n'ont pas de complément ;
- en fonction de l'élément communicativement dominant dans la relation : la FL en (97) exprime le nom du premier actant du mot-clé et celle en (99), un syntagme dans lequel le premier actant est le sujet grammatical ;
- ou en fonction des parties du discours des lexies en jeu dans une relation : les FL en (94) et (96) sont adjectivales si le mot-clé est un nom, et adverbiales dans les autres cas.

Notre objectif est donc de rendre compte de ces formes de regroupements en créant une organisation des FL qui prenne en compte ces différents axes de classement et vise à rendre explicite toutes les propriétés des relations lexicales plutôt qu'à les encapsuler dans une formule synthétique.

5.1.2 Vers une organisation systémique des fonctions lexicales

Avant d'énoncer les principes de base pour l'élaboration d'un système de FL, nous devons présenter deux travaux antérieurs dont l'initiative rejoint la nôtre : la grammaire de FL proposée par Alonso Ramos [1993] et l'encodage explicite [Kahane et Polguère, 2001].

5.1.2.1 Travaux préliminaires pour l'organisation des fonctions lexicales en système

◇ Grammaire de FL : Alonso Ramos [1993]

Alonso Ramos [1993 : 232] tente d'établir une grammaire des FL standard simples en formulant des caractéristiques propres à chaque FL. Cette grammaire vise à :

- fournir une valeur prototypique pour chaque FL ;
- faire une description sémantique ou une définition générale de chaque FL ;
- établir les liens sémantiques entre la FL et son mot-clé : s'il s'agit d'un rapport d'équivalence, d'inclusion de sens ou encore d'intersection de sens ;
- indiquer sur quelle catégorie syntaxique s'applique une FL et quelle catégorie syntaxique elle produit en retour ;
- expliquer les différents rôles que jouent les indices numériques actantiels qui accompagnent certaines FL ;
- distinguer les types de combinaison des FL et en expliquer les règles.

Des fiches descriptives sont élaborées pour chaque FL standard simple sur le même modèle de celle de S_{loc} que nous reproduisons dans une traduction française à la figure 5.4 (p. 169).

◇ L'encodage explicite : Kahane et Polguère [2001]

L'encodage explicite, déjà présenté en détail lors du troisième chapitre (section 3.3.4), permet de rendre explicite, dans un langage formel, différents types d'information pour chaque FL : leur sens, leur valence ainsi que leur partie du discours. Pour représenter le sens des FL, des universaux sémantiques sont dégagés. La valence, quant à elle, est représentée au moyen de codes précisant les éléments entrant en jeu dans la relation et leur configuration syntaxique. Nous reprendrons une partie de ce travail dans la suite de ce chapitre.

Ces deux initiatives posent les bases pour l'extraction des propriétés des relations lexicales encodées par les FL. Cependant, les FL y sont traitées une à une selon, notamment, leurs propriétés relationnelles et leur mode de combinaison aux autres

• S_{loc}

Définition : Nom typique pour désigner le circonstant de lieu de la situation dénotée par le mot-clé

Exemples :

$S_{loc}(boxe) = ring$

$S_{loc}(patiner) = patinoire$

$S_{loc}(tennis) = court$ [de \sim]

$S_{loc}(vendre) = marché$

$S_{loc}(enterrer) = cimeti\grave{e}re$

$S_{loc}(enseigner) = \acute{e}cole, universit\acute{e}$

$S_{loc}(tauromachie) = ar\^ene$

Valeur prototypique : *lieu*

Liens sémantiques : ‘C’ \subset ‘ $S_{loc}(C)$ ’^a

Prérequis sémantiques : C doit être une lexie prédicative

Catégories syntaxiques : $S_{loc}(C_{C=N,V})_{C=N}$ ^b

Capacité combinatoire :

– FL régente de FL complexe ou composées avec des FL verbales

– est située à gauche : $S_{loc}Real_4(le\grave{c}on) = \acute{e}tude, salle d’\acute{e}tude$

^aEn d’autres termes, cette équation indique que le sens de la lexie C est inclus dans celui de $S_{loc}(C)$. Par exemple, le sens ‘boxe’ est présent dans la lexie RING.

^bCette formule indique que S_{loc} prend en mot-clé un nom ou un verbe et renvoie des valeurs nominales.

Figure 5.4 – Fiche de la FL S_{loc} dans Alonso Ramos [1993 : 377]

FL. Notre objectif n’est pas de définir la dimension algébrique du système de FL, ni de calculer leurs propriétés sémantiques et syntaxiques à partir des opérations formelles explicitées par Kahane et Polguère [2001], mais d’établir la structuration d’ensemble au sein de laquelle figure individuellement chaque FL. En d’autres termes, nous cherchons à regrouper les caractéristiques communes des FL pour en dégager un modèle d’organisation, permettant d’obtenir une vision globale des différentes interactions possibles entre FL au sein d’un système structuré et explicite. Il est, en effet, difficile de dégager les propriétés d’un système si l’on se limite à l’étude de ses unités. En plus de montrer les similitudes entre FL, ce mode d’or-

ganisation permettra de mettre en avant leurs différences et de faire ressortir leurs valeurs au sein du système.

5.1.2.2 Propriétés définitoires pour un système de fonctions lexicales

Nous définissons quatre propriétés du système de FL que nous avons élaboré :

1. Fonction du système

Il s'agit d'une modélisation des propriétés attachées aux FL, elles-mêmes modélisant les relations lexicales. Le système a pour objectif de mettre en valeur les similitudes ou les aspects communs aux FL tout en soulignant leurs différences. Il est conçu pour être utilisé dans des applications de recherche d'informations à l'intérieur d'une BLEC, sous la forme de plusieurs interfaces de consultation. Il servira également de point de référence pour le calcul d'informations lexicographiques en vue de l'automatisation partielle de la rédaction de fiches BLEC, et pourra être consulté en tant qu'outil pour assister les lexicographes dans l'encodage des relations lexicales.

2. Nature des composantes du système

Le système comprend toutes les FL sans égard à leur statut. Le travail effectué sur les FLNS et présenté au chapitre précédent permet d'intégrer les FLNS dans des classes aux côtés des FL standard. Le système inclut donc les FL standard simples et complexes, les FL semi-standard, localement standard ainsi que les FL non standard.

3. Axes d'organisation du système

Le système est organisé de façon modulaire. Ces différents modules représentent les différentes perspectives sur le système. Nous distinguons :

- une perspective sémantique regroupant les FL selon les composantes sémantiques qu'elles partagent ;
- une perspective centrée sur les parties du discours des unités lexicales entrant dans une relation lexicale ;

- une perspective combinatoire présentant les différentes configurations syntaxiques des unités lexicales en jeu dans une relation lexicale ;
- une perspective communicationnelle mettant en valeur les cibles communicationnelles des relations lexicales. Par exemple, dans la relation lexicale BOIRE → BUVEUR, nous dirons que la cible communicationnelle de la relation est le premier actant sémantique de BOIRE.

4. **Système ouvert et évolutif**

Cette propriété s'applique autant au niveau des composantes du système, c'est-à-dire les FL, qu'au niveau de sa structure. Premièrement, le système est ouvert : il doit être prêt à intégrer constamment de nouvelles données au fur et à mesure que des nouvelles FL (standard ou non standard) apparaissent. Les composantes déjà présentes dans le système peuvent être amenées à être supprimées, corrigées ou complétées. Par ailleurs, la structure du système elle-même n'est pas définitive. Il est encore en cours de modélisation, ce qui implique que des classes pourront être supprimées, scindées ou ajoutées pour s'adapter aux données à venir. En d'autres termes, le système doit pouvoir s'ajuster continuellement pour que ses données rentrent dans le canevas d'ensemble.

5.2 **Perspective sémantique sur le système de fonctions lexicales**

Nous passons à présent à la présentation des différentes perspectives sur le système de FL et des structures d'organisation que nous proposons pour chacune d'entre elles. Nous commençons par la perspective sémantique, qui propose une organisation des FL en fonction des composantes sémantiques qu'elles partagent.

5.2.1 Décomposition des fonctions lexicales en diverses composantes sémantiques

5.2.1.1 Objectif

Suivant le modèle de la définition analytique, qui décompose le sens d'une unité lexicale en diverses composantes sémantiques, il est possible de décomposer le sens d'une FL. En étudiant le rapport entre les FL et les (éventuelles) composantes sémantiques qu'elles recouvrent, nous avons distingué quatre situations :

1. FL sémantiquement vides

Il s'agit des FL dites syntaxiques S_0 (100), V_0 (101), A_0 (102) et Adv_0 (103). Elles ne comportent pas véritablement de composantes sémantiques : leur rôle est de changer une unité lexicale de partie du discours selon le phénomène de translation.

(100) $S_0(\textit{craindre}) = \textit{crainte}$;

(101) $V_0(\textit{émoi}) = \textit{s'é mouvoir}$;

(102) $A_0(\textit{chat}) = \textit{félin}_{Adj}$;

(103) $Adv_0(\textit{exagérer\#1a}) = \textit{exagérément}$.

2. FL à composante sémantique unique

La définition de certaines FL ne comportent qu'une seule composante sémantique : par exemple, les fonctions $Magn$ (104), $Real_1$ (105) et S_{loc} (106) encodent respectivement les composantes 'intensité', 'utilisation' et 'endroit typique' :

(104) $Magn(\textit{aboisement}) = \textit{féroce, fort, furieux}$;

(105) $Real_1(\textit{bateau}) = \textit{naviguer [sur ART ~] ; piloter [ART ~]}$;

(106) $S_{loc}(\textit{tir\#1b}) = \textit{champ, club, stand [de ~]}$.

3. FL à composantes sémantiques multiples

Certaines FL encodent plusieurs composantes sémantiques. Par exemple,

alors que la FL Real_1 en (107) représente la composante ‘utilisation’, la FL plus complexe $\text{AntiBon.Real}_{1@}$ ($\rightarrow [X]$ utiliser \sim en manifestant de l’agressivité envers N) en (108) encode deux composantes sémantiques : ‘utilisation’ et ‘négativité’ :

(107) $\text{Real}_1(\textit{klaxon}\#\text{a}) = //\textit{klaxonner}$;

(108) $\text{AntiBon.Real}_{1@}(\textit{klaxon}\#\text{a}) = \textit{balancer}$ [ART coup de \sim à N].

4. Plusieurs FL pour une même composante sémantique

Certaines FL sémantiquement très proches n’ont toutefois pas un encodage similaire dans la mesure où elles représentent des configurations syntaxiques différentes entre un mot-clé et sa valeur. Par exemple, les fonctions Fact_i , Real_i et Labreal_{ij} , habituellement catégorisées parmi les « verbes d’utilisation ou de fonctionnement », comportent toutes, pour les lexies d’artefact, le sens d’utilisation ou de fonctionnement dévolu à l’artefact. Elles se différencient par les positions syntaxiques des actants du mot-clé. Ainsi, la lexie CHARRETTE, dont la forme propositionnelle est notée ci-dessous, contrôle trois actants : \sim utilisée par l’individu X pour transporter Y et tirée par l’animal Z.

On trouve, dans le DiCo, les relations lexicales suivantes :

(109) a. $\text{Real}_1(\textit{charrette}) = \textit{conduire, mener}$ [ART \sim];

b. $\text{Fact}_0(\textit{charrette}) = \textit{rouler}$;

c. $\text{Fact}_2(\textit{charrette}) = \textit{transporter}$ [N=Y];

d. $\text{Labreal}_{12}(\textit{charrette}) = \textit{charrier, charroyer}$ [N=Y dans ART \sim].

Ces quatre relations contiennent le sens ‘ \sim fonctionne/est utilisé’. Si l’on se place d’un point de vue purement sémantique, il semble pertinent de les regrouper afin de pouvoir accéder à ces collocations en partant d’une seule composante sémantique.

L’objectif de l’organisation sémantique de notre système est double.

- Il s’agit, d’une part, de pouvoir accéder, à partir d’une composante sémantique, à autant de FL qui peuvent l’exprimer. Par exemple, pour la lexie CHARRETTE, on souhaite pouvoir accéder, en partant de la composante ‘fonctionnement’ (ou ‘utilisation’), aux FL Labreal₁₂, Fact₂, Fact₀ et Real₁.
- D’autre part, si une FL possède plus d’une composante sémantique, on souhaite accéder à cette FL par n’importe laquelle de ces composantes. Par exemple, à partir de la lexie KLAXON#a, il doit être possible d’accéder au syntagme *balancer un coup de klaxon* par la composante ‘utilisation’ ainsi que par la composante ‘négativité’.

5.2.1.2 Identification des composantes sémantiques

Notre objectif n’est pas de dresser un relevé exhaustif et normalisé de l’ensemble des composantes sémantiques associées aux FL, mais de les identifier de façon relativement souple et explicite. Les composantes similaires seront regroupées à l’intérieur d’une même rubrique ou classe sémantique. Afin d’extraire les composantes sémantiques liées aux FL, nous avons utilisé les gloses de vulgarisation qui leur sont associées dans le DiCo et dans le LAF. Ces gloses font office de petites définitions des FL. Toutefois, comme nous l’avons mentionné à la section 3.3.1, une FL est souvent vulgarisée de plusieurs façons selon les lexies auxquelles elle s’applique. Par exemple, neuf gloses sont associées à la FL IncepPredPlus :

~ se développe	~ devient plus rapide
~ devient plus intense	~ augmente
~ deviennent plus intenses	~ devient plus marquée
~ devient plus grand	~ devient plus marqué
~ devient plus large	

Par conséquent, nous avons choisi pour identifier les composantes sémantiques, un terme général englobant l’ensemble des formulations possibles dans les différentes gloses. Par exemple, pour la FL IncepPredPlus, nous avons dégagé deux composantes sémantiques : ‘augmentation’ et ‘commencement’.

De la même façon, vingt-neuf gloses sont associées à la FL CausFunc₀ :

[Qqn/Qqch.] causer ~	[Qqn] créer ~	se manifeste
[Phénomène naturel] créer ~	[Qqch.] causer ~	[Qqn] faire commencer ~
[Qqch.] causer qu'il y ait ~	[Qqch.] causer ~ de X	[Autorité judiciaire] causer que
[Brebis] donner naissance à ~	[Qqn] causer que ~ se déve-	X soit interrogé comme ~
[Qqn] causer que ~ ait lieu	loppe à l'endroit N	[Administration] créer ~
[Qqch.] être la cause de ~	[Qqn] faire ~	[Qqn] écrire ~
[Qqn] mettre en service ~	[Qqn] causer que ~ a lieu	[Qqn] installer ~ dans le lieu
[Qqn] fabriquer ~	[Qqn] calculer ~	N
[Qqn] faire en sorte qu'il y ait	[Qqn] construire ~	[Qqch.] causer que ~ se mani-
~ à l'endroit N	[Qqn] mettre ~ en service	feste
[Qqch.] causer qu'il y a ~	[Qqch.] causer que ~ soit ap-	
[Qqn] causer qu'il y a ~	parente	
[Qqn] rendre possible ~	[Qqn/Qqch.] faire que ~ de X	

Ces gloses peuvent être paraphrasées par une formule unique : ‘création’, à laquelle on associe CausFunc_0 .

Voyons ci-dessous, à titre d’illustration, la mise en correspondance de quelques FL avec leur composantes sémantiques :

- $\text{Fact}_0(\text{ancree}) = \text{mordre, tenir} \longrightarrow$ ‘fonctionnement’ ;
- $\text{Real}_1^{\text{I}}(\text{ancree}) = \text{jeter, mouiller} [\text{ART } \sim] \longrightarrow$ ‘utilisation’ ;
- $\text{S}_2(\text{balai}) = \text{poussière} \longrightarrow$ ‘participant’ ;
- $\text{AntiBonS}_2(\text{bateau}) = \text{passager clandestin} \longrightarrow$ ‘participant’, ‘négativité’ ;
- $\text{IncepOper}_1(\text{fortune}(1)\#1a) = \text{amasser, accumuler, se constituer, etc.}$
 $[\text{ART } \sim] \longrightarrow$ ‘commencement’ ;
- $\text{S}_1\text{CausFunc}_0(\text{farine}) = \text{meunier} \longrightarrow$ ‘participant’ ;
- $\text{Syn}(\text{vaisselle}\#2) = \text{plonge} \longrightarrow$ ‘équivalent’ ;
- $\text{Contr}(\text{comptoir}\#\text{I.1b}) = \text{salle} \longrightarrow$ ‘contrastif’ ;
- $\text{Partie de } \sim(\text{pipe}\#\text{I}) = \text{bec, fourneau, tuyau}\#\text{I.1} \longrightarrow$ ‘partie de’ ;
- $\text{Equip}(\text{navire}) = \text{équipage} \longrightarrow$ ‘participant’, ‘collectif’ ;
- $\text{Hypo}(\text{chanson}\#\text{I.1}) = \text{berceuse} \longrightarrow$ ‘type de’ ;
- $\text{S}_0\text{SingReal}_1(\text{chiffon}\#a) = \text{coup} [\text{de } \sim] \longrightarrow$ ‘unité’, ‘utilisation’ ;
- $\text{Prepar}_1(\text{meurtre}) = \text{comploter, préméditer} [\text{ART } \sim] \longrightarrow$ ‘préparation’ ;
- $\text{LiquFunc}_0(\text{assemblée}\#2) = \text{clorre, lever} [\text{ART } \sim] \longrightarrow$ ‘suppression’ ;
- $\text{S}_{\text{res}}(\text{labeur}) = \text{fruit} \longrightarrow$ ‘résultat’ ;

$S_{\text{instr}}\text{Real}_1(\text{frein}\#\text{I}) = \text{pédale} [\text{de } \sim] \longrightarrow \text{'instrument'}$;

$S_{\text{loc}}\text{Fact}_0(\text{avocat}(1)\#\text{I}) = \text{cour} \longrightarrow \text{'localisation'}$.

La décomposition des FL en composantes sémantiques n'est pas exclusive aux FL standard et s'applique également aux FLNS, selon un procédé toutefois différent. Nous dirons, par exemple, que la FLNS suivante a trait à la composante sémantique 'fonction' :

Équipé pour que les passagers y dorment(*wagon*) = $\sim\text{-lit}$ \longrightarrow 'fonction'.

Les deux suivantes concernent une composante sémantique de localisation :

Qui est le dernier dans un ensemble(*wagon*) = *dernier, de queue* \longrightarrow 'localisation' ;

Qui se trouve en-dessous d'où l'on est(*étage*\#1) = *du dessous* \longrightarrow 'localisation'.

La FLNS suivante a trait à la durée :

Valable pour un an(*enveloppe*\#III) = *annuelle* \longrightarrow 'durée'.

On peut dégager et identifier encore bon nombre de composantes sémantiques à travers les exemples suivants :

Qui est plutôt fréquenté par des familles(*restaurant*) = *familial* \longrightarrow 'participant' ;

Qui est fait de feutre(*chapeau*) = *//feutre* \longrightarrow 'matériau' ;

Qui a une forme évasée(*barbe*) = $\ulcorner \text{en éventail} \urcorner$ \longrightarrow 'forme' ;

Etc.

C'est précisément cette démarche qui nous a amenée, au chapitre précédent (cf. section 4.5), à identifier et classifier les FLNS adjectivales selon des attributs de description.

En s'appuyant sur les composantes identifiées, il est possible de regrouper, sans distinction de statut et sur le seul plan sémantique, des FL standard et non standard. Ainsi, les différentes collocations portant sur l'aspect d'un chapeau, par exemple, pourront être regroupées dans la même classe QUALIFICATIFS :

AntiBon(*chapeau*) = *cabossé* ;

Matériau : paille(*chapeau*) = *de paille* ;

Matériau : feutre(*chapeau*) = *de feutre, //feutre* ;

Forme : rond(*chapeau*) = *rond* ;

Forme : pointu(*chapeau*) = *pointu*.

De la même façon, les deux FL suivantes se côtoieront sous la classe PARTICIPANTS :

$S_{1/2}$ (*bataille#II.1*) = *adversaire* ;

Participants : chefs de deux partis(*bataille#II.1*) = *de chefs*.

5.2.2 Organisation des composantes sémantiques dans des classes sémantiques

Nous avons créé dix classes sémantiques. Le choix de ces classes a été motivé par plusieurs critères liés à la problématique de la recherche d'information. Premièrement, nous avons tenté de limiter au maximum le nombre de classes créées pour que l'on puisse, d'une part, les visualiser facilement (sur une interface de consultation d'une base de données lexicales, par exemple) et, d'autre part, les mémoriser de façon à rendre la recherche de FL la plus spontanée et rapide possible.

Deuxièmement, nous avons tenté d'élaborer des regroupements les plus intuitifs possibles du point de vue d'un utilisateur d'une base de données lexicale. Par exemple, la composante sémantique 'conversif' liée à la FL Conv_{ij} (par exemple : $\text{Conv}_{21}(\textit{suivre}) = \textit{précéder}$ [N=X]) est présente à la fois dans la classe ÉQUIVALENCE et dans la classe OPPOSITION. Conv_{ij} permet, en effet, une inversion des actants du mot-clé et relie deux unités lexicales à la fois sémantiquement équivalentes, mais opposées au niveau de leurs positions actantielles.

Par ailleurs, la décomposition des FL en composantes sémantiques rend possible la recherche de FL via deux chemins différents. Par exemple, la FL $\text{AntiBon.Real}_{1@}$, citée dans l'exemple (108) (p. 173), comporte deux composantes sémantiques : 'utilisation' et 'négativité', répertoriées dans deux classes sémantiques différentes. Il est donc possible d'y accéder par l'une ou l'autre de ces classes. Notons que, dans

notre terminologie, nous disons parfois, de façon elliptique, qu'une FL *pointe* vers une classe. Nous détaillons, à présent, chacune des classes créées et en proposons un schéma récapitulatif à la figure 5.5 (p. 188).

5.2.2.1 Classe ÉQUIVALENCE

La classe ÉQUIVALENCE concerne les FL véhiculant une identité sémantique. Elle comprend trois sous-classes : les 'lexies similaires', les 'conversifs' et la 'reformulation syntaxique'.

- Les composantes sémantiques des FL **Syn**, **QSyn** et des renvois **Cf** et **=**, bien que sémantiquement différents — **Syn** représente la synonymie exacte, **QSyn** la synonymie approximative, **Cf** relève davantage de l'association d'idée et **=** représente la stricte équivalence —, sont regroupées dans la première sous-classe.

$\text{Syn}(\text{abeille}) = \text{mouche à miel}$;

$\text{QSyn}(\text{barricade}) = \text{barrage, barrière, défense}$;

$\text{Cf}(\text{abeille}) = \text{frelon, guêpe, bourdon}$;

$\text{=}(ceinture\#I.2) = \sim \text{de sécurité}$.

- La composante sémantique 'conversif', encodée par la FL **Conv_{ij}**, constitue la deuxième sous-classe d'équivalence, par exemple :

$\text{Conv}_{21}(\text{craindre}) = \text{effrayer [N=X]}$.

- La troisième sous-classe regroupe les cas de composantes sémantiques « zéro » liées aux FL dites syntaxiques, qui permettent d'effectuer des reformulations : **S₀**, **V₀**, **A₀** et **Adv₀**. Ces FL servent respectivement à nominaliser, verbaliser, adjectivaliser et adverbialiser une unité lexicale. On peut donc les considérer comme des équivalences sémantiques d'un mot-clé :

$\text{V}_0(\text{acclamation}) = \text{acclamer}$;

$\text{S}_0(\text{applaudir}\#I) = \text{applaudissements}$;

$\text{A}_0(\text{catastrophe}\#I) = \text{catastrophique}$;

$\text{Adv}_0(\text{exagérer}\#2) = \text{exagérément}$.

5.2.2.2 Classe OPPOSITION

Sous la classe OPPOSITION, nous avons regroupé les composantes sémantiques d'opposition. Cette classe comporte trois sous-classes.

- La première sous-classe représente différentes formes d'opposition sémantique : l'antonymie (encodée par la FL *Anti*), l'absence (représentée par la FL *Non*) ainsi qu'une expression contrastive consacrée par l'usage (encodée par la FL *Contr*) :

$Anti(\textit{assentiment}) = \textit{interdiction}$;

$Non(\textit{desir}\#1) = \textit{désintéret}, \textit{indifférence}$;

$Contr(\textit{comptoir}\#I.1b) = \textit{salle}$.

- La seconde sous-classe représente la conversion, déjà illustré plus haut.
- La troisième sous-classe représente les équivalents masculins ou féminins (pouvant correspondre à l'opposition mâle/femelle dans certains cas). Ces composantes sémantiques sont représentées par les FLNS suivantes :

$Fem(\textit{avocat}(1)\#I) = \textit{avocate}$;

$Masc(\textit{vache}) = \textit{taureau}, \textit{bœuf}$.

5.2.2.3 Classe ÉLÉMENT/ENSEMBLE

Nous regroupons, sous la classe ÉLÉMENT/ENSEMBLE, plusieurs sous-classes relativement différentes les unes des autres. Ces sous-classes ont, cependant, toutes en commun une mise en relation entre un élément et son ensemble. Les composantes sémantiques qui y sont représentées fonctionnent par paires, par exemple : 'partie-tout', 'chef-équipe', 'générique-spécifique', etc.

Les deux premières sous-classes regroupent les FL représentant des relations habituellement présentes dans les ontologies de type terminologique :

- il s'agit de composantes servant à classer une unité lexicale dans une hiérarchie, c'est-à-dire le terme générique (représentée par la FL *Gener*) et l'hyponyme (souvent représenté par la FL *Hypo* ou certaines FLNS), par exemple :

$Gener(\textit{avocat}(1)) = \lceil \textit{homme de loi} \rceil$;

$Hypo(\textit{chanson}\#I.1) = \textit{berceuse}$;

Qui sert pour toutes les portes d'un bâtiment(*clé#I.1*) = *passe, passe-partout*.

- Il s'agit également de composantes servant à décrire les parties constitutives d'une entité dénotée par une unité lexicale, soit les méronymes et hollonymes ainsi que les composantes 'ingrédient' et 'matériau typique' :

Mero(*cerise*) = *queue, noyau* ;

Hollo(*train d'atterrissage*) = *avion* ;

Compos(*yaourt#a*) = *lait#b fermenté*.

- On inclut, par ailleurs, dans cette classe, les composantes sémantiques désignant les « relations hiérarchiques entre individus ou animaux ». Bien que ces relations fonctionnent comme une paire, on ne trouve, dans le DiCo, que des occurrences de la relation « descendante ». Celle-ci est encodées par la FLNS Petit de \sim , par exemple :

Petit de \sim (*poule#I*) = *poussin ; poulet*.

- Sont également représentées dans cette classe, la paire de composantes sémantiques 'élément régulier/ensemble de', par exemple :

Sing(*chaîne#I*) = *chaînon, maillon* ;

Mult(*assiette#a*) = *pile* ;

- ainsi que la paire 'chef/équipe', par exemple :

Cap(*majorité#I.3*) = *chef, leader* ;

Equip(*navire*) = *équipage*.

- La dernière sous-classe se situe légèrement à l'écart des précédentes : elle regroupe les composantes sémantiques liées à la mesure, encodées par la FLNS Metr et ses différentes variantes, dont nous avons suggéré la création au chapitre précédent (section 4.3.2.1, p. 127), par exemple :

Metr_{capacité}(*clinique#1*) = *lit* ;

Metr_{capacité}(*restaurant*) = *couvert*.

5.2.2.4 Classe PARTICIPANTS

Nous regroupons sous la classe PARTICIPANTS, les composantes sémantiques liées à l'expression d'un participant de la situation dénotée par le mot-clé. Il s'agit bien souvent des actants du mot-clé, mais l'on retrouve également des participants extérieurs. Nous distinguons donc deux sous-classes.

- La première sous-classe rassemble les composantes sémantiques liées aux actants du mot-clé, par exemple :

$S_1(\text{chanter}(1)\#1a) = \text{chanteur}$;

$\text{Bon}S_{1/2}(\text{débat}\#2) = \text{débatteur}$;

$\text{AntiBon}S_1(\text{compliment}\#a) = \text{flatteur-n}, \text{complimenteur}, \lceil \text{faiseur de } \sim s \rceil$.

$A_1(\text{angoisse}) = \lceil \text{en proie à} \rceil [\text{ART } \sim], \text{ saisi} [\text{de (ART) } \sim], \text{ angoissé}$;

$\text{Able}_1(\text{bagarre}\#a) = \text{bagarreur-adj}$;

$\text{Qual}_2(\text{haine}) = \text{odieux}$.

- La seconde sous-classe regroupe les composantes identifiant les participants extérieurs, dont ceux représentés par les FL de la sous-classe ‘Chef/équipe’ décrite plus haut, par exemple :

Individu qui a l'habitude d'essayer de se faire offrir à manger par les gens,
contre leur gré($\text{assiette}\#a$) = *pique-assiette* ;

Individu qui aide la femme Y au moment de $\sim(\text{naissance}\#I.1) = \text{accou-}$
cheuse, sage-femme ;

$\text{Cap}(\text{majorité}\#I.3) = \text{chef, leader}$;

$\text{Equip}(\text{navire}) = \text{équipage}$.

5.2.2.5 Classe QUALIFICATIFS

Le terme *qualificatif* doit être ici interprété au sens large. Il ne s'agit pas uniquement d'adjectifs qualificatifs. Certaines FL nominales ou verbales peuvent tout à fait entrer en ligne de compte si elles sous-entendent un jugement ou une caractérisation du mot-clé ou de ses participants. La classe intensité recouvre quatre sous-classes.

- La première sous-classe regroupe les composantes sémantiques qui ont trait au jugement

- qu'il soit positif, par exemple :

$\text{VerFact}_0(\text{montre}) = \text{\textit{\textit{être exacte, fonctionner bien}}}$;

$\text{BonS}_{1/2}(\text{débat\#2}) = \text{\textit{\textit{débatteur}}}$;

- ou négatif, par exemple :

$\text{AntiVer}(\text{piste\#II}) = \text{\textit{\textit{vague, embrouillée}}}$;

$\text{AntiBonS}_1(\text{théâtre\#I.1}) = \text{\textit{\textit{cabotin, théâtraux}}}$.

– La deuxième sous-classe recouvre les qualificatifs ayant trait à l'intensité,

- qu'il s'agisse d'une forte intensité, par exemple :

$\text{Magn}(\text{aboitement\#I}) = \text{\textit{\textit{fort, furieux, féroce}}}$;

$\text{Magn.IncepOper}_1(\text{abattement(1)}) = \text{\textit{\textit{sombrer}}}$ [dans ART \sim] ;

$\text{CausPredPlus}(\text{crainte}) = \text{\textit{\textit{attiser, exciter}}}$ [ART \sim] ;

$\text{Culm}(\text{haine}) = \text{\textit{\textit{paroxysme}}}$ [de ART \sim] ;

- ou d'une faible intensité, par exemple :

$\text{AntiMagn}(\text{pécule}) = \text{\textit{\textit{modeste, petit}}}$;

$\text{AntiBon.AntiMagn.Oper}_1(\text{conversation\#b}) = \text{\textit{\textit{avoir peu, ne pas avoir beaucoup}}}$ [de \sim] ;

$\text{CausPredMinus}(\text{fardeau\#II}) = \text{\textit{\textit{alléger}}}$ [ART \sim].

– La troisième sous-classe regroupe les composantes sémantiques concernant l'apparence physique d'une entité dénotée par le mot-clé. De nombreuses FLNS pointent vers cette sous-classe. Nous les organisons en fonction des attributs élaborés au chapitre précédent (section 4.5), soit : taille, forme, couleur, matériau, etc. On note, par exemple, les FLNS suivantes décrivant les différentes formes de jupe :

Forme : Qui descend jusqu'aux mollets de $X(\text{jupe}) = \text{\textit{\textit{à mi-mollet}}}$;

Forme : Qui descend jusqu'aux genoux de $X(\text{jupe}) = \text{\textit{\textit{au genoux}}}$;

Forme : Qui a des plis comme décoration($\text{jupe}) = \text{\textit{\textit{à plis, froncée, plissée}}}$.

5.2.2.6 Classe PHASE/ASPECT

Cette classe regroupe les composantes sémantiques ayant trait au déroulement d'un fait. Nous avons distingué sept sous-classes.

- Les quatre premières sous-classes correspondent à différentes phases d'une action ou d'un événement, soit, la préparation, le début, le milieu, la fin, par exemple :

$\text{Prepar}_1(\textit{assassinat}) = \textit{comploter, préparer}$;

$\text{IncepFunc}_2(\textit{abcès}\#\text{I}) = \textit{se développer, se former, naître}$ [Loc-in N=Y] ;

$\text{Centr}(\textit{débat}\#\text{1}) = \textit{centre, cœur}$ [de ART \sim] ;

$\text{FinOper}_1(\textit{abîme}\#\text{II.1}) = \textit{sortir}$ [de ART \sim].

- Les cinquième et sixième sous-classes regroupent les composantes sémantiques exprimant un aspect tel que le résultat ou la répétition, par exemple :

$\text{S}_{\text{res}}(\textit{tirage}(1)\#\text{1}) = \textit{copie, exemplaire, } \lceil \textit{tiré à part} \rceil$;

$\text{De_nouveauOper}_1(\textit{ami}\#\text{I.1}) = \textit{redevenir}$ [ART \sim].

- Enfin, la septième sous-classe comporte les composantes sémantiques ayant trait à l'expression de la durée d'une action ou d'un événement, par exemple :

$\text{Magn}^{\text{temp}}(\textit{applaudissements}) = \textit{prolongés}$.

5.2.2.7 Classe MANIÈRE

La classe MANIÈRE regroupe les composantes sémantiques liées à la manière au sens large, c'est-à-dire autant au mode de déroulement d'un fait qu'aux instruments permettant le déroulement de ce fait. Cette classe contient trois sous-classes :

- la première a trait aux instruments, aux moyens et façons permettant le déroulement d'un fait. Ces composantes sémantiques sont, le plus souvent, représentées par les trois FL de dérivation sémantique circonstancielle S_{instr} , S_{med} et S_{mod} , par exemple :

$\text{S}_{\text{instr}}(\textit{crime}\#\text{I.b}) = \textit{arme}$ [de ART \sim] ;

$\text{S}_{\text{med}}(\textit{excuse}\#\text{II}) = \textit{lettre, mot}$ [d' \sim s] ;

$\text{S}_{\text{mod}}(\textit{vie}\#\text{I.2}) = \textit{genre, mode, style}$ [de \sim].

- On trouve également, dans cette classe, des composantes sémantiques indiquant la façon dont l’entité dénotée par le mot-clé est utilisée ou exprimée :

$\text{Instr}(\text{couteau}) = \text{au } [\sim]; \text{ avec } [\text{ART } \sim]; \text{ } \ulcorner \text{à coups de} \urcorner [\sim];$

$\text{Adv}_1(\text{gratitude}) = \text{avec } [\sim];$

$\text{NonAdv}_1(\text{animosité}) = \text{sans } [\sim];$

$\text{Adv}_1\text{Real}_1(\text{flèche}\#\text{III}) = \text{dans le sens } [\text{de ART } \sim];$

En regardant à travers l’orifice de $\sim(\text{serrure}) = \text{par le trou de la } \sim.$

- Enfin, cette classe regroupe diverses composantes sémantiques (encodées par des FL standard ou non standard) indiquant les différentes manières dont se déroule l’action ou l’événement dénoté par le mot-clé, par exemple :

$\text{Adv}_0\text{NonManif}(\text{admirer}\#\text{I}) = \text{secrètement};$

En s’asphyxiant avec du gaz(se suicider) = *au gaz.*

5.2.2.8 Classe CAUSE

La classe CAUSE rassemble les composantes sémantiques exprimant la cause, par exemple :

$\text{Propt}(\text{animosité}) = \text{par } [\sim];$

Dont la raison est l’amour($\text{crime}\#\text{I.b}$) = *passionnel.*

5.2.2.9 Classe LOCALISATION

La classe LOCALISATION regroupe toutes les composantes sémantiques de localisation, qu’elles soient spatiales ou temporelles. On distingue deux sous-classes :

- une première, qui regroupe les composantes sémantiques liées au lieu où se déroule un fait, par exemple :

$\text{S}_{\text{loc}}\text{Fact}_0(\text{avocat}(1)\#\text{I}) = \text{cabinet, cour};$

$\text{S}_{\text{loc}}\text{PreparOper}_1(\text{chant}\#\text{I.a}) = \text{conservatoire, école.}$

- la deuxième sous-classe concerne la localisation spatiale et temporelle. Elle inclut les quatre premières sous-classes de la classe PHASE/ASPECT, indiquant les phases de déroulement d’un fait, soit la préparation, le début, le milieu et la fin,

ainsi que d'autres sous-classes permettant de positionner dans l'espace ou dans le temps ce qui est dénoté par le mot-clé, par exemple :

Préparation : $\text{PreparReal}_1(\text{balle}\#\text{I}) = \textit{gonfler}$ [ART \sim];

Début : $\text{IncepFunc}_2(\text{abcès}\#\text{I}) = \textit{se développer, se former}$ [Loc-in N=Y];

Milieu : $\text{Culm}(\textit{haine}) = \textit{paroxysme}$;

Milieu : $\text{Loc}_{\text{in}}(\textit{bataille}\#\text{I.1}) = \textit{durant, pendant, } \ulcorner \textit{au cours de} \urcorner$ [ART \sim];

Fin : $\text{S}_0\text{FinFunc}_0(\textit{hostilité}\#\text{III}) = \textit{arrêt, fin}$ [de ART \sim s].

Dans : $\text{Loc}_{\text{in}}(\textit{bateau}) = \ulcorner \textit{à bord de} \urcorner, \textit{dans, sur}$ [ART \sim];

Depuis : $\text{Loc}_{\text{ab}}(\textit{balcon}) = \textit{de, depuis}$ [ART \sim]; $\textit{du haut}$ [de ART \sim].

5.2.2.10 Classe ACTION/ÉVÉNEMENT

La classe ACTION/ÉVÉNEMENT regroupe les composantes sémantiques ayant trait à une action, un événement ou une activité. Les FL pointant vers cette classe, c'est-à-dire, dont les composantes sémantiques sont répertoriées dans cette classe, sont généralement, mais non exclusivement, verbales. Nous divisons cette classe en un certain nombre de sous-classes :

- Premièrement, la sous classe 'création/apparition' regroupe les composantes liées à l'apparition ou à la création d'un fait ou d'une entité, par exemple :

$\text{CausFunc}_0(\textit{bijou}\#\text{I}) = \textit{ciseler, créer, fabriquer, monter}$ [ART \sim];

$\text{S}_0\text{CausFunc}_0(\textit{bateau}) = \textit{lancement}$ [de ART \sim].

- Une deuxième sous-classe rassemble les composantes sémantiques exprimant une forme de manifestation, qu'il s'agisse de l'expression d'un symptôme (Sympt_i) ou de la manifestation d'un sentiment ou d'une émotion (Manif , $\text{Perm}_1\text{Manif}$), par exemple :

$\text{Sympt}_1(\textit{émoi}) = \textit{frémir, palpiter, trembler}$ [d' \sim];

$\text{Manif}(\textit{pleurs}) = \textit{couler, jaillir}$.

- La troisième sous-classe permet de regrouper les composantes sémantiques de réalisation, fonctionnement ou utilisation, illustrées à plusieurs reprises dans le chapitre, par exemple :

$\text{Real}_2(\text{avocat}(1)\#\text{I}$ ‘individu qui pratique un métier’) = *consulter*

[ART \sim]; *utiliser les services* [de ART \sim];

$\text{Fact}_0(\text{ancre}) = \textit{mordre, tenir}$;

$\text{Involv}(\text{plaque}\#\text{I.1}$ ‘accumulation de matière’) = *recouvrir* [N];

$\text{Son}(\text{abeille}) = \textit{bourdonner}$.

- La quatrième sous-classe regroupe les composantes sémantiques d’augmentation ou d’amélioration, par exemple :

$\text{CausPredPlus}(\text{crainte}) = \textit{attiser, exciter}$ [ART \sim].

- La cinquième sous-classe, complémentaire de la précédente, regroupe les composantes sémantiques exprimant une diminution ou une dégradation, soit, par exemple :

$\text{CausPredMinus}(\text{fardeau}\#\text{II}) = \textit{alléger}$ [ART \sim].

- On note une autre sous-classe regroupant les composantes sémantiques ayant trait à la destruction, la fin ou la disparition d’un fait ou d’une entité, par exemple :

$\text{LiquOper}_1(\text{pleurs}) = \textit{sécher, tarir}$ [ART \sim s];

$\text{FinFunc}_0(\text{barrière}\#\text{II.1}) = \textit{tomber, voler}$.

- La septième sous-classe regroupe les composantes sémantiques ayant trait au non-fonctionnement d’une entité, à la non réalisation d’un événement ou encore à l’interdiction ou au refus. Ces composantes sont illustrées à travers les FL suivantes :

$\text{NonPermFact}_0(\text{zone}) = \textit{boucler}$ [ART \sim]; *interdire* [ART \sim à N=@];

$\text{NonFact}_0(\text{usine}) = \textit{chômer, être arrêtée, être en grève}$;

$\text{AntiReal}_2(\text{ultimatum}) = \textit{refuser, rejeter, repousser}$ [ART \sim];

$\text{NonOper}_1(\text{tact}) = \textit{manquer}$ [de \sim].

- Enfin, dans la dernière sous-classe sont rassemblées les composantes sémantiques exprimant une tentative ou un essai. Par exemple :

$\text{Essayer de Oper}_{12}(\text{victoire}\#\text{I}) = \textit{disputer}$ [ART \sim à N=Y];

$\text{Essayer de IncepReal}_1(\text{fauteuil}) = \textit{briguer}$ [ART \sim].

L'ensemble des classes proposées est récapitulé dans la figure 5.5 (p. 188).

5.2.3 Bilan et illustration

Les dix classes qui viennent d'être présentées permettent d'organiser l'ensemble des FL d'un point de vue sémantique. Cette représentation est novatrice dans la mesure où elle permet :

- d'organiser les FL sur le plan sémantique, indépendamment de la dimension combinatoire et des parties du discours des lexies en mots-clés et valeurs ;
- d'accéder aux FL de façon la plus intuitive possible grâce à une structuration souple des données : les sous-classes ne sont pas figées à l'intérieur d'une classe donnée, mais peuvent se trouver sous plusieurs classes ;
- de mettre sur le même plan les FL standard (simples et combinées), semi-standard, localement standard et non standard.

Cette organisation sémantique a un rôle majeur à jouer dans une optique de recherche d'informations lexicales. Un des objectifs du système de FL est, en effet, de servir à la confection d'interfaces de consultation dans une base de données lexicale. De cette façon, l'organisation sémantique des FL peut donner lieu à une interface sémantique, dont le rôle est de fournir, à partir d'une entrée de dictionnaire, une classification de ses collocatifs et dérivés sémantiques selon des classes sémantiques. Nous illustrons ce type d'interface à l'aide de la lexie BATEAU dans la figure 5.6.

- ▼ ÉQUIVALENCE
 - ▷ Lexies similaires
 - ▷ Conversifs
 - ▷ Reformulation syntaxique
- ▼ OPPOSITION
 - ▷ Contraires
 - ▷ Conversifs
 - ▷ Équivalent masculin/féminin
- ▼ ÉLÉMENT/ENSEMBLE
 - ▷ Hiérarchiques
 - ▷ Partie/tout
 - ▷ « Famille »
 - ▷ Unité régulière/ensemble régulier
 - ▷ Chef/équipe
 - ▷ Mesure
- ▼ PARTICIPANTS
 - ▷ Actants du mot-clé
 - ▷ Participants extérieurs
- ▼ QUALIFICATIFS
 - ▷ Qualificatif de jugement
 - jugement positif
 - jugement négatif
 - ▷ Qualificatif d'intensité
 - Forte intensité
 - Faible intensité
 - ▷ Autres qualificatifs portant sur l'apparence physique
- ▶ CAUSE
- ▼ PHASES/ASPECT
 - ▷ Préparation
 - ▷ Début
 - ▷ Milieu
 - ▷ Fin
 - ▷ Résultat
 - ▷ Répétition
 - ▷ Expression de la durée
- ▼ MANIÈRE
 - ▷ Instrument/moyen/manière
 - ▷ Manière dont l'entité ~ est utilisée/exprimée
 - ▷ Manière de réaliser l'action/événement ~
- ▼ LOCALISATION
 - ▷ Lieu typique
 - ▷ Positionnement spatial/temporel
- ▼ ACTION/ÉVÉNEMENT
 - ▷ Création/causation d'existence
 - ▷ Manifestation
 - ▷ Utilisation/fonctionnement typique
 - ▷ Augmentation/amélioration
 - ▷ Diminution/dégradation
 - ▷ Disparition/Fin d'existence
 - ▷ Non-fonctionnement/Refus
 - ▷ Tentative

Figure 5.5 – Liste des classes sémantiques

BATEAU

▼ ÉQUIVALENCE

QSyn *navire*, *bâtiment*, *vaisseau*, [engin de navigation]¹
Cf *barque*, *embarcation*, *esquif*, *radeau*

▼ ÉLÉMENT/ENSEMBLE

▼ Hyponymes

Hypo *bac*, *bateau-mouche*, *bateau-citerne*, *barge*, *brise-glace*, *cargo*, *chalutier*, *drakkar*, *ferry*, *frégate*, *galère*, *jonque*, *paquebot*, *péniche*, *pétrolier*, *transbordeur*, *yacht*

Qui sert à la pêche de pêche

Qui sert à faire des croisières de croisière

Qui a un usage non commercial de plaisance

Qui sert à transporter Y d'un côté à l'autre de l'Atlantique transatlantique-adj, // *transatlantique-n*

Qui est utilisé par des pirates *pirate#1*

Qui Fact₀ grâce à N à [N]

▼ Méronymes

Mero *bord*, *coque*, *quille*, *passerelle*, *pont*, *cabine*, *carré des officiers*, *carré*, *cockpit*, *mess*, *poste d'équipage*, *écoutille*, *hublot*, *cale*, *ballast*, *soute*, *mât*, *cheminée*

à l'avant Mero *proue*, *étrave*

à l'arrière Mero *étambot*, *poupe*

à droite Mero *tribord*

sculpture à l'avant Mero [figure de proue]¹

Dispositif à bord de ~ servant à émettre un fort signal sonore *sirène*

Dispositif servant à diriger ~ *gouvernail*, *barre*

Instrument servant à immobiliser ~ *ancre*, *amarre*

Passerelle servant à accéder à ~ *passerelle (d'embarquement)*

Drapeau utilisé sur ~ *pavillon*

Petite embarcation de secours transportée sur ~ *canot de sauvetage*

Charge que l'on met sur ~ pour le stabiliser *lest*

▼ Ensemble régulier

MultS₁/Equip *équipage*

Mult *flottille* [de ~x], *flotte*

▼ Mesure

Quantité de marchandise que ~ peut transporter *capacité*, *tonnage* [de ART ~]

Mesure de la taille de ~ *calaison*, *déplacement*, [tirant d'eau]¹ [de ART ~]

Unité de mesure pour la vitesse de ~ *nœud*

▼ PARTICIPANTS

▼ Dérivés nominaux actantiels

S₁ *membre d'équipage*, *marin*, *matelot*, *mousse*, *timonier*, *pilote*

MultS₁/Equip *équipage*

S₂ *passager*, *cargaison*, *chargement*

AntiBonS₂ *passager clandestin*

▼ Participants extérieurs

Cap *capitaine* [de ART ~], *commandant* [de ART ~]

Propriétaire de ~ qui en fait l'exploitation commerciale *armateur*

Entreprise faisant du transport par ~ *compagnie de navigation*, *compagnie maritime*, *société maritime*

- ▼ **QUALIFICATIFS**
- ▼ **Qualification d'intensité**
 - forte intensité
 - Magn *grand, gros, énorme*
 - faible intensité
 - AntiMagn *petit, « coquille de noix »*
 - ▼ **Qualification négative**
 - AntiBon *mauvais, // rafiot*
 - B. se retourne sur lui-même *chavirer*
 - AntiBonS₂ *passager clandestin*
- ▼ **ACTION/ÉVÉNEMENT**
- ▼ **Création/causation d'existence**
 - CausFunc₀ *lancer, « mettre à l'eau »* [ART ~]
 - S₀CausFunc₀ *lancement* [de ART ~]
 - ▼ **Fonctionnement/Utilisation**
 - IncepFact₀ *appareiller, « larguer les amarres », « lever l'ancre », quitter [le port], « prendre la mer », « prendre le large », « mettre le cap »* [sur N]
 - individu Y Prepar₁Real₂ *prendre* [ART ~], *s'embarquer, embarquer, monter* [« à bord de »/dans/sur ART ~], *monter* [« à bord de »]
 - S₀ individu Y Prepar₁Real₂ *embarquement*
 - cargaison Y PreparReal₂ *charger, embarquer* [N=Y « à bord de »/dans/sur ART ~]
 - S₀ cargaison Y PreparReal₂ *chargement, embarquement* [de N=Y]
 - LiquFinAble1Fact₀ « remettre à flot », *renflouer* [ART ~]
 - Real₁ *naviguer* [sur ART ~]; *piloter* [ART ~]
- individu Y Real₂ *aller, se déplacer, voyager* [en/par ~], *prendre* [le ~] étendue d'eau © Fact© *naviguer, voguer* [Loc-in N=©], *parcourir, sillonner* [N=©]
- endroit © Fact© *naviguer, voguer, cingler* [« en direction de »/vers N=©]
- S₀ endroit © Fact© *traversée*
- Fact₂ *transporter* [N=Y]
- Transport effectué par ~ *transport maritime*
- ~ se positionne le long d'un autre ~ *aborder* [N]
- ~ oscille *se balancer, tanguer*
- FinFact₀-I *arriver* [au port], *toucher* [Nom-propre <le port>]
- FinFact₀-II *accoster, « jeter l'ancre », mouiller, « faire escale »* [Loc-in N]
- Enregistrement officiel de ~ *immatriculation* [de ART ~]
- ▼ **Dégradation**
- AntiVerFact₀/AntiBonFact₀ *dérriver, partir à la dérive*
 - ~ se retourne sur lui-même *chavirer*
 - IncepFinFunc₀ « prendre l'eau »
- ▼ **Disparition/Fin d'existence**
- FinAble1Fact₀ (*s'*) *échouer* [sur N]
 - S₀FinAble1Fact₀ *échouage* [sur N]
 - CausIFinAble1Fact₀ *échouer* [ART ~ sur N]
 - FinFunc₀ *couler, faire naufrage, sombrer*
 - S₀FinFunc₀ *naufrage*
 - CausFinAble1Fact₀ *couler, « envoyer par le fond », faire sombrer* [ART ~]

<p>▼ LOCALISATION</p> <p>Loc_m \lceil à bord de \lceil [ART \sim]; dans, sur [ART \sim] Endroit où l'on construit \sim \lceil chantier naval \lceil Endroit où l'on répare \sim cale, \lceil cale sèche \lceil SlocFinFact₀[!] port, rade, \lceil base navale \lceil, dock, gare#b fluviale, gare#b maritime construction SlocFinFact₀[!] appontement, débarcadère, embarcadère, jetée, quai, wharf</p>	<p>▼ Début</p> <p>IncepFact₀ appareiller, \lceil larguer les amarres \lceil, \lceil lever l'ancre \lceil, quitter [le port], \lceil prendre la mer \lceil, \lceil prendre le large \lceil, \lceil mettre le cap \lceil [sur N]</p> <p>▼ Fin</p> <p>FinFact₀[!] arriver [au port], toucher [Nom-propre <le port>] FinFact₀[!] accoster, \lceil jeter l'ancre \lceil, mouiller, \lceil faire escale \lceil [Loc-in N] FinAbleIFact₀ (s') échouer [sur N] S0FinAbleIFact₀ échouer [sur N] CausIFinAbleIFact₀ échouer [ART \sim sur N] FinFunc₀ couler, faire naufrage, sombrer S0FinFunc₀ naufrage CausFinAble₁Fact₀ couler, \lceil envoyer par le fond \lceil, faire sombrer [ART \sim]</p> <p>▼ Réitération</p> <p>LiquFinAble₁Fact₀ \lceil remettre à flot \lceil, renflouer [ART \sim]</p> <p>▼ Résultat</p> <p>Ce qui reste de \sim après FinFunc₀ épave [de ART \sim]</p>
<p>▼ PHASE/ASPECTS</p> <p>▼ Préparation</p> <p>CausFunc₀ lancer, \lceil mettre à l'eau \lceil [ART \sim] S0CausFunc₀ lancement [de ART \sim] individu Y Prepar1Real2 prendre [ART \sim], s'embarquer, embarquer, monter \lceil à bord de \lceil/dans/sur ART \sim, monter \lceil à bord \lceil S0 individu Y Prepar1Real2 embarquement cargaison Y PreparReal2 charger, embarquer [N=Y \lceil à bord de \lceil/dans/sur ART \sim] S0 cargaison Y PreparReal2 chargement, embarquement [de N=Y]</p>	

Figure 5.6 – Relations lexicales de BATEAU organisées par classes sémantiques

5.3 Perspective combinatoire sur le système de fonctions lexicales

5.3.1 Objectif

Cette section présente les objectifs visés pour l'organisation des FL en fonction de leurs propriétés combinatoires. Nous envisageons ici une relation entre un mot-clé et sa valeur (par le biais d'une FL) du point de vue de leur configuration syntaxique l'un vis-à-vis de l'autre. Une telle organisation permet un accès aux patrons de dépendance syntaxique indépendamment de toute autre information.

Classifier les FL en fonction de leur patron de dépendance syntaxique facilite la sélection d'une FL lors d'une requête sur les différentes relations associées à une lexie donnée. Par exemple, sous la lexie MEURTRE, on observe les relations syntagmatiques suivantes :

- Syntagmes dans lesquels MEURTRE subit une modification syntaxique :

Magn atroce, affreux, brutal, horrible, inqualifiable, odieux ;
A₂Prepar₁ avec préméditation, prémédité.

- Syntagmes dans lesquels MEURTRE est un actant syntaxique :

Oper₁ accomplir, commettre [ART ~] ; tremper [dans ART ~] ;
CausOper₁ pousser [N=X au ~] ;
Real^I_@ enquêter [sur ART ~]² ;
Real^{II}_@ élucider [ART ~] ;
Real^{III}_@ punir, châtier, venger [ART ~] ;
Prepar₁ comploter, préparer, préméditer [ART ~].

Ce type d'organisation des collocations a une grande portée pédagogique dans le cadre d'une application d'aide à la rédaction. Un utilisateur peut, par exemple, chercher quel collocatif conviendrait le mieux pour s'insérer dans une configuration syntaxique souhaitée.

Notre objectif est d'élaborer la liste des différents patrons de dépendance liés aux

²Les exposants I, II et III renvoient aux différentes phases de l'action encodée par la FL Real_@ ; @ désigne un participant extérieur, c'est-à-dire qui n'est pas actant de la lexie MEURTRE.

FL. Nous devons toutefois, dans un premier temps, resituer les composantes syntaxiques à l'intérieur du modèle Sens-Texte et préciser les différentes relations syntaxiques prises en compte (section 5.3.2). Par la suite, nous présenterons la méthodologie adoptée pour dégager les différents patrons de dépendance modélisés par les FL ainsi que la liste de ces patrons elle-même (section 5.3.3).

5.3.2 Les relations syntaxiques modélisées par les fonctions lexicales

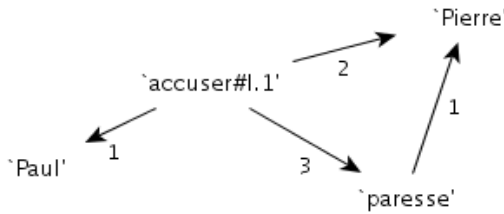
Dans le modèle Sens-Texte, sept niveaux de représentation linguistique se succèdent (cf. figure 5.7 ci-dessous). À ces représentations s'ajoutent normalement des structures communicatives et rhétoriques, qui représentent les oppositions du type thème-rhème, explicite-sous-entendu, etc., dont nous ne tiendrons pas compte ici. Dans cette section, nous ne détaillons que les éléments pertinents pour comprendre le fonctionnement des FL, soit les trois premiers niveaux de représentation. Nous renvoyons le lecteur aux textes introductifs tels que Mel'čuk [1997], Polguère [1998] ou Kahane [2003a] pour une présentation plus détaillée de la Théorie Sens Texte.

Représentation sémantique
Représentation syntaxique profonde
Représentation syntaxique de surface
Représentation morphologique profonde
Représentation morphologique de surface
Représentation phonologique
Représentation phonétique

Figure 5.7 – Niveaux de représentation de la Théorie Sens-Texte

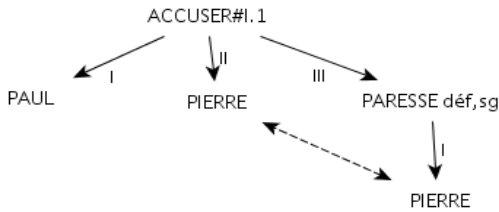
Nous nous servons de la phrase *Paul accuse#I.1 Pierre de paresse*, pour illustrer ces trois niveaux. Le premier niveau, constituant la **représentation sémantique**, décrit la structure sémantique d'un ensemble potentiel de phrases quasiment synonymes. Il s'agit d'un graphe connexe orienté, dont les nœuds sont étiquetés par des unités sémantiques correspondant à des acceptions d'unités lexicales. Certaines de

ces unités sémantiques sont des prédicats (comme le verbe ACCUSER#.1 et le substantif PARESSE), d'autres, des noms sémantiques (par exemple : PAUL, CAILLOU). Les arcs du graphe sont étiquetés par des numéros distinctifs spécifiant les différents actants sémantiques du prédicat. Par exemple, le verbe ACCUSER#I.1 contrôle trois positions actantielles : *X accuse Y du fait Z*. On représentera donc la phrase *Paul accuse#I.1 Pierre de paresse*, de la façon suivante :

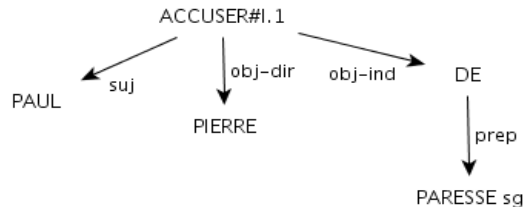


Un deuxième niveau de **représentation syntaxique profonde** représente les relations syntaxiques profondes entre les éléments sous la forme d'un arbre de dépendance non ordonné. Les relations syntaxiques profondes sont universelles (Mel'čuk 1997). Il s'agit des six relations actantielles (I, II, III, IV, V et VI), de la relation attributive (ATTR), qui couvre les cas de modification au sens large, de la relation coordinative (COORD), représentant toutes les formes de coordination ainsi que de la relation appenditive (APPEND) pour les éléments de la phrase tels que les interjections, les formes d'adresse et les adverbes de phrase. Nous ne retenons que les relations actantielles et la relation attributive, les relations coordinatives et appenditives n'étant pas concernées par notre problématique.

L'arbre de dépendance a pour nœuds des unités lexicales. Les arcs de l'arbre représentent des relations syntaxiques. Notons que l'on ne tient pas compte, à ce niveau, des prépositions régies ni des éventuelles formes de flexion et de pronominalisation. On observe dans l'exemple ci-dessous, que les actants syntaxiques profonds I, II et III de ACCUSER#I.1 sont en correspondance avec les actants sémantiques X, Y et Z de 'accuser#I.1'.



Le niveau de **représentation syntaxique de surface** présente la façon dont les éléments sont employés dans une phrase sous la forme d'un arbre de dépendance non ordonné. C'est à ce niveau qu'apparaissent les relations syntaxiques de sujet, complément d'objet direct, complément d'objet indirect, ainsi que les prépositions.



5.3.3 Établissement des différents patrons de dépendance syntaxique associés aux fonctions lexicales

5.3.3.1 Utilisation de l'encodage explicite

Afin d'établir une liste des différents patrons syntaxiques des relations lexicales syntagmatiques, nous procédons à la conversion de l'encodage traditionnel des FL en encodage explicite (Kahane et Polguère 2001). Cet encodage a déjà été présenté au chapitre 3 et illustré dans la figure 3.7 (p. 106). Rappelons qu'il vise à expliciter le contenu sémantique d'une fonction ainsi que sa structure syntaxique. Cette dernière est exprimée au moyen de deux éléments : la partie du discours et la valence active, c'est-à-dire la liste des dépendants syntaxiques par ordre croissant d'oblicité.

Par exemple, la FL Fact_1 est transcrite selon l'encodage explicite syntaxique de la manière suivante : $V[\#,1]$. « V » indique que la valeur est verbale, « # » en première position, qu'elle a pour premier actant syntaxique profond le mot-clé lui-même et

« 1 » en seconde position, qu'elle a pour second actant syntaxique profond le premier actant syntaxique profond du mot-clé. Par exemple, la forme propositionnelle de la lexie ABAT-JOUR étant la suivante :

~ MIS SUR L'artefact X, on a comme valeur de Fact_1 : $\text{Fact}_1(\text{abat-jour}) = \text{coiffer}$ [N=X], que l'on peut illustrer dans la figure 5.8.

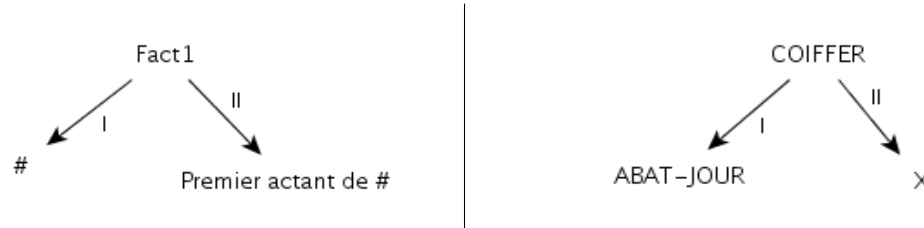


Figure 5.8 – Valence de Fact_1

Nous ne conservons de l'encodage explicite que la partie syntaxique sans l'indication de partie du discours, dans la mesure où nous souhaitons l'isoler des autres types d'information contenus dans les FL. Précisons à présent certaines informations pour la lecture de l'encodage explicite :

- # représente le mot-clé ;
- **1, 2, 3, 4, ...** représentent les actants syntaxiques profonds I, II, III, IV, ... du mot-clé ;
- @ représente un actant extérieur à la structure actantielle du mot-clé, par exemple, dans la FL suivante : $\text{Fact}_@(\text{cri}\#\text{I.1}) = \text{parvenir}$ [à N], le complément du verbe représente un participant de la situation qui ne correspond toutefois pas à un actant du mot-clé.
- ^ est un signe de modification : le symbole précédé de ^ est l'élément modifié par la valeur retournée par la FL. Par exemple, pour la FL A_1 , ($A_1(\text{haine}) = \text{haineux}$), on aura la formule [1^], qui indique que l'application de la FL modifie le premier actant du mot-clé, soit X dans sa structure actantielle de *haine* : ~ DE LA personne OU DE L'animal X ENVERS Y.

5.3.3.2 Identification des différents patrons de dépendance syntaxique liés aux liens lexicaux encodés par les FL

L'analyse des liens lexicaux syntagmatiques du DiCo encodés par les FL a permis de mettre en évidence quarante et un patrons de dépendance syntaxique. Nous les avons triés et classifiés selon deux critères : le type de relation (attributive ou actantielle) et le nombre d'actants en jeu, dans le cas des relations actantielles. De la même façon que pour l'organisation sémantique des FL, nous avons regroupé l'ensemble des patrons obtenus dans des classes, dont chaque patron constitue une sous-classe et vers lesquelles peuvent pointer les FL (plus précisément, l'ensemble des FL syntagmatiques et certaines FL paradigmatiques (cf. p. 200)). Nous illustrons ci-dessous les différentes classes établies à l'aide d'un exemple de patron. Notons qu'une liste exhaustive et illustrée est fournie à la figure 5.9 (p. 203).

Classe 1 : Structures actantielles à un seul actant

La première classe regroupe les structures actantielles à un actant. L'actant en question est le mot-clé de la FL représentant la relation. Ce patron est représenté par la formule : [#]. Les FL Degrad ou Func₀, par exemple, pointent vers cette classe :

Degrad(*lait*#b) = *tourner* ;

Func₀(*commérage*#a) = *circuler*.

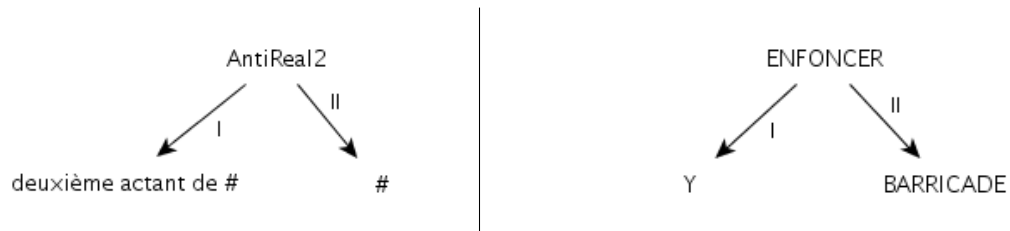


Classe 2 : Structures actantielles à deux actants

Dans la deuxième classe figurent les patrons représentant des structures actantielles à deux actants. On observe plusieurs configurations possibles. Les FL Func_1 , Fact_1 , par exemple, sont associées au patron suivant : $[\#,1]$; les FL Func_2 , Fact_2 , au patron $[\#,2]$; Oper_1 , Real_1 sont associées au patron $[1,\#]$, etc. Prenons l'exemple de la FL AntiReal_2 (\rightarrow ne pas réaliser l'action inhérente à \sim) illustrée ci-dessous :

$\text{AntiReal}_2(\text{barricade}) = \text{enfoncer, franchir, forcer}$ [ART \sim].

On lui associe le patron $[2,\#]$ que l'on peut illustrer de la façon suivante :



Avant de poursuivre vers les structures actantielles à trois actants, nous devons préciser quelques aspects de certaines constructions à verbe support. Comme l'a décrit Alonso Ramos [2007], la diathèse de certaines FL modélisant des verbes supports, c'est-à-dire la correspondance entre leur actants sémantiques et leurs actants syntaxiques profonds, n'est pas prévisible. Selon les valeurs qu'elles retournent, leur nombre d'actants syntaxiques profonds diffère. Prenons l'exemple des relations (110) et (111) pour la lexie INTERROGATOIRE, dont la structure actantielle est la suivante : \sim PAR LA personne X DE L'individu Y À PROPOS DE Z.

(110) $\text{Oper}_1(\text{interrogatoire}) = \text{conduire, effectuer, mener}$ [ART \sim auprès de N=];

(111) $\text{Oper}_1(\text{interrogatoire}) = \text{faire subir}$ [ART \sim à N=Y].

En (110), les valeurs sont bi-actantielles, elles prennent pour actant syntaxique profond le premier actant syntaxique profond du mot-clé et, pour second actant, le mot-clé lui-même. Le deuxième actant de INTERROGATOIRE n'est pas un actant syntaxique profond des valeurs de la FL, mais dépend de son mot-clé. En (111), la valeur est tri-actantielle : elle a les deux mêmes premiers actants syntaxiques profonds que les valeurs en (110) ainsi qu'un troisième actant syntaxique profond

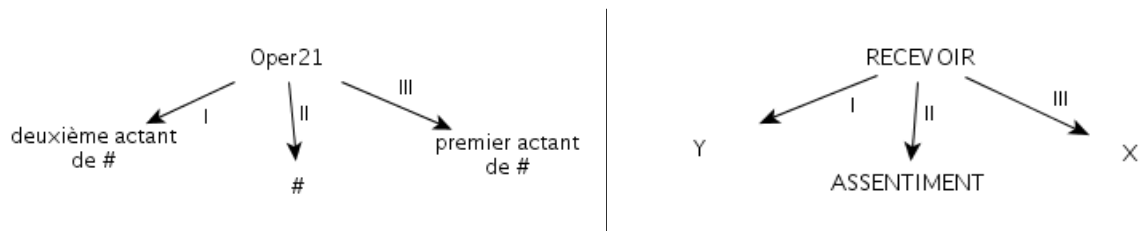
correspondant au deuxième actant syntaxique profond du mot-clé. Au niveau syntaxique profond, les deux relations seraient représentées par deux arbres distincts. Cette observation a conduit à augmenter la granularité de cette FL en proposant une FL Oper_{12} pour représenter les cas où, comme en (111), la valeur prend l'actant j du mot-clé pour deuxième complément. Les différentes FL Oper pointent donc vers des classes différentes selon le type de valeurs qu'elles retournent : alors que les Oper_i pointent vers la classe 2, les Oper_{ij} pointent vers la classe 3.

Classe 3 : Structures actantielles à trois actants

Les relations actantielles à trois actants sont regroupées dans la troisième classe. On observe un grand nombre de patrons avec plusieurs permutations possibles du mot-clé et de ses différents actants, par exemple : $[\#,1,2]$ pour les FL Fact_{12} et Func_{12} ; $[2,1,\#]$ pour les FL Labreal_{21} et Labor_{21} , etc.

On associe à la FL Oper_{21} , par exemple, le patron $[2,\#,1]$ que l'on peut illustrer comme suit :

$\text{Oper}_{21}(\textit{assentiment}) = \textit{obtenir}, \textit{recevoir}, \textit{recueillir}$ [ART \sim de N=X].



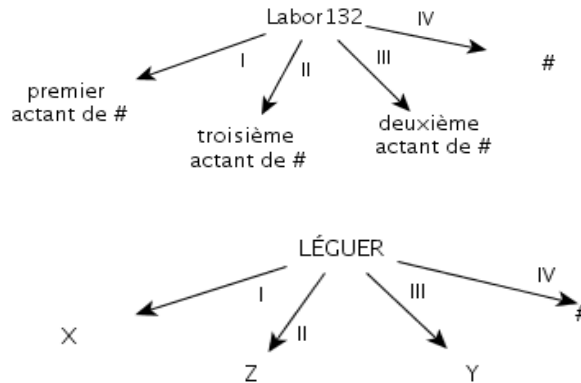
Classe 4 : Structures actantielles à quatre actants

La quatrième classe rassemble les structures actantielles comprenant quatre actants. Parmi l'ensemble des relations lexicales du DiCo, nous n'avons relevé que deux patrons de valence syntaxique profonde à quatre actants, encodées par les FL Labreal_{123} et Labor_{132} :

$\text{Labreal}_{123}(\textit{enveloppe}\#\text{II}) = \textit{acheminer}, \textit{expédier}$ [N=Y à N=Z sous \sim];

$\text{Labor}_{132}(\textit{testament}\#\text{I}) = \textit{léguer}$ [N=Z à N=Y par \sim].

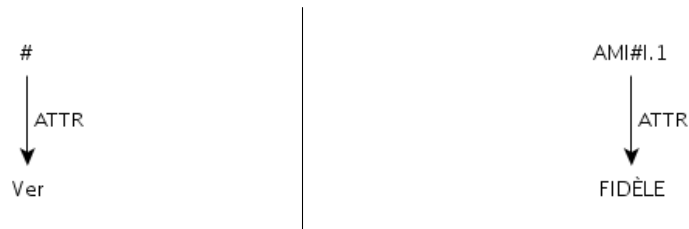
Ces deux FL pointent respectivement vers les patrons [1,2,3,#] et [1,3,2,#]. Le patron [1,3,2,#] peut être représenté par le schéma ci-dessous.



Nous n'avons pas relevé, parmi les relations du DiCo, de relations actantielles à plus de quatre actants. Nous passons, par conséquent, à la présentation des relations attributives.

Classe 5 : Relations attributives simples

Le premier type de relation attributive correspond à une modification du mot-clé. Elle est représentée par le patron [#^]. Rappelons que le symbole « ^ » est toujours situé à droite du symbole représentant l'élément qui subit la modification, soit, ici, le mot-clé. Ce patron s'applique aux FL telles que *Ver*, *Bon*, *Magn*, etc., par exemple : $Ver(ami\#I.1) = fidèle$, que l'on peut représenter par le schéma suivant :



Classe 6 : Relations attributives à un actant

Les FL paradigmatiques ne sont logiquement pas concernées par l'organisation combinatoire des FL, puisqu'elles ne retournent pas de valeurs formant un syntagme

avec le mot-clé. Cependant, comme nous l'avons mentionné à plusieurs reprises dans la thèse (aux sections 2.2.1 et 3.3.4), il arrive que les liens lexicaux encodés par certaines FL paradigmatiques soient syntagmatiques. Par exemple, la FL $Able_1$ représente une relation paradigmatique de dérivation adjectivale « potentielle » du mot-clé. Or, dans l'exemple ci-dessous,

$$Able_1(\textit{rancune}) = \textit{rancunier} ; \textit{porté} [\textit{à ART} \sim],$$

on s'aperçoit que cette FL appliquée à la lexie RANCUNE a comme valeurs, d'une part, un dérivé adjectival en relation paradigmatique : *rancunier*, d'autre part, un adjectif prenant RANCUNE en tant qu'actant : *porté à la rancune*.

On recense de nombreux cas similaires. Même si ces cas ne sont pas systématiques, on observe que certaines FL paradigmatiques ont tendance à retourner des valeurs en lien syntagmatique avec le mot-clé. Il s'agit notamment des dérivés actantiels adjectivaux (A_1 , A_2 , etc.), adverbiaux (Adv_1 , Adv_2 , etc.) ainsi que des dérivés adjectivaux potentiels ($Able_1$, $Able_2$, etc.). On note également que les FL *Mult* et *Sing* ont régulièrement des valeurs prenant le mot-clé en tant qu'actant, par exemple :

$$Mult(\textit{chien}\#I) = \textit{horde}, \textit{meute} [\textit{de} \sim s];$$

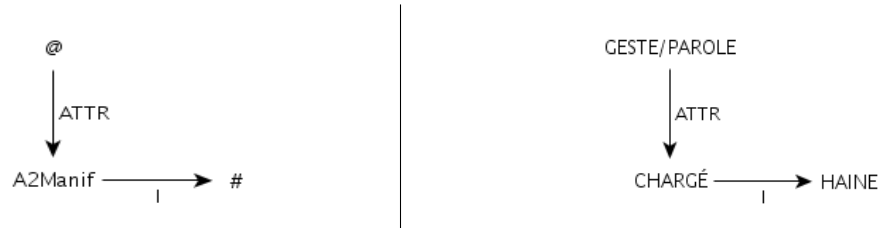
$$Sing(\ulcorner \textit{bon sens} \urcorner) = \textit{dose} [\textit{de} \sim].$$

La dernière classe de patrons de dépendance syntaxique profonde concerne des liens lexicaux syntagmatiques, dont certains sont encodés par des FL paradigmatiques. Les valeurs de ces liens particuliers modifient un actant du mot-clé et prennent le mot-clé comme actant. Par exemple, dans l'exemple ci-dessous, la valeur de la FL modifie un participant extérieur à la lexie HAINE et prend HAINE pour actant :

$$A_2\textit{Manif}(\textit{haine}) = \textit{chargé} [\textit{de (ART)} \sim] : \rightarrow [\textit{Geste ou parole de X}] \textit{qui} \\ \textit{est une manifestation de} \sim$$

Nous avons mentionné, au chapitre 3 (section 3.3.4, p. 107), l'opération de fusion, mise en valeur dans les encodages algébrique et explicite, qui associe à chaque FL syntagmatique f , une relation paradigmatique $//f$. Dans le système de FL, nous associons à certaines FL syntagmatiques, une relation paradigmatique fusionnée potentielle et à certaines FL paradigmatiques, un patron de dépendance syntaxique potentiel. Ainsi, $Able_1$, habituellement catégorisée comme une FL paradigmatique

est potentiellement associée au patron $[1^{\wedge}, \#]$; **Mult** et **Sing** se verront attribuer le patron potentiel $[\#]$, etc. Bien que la FL **A2Manif** soit considérée comme une FL paradigmatique, on lui attribue le patron potentiel $[@^{\wedge}, \#]$ illustré comme suit :



5.3.4 Bilan et illustration

La liste des différents patrons est représentée dans la figure 5.9, chacun d'entre eux y est illustré à l'aide d'un exemple. Cette répartition des FL offre une perspective nouvelle, qui pourra être exploitée, à la fois pour l'aide à l'encodage des FL, et pour la réalisation d'un environnement de consultation d'une base de données ou d'aide à la rédaction. Nous en proposons une illustration dans la figure 5.10 à l'aide de la lexie **RANCUNE** et de ses relations lexicales.

Figure 5.9 – Liste des valences syntaxiques associées aux valeurs de FL

- [#] Degr₁(lait#b) = tourner
-
- **Deux actants**
- [#1] Func₁(catastrophe#I) = s'abattre [sur N=X]
- [#2] Fact₂(charrette) = transporter [N=Y]
- [#3] Func₃(conversation#a) = porter [sur N=Z]
- [#4] Fact₄(cadeau#I) = souligner [N=W]
- [#@] Invol_v(forêt#I) = s'étendre [Loc-in N], couvrir [N]
- [1,#] Real₁(bijou#I) = avoir, porter [ART ~]
- [2,#] Oper₂(faute#I.2) = avoir, contenir [ART ~]
- [3,#] Real₃(charrette) = tirer [ART ~]
- [4,#] Oper₄(commérage#a) = provoquer, susciter [ART ~s]
- [@,#] LiquFact₀(ampoule#I.1) = éteindre [ART ~]
-
- **Trois actants**
- [#1,2] Func_{1,2}(lutte#I.1) = opposer [N=X à N=Y]
- [#2,1] Fact_{2,1}(piste#II) = mener [N=Y à N=X]
- [#2,3] Fact_{2,3}(facteur(1)) = apporter, distribuer, livrer [N=Y à N=Z]
- [#2,4] Fact_{2,4}(tuyau#II) = permettre [à N=Y de V-inf=W]
- [#3,2] Fact_{3,2}(clinique#1) = offrir [N=Z à N=Y]
- [1,#,2] Oper_{1,2}(applaudissements) = adresser [A-poss=X ~ à N=Y]
- [1,#,3] Real_{1,3}(balai) = passer [le ~ Loc-in N=Z]
- [1,#,4] Real_{1,4}(plaque#II.1) = apposer, poser [ART ~ sur N=W]
- [1,2,#] Labor_{1,2}(affection) = avoir [N=Y en ~]
- [1,3,#] Labreal_{1,3}(peine#II) = sanctionner [N=Z de ART ~]
- [2,#,1] Oper_{2,1}(défrite) = faire subir, infliger [ART ~ à N=X]
- [2,#,3] Oper_{2,3}(enveloppe#III) = allouer, verser [ART ~ à N=Z]
- [2,1,#] AntiLabreal_{2,1}(amour-propre) = blesser [N=X dans A-poss ~]
- [3,#,1] Oper_{3,1}(rancune) = inspirer [ART ~ à N=X]
- [3,#,2] Real_{3,2}(peine#II) = valoir [ART ~ à N=Y]
- [3,#,4] Oper_{3,4}(témoignage#II.3) = accorder, donner [ART ~ à N=W]
- [3,1,#] Magn.IncepLabor_{3,1}(admiration) = frapper, saisir [N=X d'~]
- [@,#,1] CausFunc₁(espérance#I) = donner [ART ~ à N=X]
- [@,#,2] CausFunc₂(soupon#I) = diriger [ART ~ sur N=Y]
- [@,#,3] CausFunc₃(indice(1)) = donner, fournir [ART ~ à N=Z]
- [@,1,#] Magn^{quant}.CausReal₁(bijou#I) = couvrir [N=X de ~x]
- [@,2,#] CausReal₂(fardeau#II) = faire peser [ART ~ sur N=Y]
-
- **Quatre actants**
- [1,2,3,#] Labreal_{1,2,3}(enveloppe#II) = expédier [N=Y à N=Z sous ~]
- [1,3,2,#] Labor_{1,3,2}(testament#I) = léguer [N=Z à N=Y par ~]
-
- **Modification simple**
- [#^] Magn(ami#I.1) = fidèle
-
- **Modification à un actant**
- [#^,1] A₂Oper₁(forêt#I) = peuplée [de N=X]
- [#^,2] A₁Fact₂(camion) = chargé [de N=Y]
- [#^,3] A₁Func₃(peine#II) = applicable [^ dans le cas de ^/pour N=Z]
- [1^,#] Adv_v(appréhension) = avec [~]
- [2^,#] A₂(ordonnance#II) = muni [de ART ~]
- [3^,#] A₃(^ levée de boucliers^) = confronté, exposé [à ART ~]
- [@^,#] A₂Manif(défiance) = de [~]

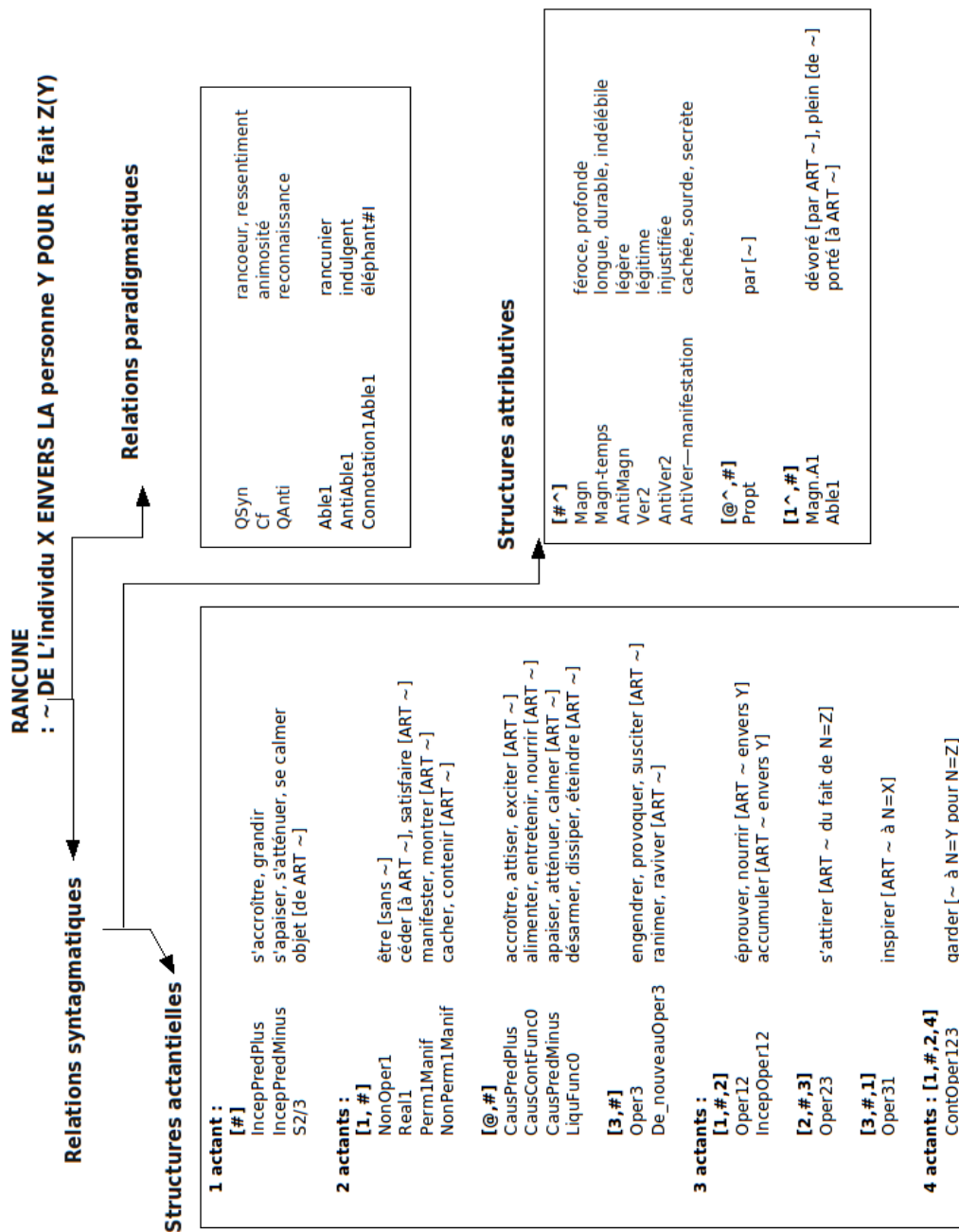


Figure 5.10 – Collocatifs de RANCUNE organisés par patrons de dépendance syntaxique

5.4 Perspective sur le système de fonctions lexicales orientée sur les parties du discours

5.4.1 Objectif

Nous mettons en avant une troisième perspective pour l'organisation du système de FL, centrée sur les parties du discours. Celle-ci offre la possibilité d'organiser les FL en fonction des parties du discours associées à leurs mots-clés et valeurs, et permettra ainsi d'effectuer des requêtes orientées sur les parties du discours des lexies en jeu dans une relation. Il est possible, par exemple, à partir de la lexie TÊTU, d'accéder à l'ensemble de ses dérivés sémantiques et collocatifs organisés de la façon suivante :

– **Adjectifs :**

QSyn *cabochard*_{Adj}, *entêté*_{Adj}, *obstiné*, *intraitable*, *intransigent*, *inflexible*,
borné

QAnti *accommodant*, *arrangeant*, *conciliant*, *flexible*, *souple*, *raisonnable*

– **Verbes :**

V₀ « *avoir la tête dure* »

Caus₁Manif *s'entêter*

– **Noms :**

S₀ *entêtement*

S₁ *têtu*_N, *cabochard*_N, *entêté*_N, « *tête de mule* », « *tête de cochon* »

– **Adverbes :**

Magn « *comme une bourrique* », « *comme une mule* »

On note, pour certaines unités lexicales, la présence d'interjections parmi les dérivés sémantiques. Par exemple, pour FUSIL(1)#a :

– **Interjections :**

$\text{ImperReal}_1(\text{fusil}(1)\#a) = \ll \text{Feu!} \gg, \ll \text{Tirez!} \gg, \ll \lceil \text{En joue!} \rceil \gg.$

Dans la suite de cette section, nous exposons la façon particulière dont la notion de partie du discours est traitée dans la théorie Sens-Texte, et sur laquelle repose notre travail de classification. La TST distingue, en effet, les parties du discours profondes des parties du discours de surface, qui sont respectivement impliquées dans les représentations syntaxiques profondes et de surface (cf. section 5.7 p. 193). Nous traitons également la question des parties du discours associées aux locutions. Une fois ces concepts définis, nous passerons à l'exposé des différentes classes élaborées pour le système de FL.

5.4.2 Les parties du discours dans l'approche Sens-Texte

5.4.2.1 Parties du discours profondes et parties du discours de surface

Les parties du discours dans la Théorie Sens-Texte sont considérées d'une façon particulière. On observe, pour certaines unités lexicales, un manque de correspondance entre leur fonctionnement syntaxique profond et leur fonctionnement syntaxique de surface. Par exemple, la locution $\lceil \text{BEC ET ONGLE} \rceil$ se comporte en syntaxe profonde comme un adverbe modifiant un verbe alors qu'en surface, elle prend la forme d'un syntagme nominal. Mel'čuk [2006 : 33-34] — et, avant lui, Zholkovsky [1971 : 197-209] — propose alors de séparer deux ensembles de parties du discours :

Dans l'approche Sens-Texte, conformément à la distinction entre les structures syntaxiques profondes et les structures syntaxiques de surface, on distingue deux familles de parties du discours : les parties du discours PROFONDES et les parties du discours DE SURFACE (cette distinction a été introduite et discutée dans Zholkovsky [1971 : 197-209]). Les premières sont assignées aux lexies qu'on trouve dans la structure syntaxique profonde, les secondes, aux lexies qu'on trouve dans la structure syntaxique de surface (et dans la Structure Morphologique Profonde).

◇ Les parties du discours profondes

Cinq parties du discours profondes ont été déterminées en se fondant sur les rôles

syntaxiques passifs et actifs profonds pouvant être remplis au niveau syntaxique profond, soit le verbe, le nom, l'adjectif, l'adverbe et le clausatif [Mel'čuk, 2006 : 41-43]. Donnons quelques précisions sur cette dernière partie du discours : un clausatif profond est typiquement le sommet d'une proposition, il est syntaxiquement équivalent à une proposition, qui peut être complète comme *Oui, Non, Pas du tout*, ou incomplète comme *N'empêche que ...*. Les clausatifs peuvent être également des phrases entières comme des proverbes ou des maximes.

◇ Les parties du discours de surface

Les parties du discours de surface regroupent que les unités lexicales monolexémiques. Elles sont au nombre de huit : verbe, nom, adjectif, adverbe, clausatif, préposition, conjonction et numéral (cardinal). Ce sont des classes disjointes formées à partir de trois critères : la valence syntaxique passive des lexèmes (c'est-à-dire les rôles de dépendants qu'ils peuvent remplir), la valence syntaxique active des lexèmes (c'est-à-dire les types de dépendants qu'ils peuvent avoir eux-mêmes) ainsi que les catégories flexionnelles syntaxiques des lexèmes.

Parmi les huit parties du discours de surface, les cinq premières correspondent aux parties du discours profondes du même nom. Quant aux trois autres, elles sont considérées comme des types particuliers d'adverbes et d'adjectifs profonds, mais dont le rôle et le comportement est si particulier en surface, qu'elles méritent une catégorie propre. Selon Mel'čuk [2006 : 45-47], une préposition est un adverbe qui régit, en tant que complément obligatoire, un nom ou une forme nominale du verbe, c'est-à-dire un infinitif. Une conjonction est *grosso modo* un adverbe régissant, en tant que complément obligatoire, soit un verbe fini, c'est-à-dire la tête d'une proposition (c'est alors une conjonction subordinative), soit une lexie de la même partie du discours que le gouverneur de la conjonction (c'est alors une conjonction coordinative). On note que les prépositions et les conjonctions ont une distribution assez similaire à celle des adverbes : elles se rattachent à leur gouverneur syntaxique à peu près par les mêmes règles que les adverbes. Au niveau syntaxique profond, on les encode comme des adverbes. Cependant, les prépositions ainsi que les conjonc-

tions ont tellement de propriétés syntaxiques particulières et jouent, en syntaxe de surface, un rôle si spécial et important qu'on doit considérer qu'elles représentent des parties du discours de surface à part entière.

Un numéral est un nom de nombre qui s'utilise de façon assez spéciale pour quantifier les noms et pour construire les noms de nombres complexes. Bien que considéré comme relevant d'un adjectif profond, il ne partage pas, en français, suffisamment de propriétés avec l'adjectif pour être considéré comme une sous-classe de l'adjectif de surface.

Suite à ce qui vient d'être exposé, on peut établir une correspondance entre les parties du discours profondes et les parties du discours de surface [Mel'čuk, 2006 : 47].

Partie du discours profonde	N	V	Adj	Adv	Claus
Partie du discours de surface	N	V	Adj Num	Adv Prép Conj	Claus

Figure 5.11 – Correspondance entre les parties du discours profondes et les parties du discours de surface

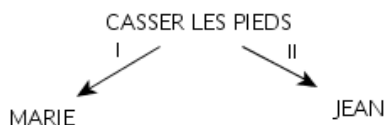
Nous devons signaler que toutes les parties du discours ne semblent pas universelles. Par exemple, dans certaines langues asiatiques, comme le chinois, le vietnamien et le coréen, il n'existe pas à proprement parler de classe d'adjectifs. Ces derniers sont considérés comme des verbes qualificatifs, qui représentent une sous-classe de verbes (cf. Nguyen [2006]).

5.4.2.2 Parties du discours et locutions

La distinction exposée plus haut est, en fait, issue de la problématique d'assigner une partie du discours à une locution. En effet, dans une représentation syntaxique profonde, une locution n'occupe qu'un seul nœud, mais au niveau syntaxique de surface, elle va correspondre à un sous-arbre comme on le voit dans la figure 5.12 représentant la phrase : *Marie casse les pieds à Jean* [Mel'čuk, 2006 : 31].

Ces observations ont été faites par Igor Mel'čuk, Alain Polguère et Sara-Anne Leblanc dans le cadre de l'élaboration du LAF (et du travail de doctorat de Sara-Anne

Structure syntaxique profonde



Structure syntaxique de surface

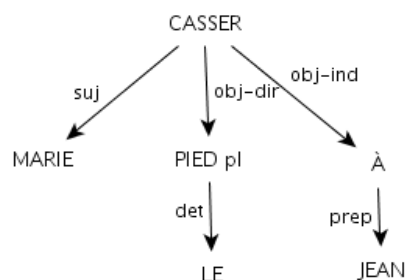


Figure 5.12 – Structures syntaxiques profonde et de surface de *Marie casse les pieds à Jean*

Leblanc [en cours]). Pour traiter le problème, ils distinguent deux types de locutions : les locutions syntaxiquement endocentriques et les locutions syntaxiquement exocentriques.

1. **Les locutions endocentriques** ont un comportement syntaxique déterminé par leur tête. Par exemple, «BAIN DE FOULE», «CERCLE VICIEUX» et «SAC À DOS» se comportent syntaxiquement comme des noms alors que les locutions «SE BOUFFER LE NEZ» et «METTRE LES BOUTS» s'utilisent comme des verbes. Les locutions sont distribuées entre des classes syntaxico-lexicales qui sont parallèles aux parties du discours de surface des lexèmes, soit les locutions nominales, verbales, adjectivales, adverbiales, clausales, prépositionnelles, conjonctionnelles et numérales.
2. **Les locutions exocentriques** ont un comportement syntaxique qui ne peut être déterminé par leur structure interne. Par exemple, «BEC ET ONGLE» est une locution nominale, c'est-à-dire syntaxiquement gouvernée par un nom, mais s'utilise comme un adverbe (ex : *se défendre bec et ongle*), de même «EN MIETTES» est une locution prépositionnelle occupant la position syntaxique d'un adjectif (ex : *une voiture en miettes*). Pour résoudre cette question, Mel'čuk propose de fournir la caractérisation syntaxique de surface ainsi que sa caractéristique syntaxique profonde sous la forme « emploi ...-al » comme l'illustrent les exemples suivants [Mel'čuk, 2006 : 31] :

「À PART ENTIÈRE」 : locution prépositionnelle, emploi adjectival ;

「COMME DU POISSON POURRI」 : locution conjonctionnelle, emploi adverbial ;

「DANS LE SENS DES AIGUILLES D'UNE MONTRE」 : locution prépositionnelle, emploi adverbial.

Ces précisions sur la façon dont les parties du discours sont vues dans la théorie Sens-Texte permettent à présent d'établir la classification des FL visée dans cette section.

5.4.3 Les différents couples de parties du discours associés aux liens lexicaux représentés par les fonctions lexicales

Une FL est considérée, dans les représentations syntaxiques profondes, comme une lexie. On peut donc lui attribuer une partie du discours profonde. Cependant, si certaines FL correspondent à une et une seule partie du discours profonde : par exemple, $Oper_1$ et $Real_1$ sont des verbes, S_{loc} et S_1 des noms, A_1 et A_0 des adjectifs, Adv_1 et Loc_{in} des adverbes et $Imper$ un clausatif profond, d'autres peuvent être liées à deux parties du discours : $Magn$ et Ver sont des adjectifs ou des adverbes profonds en fonction du mot-clé auquel elles sont appliquées :

mot-clé nominal $\rightarrow Magn(aboiement) = fort, furieux, féroce ;$

mot-clé adjectival $\rightarrow Magn(armé) = bien, fortement, 「jusqu'au dents」.$

D'autres encore ont un comportement « caméléon » dans le sens où elles épousent la partie du discours profonde du mot-clé auquel elles sont appliquées : c'est le cas de Syn et $Anti$, par exemple.

Dans le cas des locutions endocentriques, la partie du discours profonde associée à la valeur d'une FL correspondra à la partie du discours profonde de l'élément tête, et, dans le cas des locutions exocentriques, la partie du discours profonde associée à la valeur d'une FL correspondra au type d'emploi (nominal, adverbial, etc.) de la locution.

Aussi, sachant qu'il existe cinq parties du discours profondes, on dispose potentiellement de 5^2 , c'est-à-dire vingt-cinq couples de parties du discours profondes associés aux liens lexicaux représentés par les FL. Ces vingt-cinq couples forment des classes vers lesquelles pointent les FL. Comme nous l'avons mentionné plus haut, de nombreuses FL pointent vers plus d'une de ces classes.

Nous avons organisé les couples de parties du discours en fonction de la partie du discours des valeurs retournées par les FL. Ce choix est motivé par l'un des usages qui sera fait du système de FL, pour lequel on souhaite organiser les dérivés sémantiques et collocatifs d'une unité lexicale en fonction de leurs parties du discours.

Valeur : N	Valeur : V	Valeur : A	Valeur : Adv	Valeur : Claus
N → N	N → V	N → A	N → Adv	N → Claus
A → N	A → V	A → A	A → Adv	A → Claus
V → N	V → V	V → A	V → Adv	V → Claus
Adv → N	Adv → V	Adv → A	Adv → Adv	Adv → Claus
Claus → N	Claus → V	Claus → A	Claus → Adv	Claus → Claus

Tableau 5.I – Classes des couples de parties du discours associés aux liens lexicaux encodés par les FL

Nous illustrons dans la figure 5.13 l'organisation des dérivés sémantiques et collocatifs de RANCUNE selon leur parties du discours, telle qu'elle pourrait figurer dans une base de données lexicale.

RANCUNE

- ▶ **NOMS**
- QSyn ranceur, ressentiment
- Cf amiosité
- QAnti reconnaissance
- S_{2/3} objet [de ~]
- Connotation₁Able₁ éléphant#I

- ▶ **VERBES**
- IncepPredPlus s'accroître, grandir
- CausPredPlus accroître, attiser, exciter, renforcer [ART ~]
- IncepPredMinus s'apaiser, s'atténuer, se calmer
- CausPredMinus apaiser, atténuer, calmer [ART ~]
- Oper_{1,2} avoir, éprouver, nourrir, ressentir [ART ~ Prep-envers N=Y]
- IncepOper_{1,2} accumuler [ART ~ Prep-envers N=Y]
- ContOper_{1,2,3} garder [~ à N=Y pour N=Z]
- NonOper₁ être [sans ~]
- Oper_{2,3} s'attirer [ART ~ avec/↑ du fait de ↓ / par N=Z < en V-pprés=Z>]
- Oper₃ engendrer, provoquer, susciter [ART ~]
- Oper_{3,1} inspirer [ART ~ à N=X]
- De_nouveauOper₃ ranimer, raviver [ART ~]

- ▶ **ADJECTIFS**
- Magn.A₁ plein [de ~], dévoré [par ART ~]
- Able₁ porté [à la ~], rancunier
- AntiAble₁ indulgent
- Magn profonde, féroce
- AntiMagn légère
- Magn_{temp} longue, durable, tenace, vivace, indéfectible, vieille
- Ver₂ légitime
- AntiVer₂ injustifiée
- AntiVer₁manifestation sourde, cachée, secrète

- ▶ **ADVERBES**
- Propt par [~]

Figure 5.13 – Relations lexicales de RANCUNE organisées par parties du discours profondes des valeurs de FL

5.5 Perspective communicationnelle sur le système de fonctions lexicales

5.5.1 Objectif

Le mode d'organisation que nous présentons dans cette section vise à déterminer la cible communicationnelle d'une FL. Nous appelons *cible communicationnelle* d'une FL, l'élément (soit, le mot-clé de la FL, l'un de ses actants sémantiques, un participant extérieur à sa structure actantielle, ou encore, la valeur de la FL), qui est mis en avant au niveau communicationnel dans la relation encodée par la FL. Illustrons notre propos à l'aide de la lexie RANCUNE, dont la structure actantielle est la suivante :

~ DE L'individu X ENVERS LA personne Y POUR LE fait Z(Y).

Nous postulons que, parmi les dérivations sémantiques et les collocations associées à cette lexie, certaines prennent pour cible communicationnelle le mot-clé, par exemple :

Magn(*rancune*) = *féroce, profonde* ;

IncepPredPlus(*rancune*) = *s'accroître, grandir*.

D'autres collocations ou dérivations sémantiques visent un actant sémantique particulier du mot-clé, soit son premier actant :

Able₁(*rancune*) = *porté [à la ~], rancunier* ;

Real₁(*rancune*) = *céder [à la ~]* ;

son deuxième actant :

Oper₂₃(*rancune*) = *s'attirer [ART ~ avec/Γ du fait de ∇/par N=Z <en V-pprés=Z>]* ;

son troisième actant :

Oper₃(*rancune*) = *engendrer, provoquer, susciter [ART ~]* ;

ou un participant extérieur ne correspondant pas à un de ses actants :

LiquFunc₀(*rancune*) = *désarmer, dissiper, éteindre [ART ~]*.

D'autres encore visent la valeur de la FL, par exemple :

$$\text{Propt}(\text{rancune}) = \text{par } [\sim].$$

Cette répartition des dérivés sémantiques et des collocatifs d'une unité lexicale a un intérêt pour guider un processus de rédaction, où un utilisateur cherche, par exemple, à encoder un message mettant en valeur un élément particulier plutôt qu'un autre. Si, pour exprimer un énoncé signifiant 'X n'éprouve plus de rancune pour Y', on cherche à mettre l'accent sur un participant extérieur, qui est responsable de ce fait, on choisira les collocations : *[Qqn.] désarmer/dissiper/éteindre la rancune*. En revanche, si l'on souhaite ne pas évoquer la cause de ce fait et mettre en avant le nom RANCUNE, on choisira les collocations : *la rancune s'éteint/disparaît/s'évanouit*.

On peut également, dans une optique de préparation de texte, rassembler tous les dérivés sémantiques identifiant ou qualifiant un actant sémantique particulier d'un mot-clé et les collocatifs désignant les actions dont il est l'auteur. Par exemple, pour la lexie BAGARRE#a, on récolterait les relations suivantes, dans lesquelles le premier actant de BAGARRE#a est la cible communicationnelle :

$$\begin{aligned} \text{Able}_1(\text{bagarre}\#a) &= \text{bagarreur}_{\text{Adj}}; \\ \text{Qual}_1(\text{bagarre}\#a) &= \text{agressif}; \\ \text{Oper}_1(\text{bagarre}\#a) &= \text{participer} [\text{à ART } \sim]; \\ \text{sans le vouloir } \text{Oper}_1(\text{bagarre}\#a) &= \text{être pris} [\text{dans ART } \sim]; \\ \text{Caus}_1\text{Oper}_1(\text{bagarre}\#a) &= \text{se lancer} [\text{dans ART } \sim]; \\ \text{essayer de } \text{Caus}_1\text{Func}_2(\text{bagarre}\#a) &= \text{chercher} [\text{la } \sim \text{ avec N=Y}]. \end{aligned}$$

Cette perspective sur le système est donc particulièrement pertinente pour la consultation d'une BLEC en tant qu'outil d'aide à la rédaction.

Dans la suite de cette section (section 5.5.2), nous devons définir la notion de cible communicationnelle d'une FL et la façon dont elle est déterminée. Nous montrons, par la suite, comment la perspective communicationnelle sur le système de FL est formalisée à travers plusieurs classes (section 5.5.3).

5.5.2 Comment déterminer la cible communicationnelle d'une FL ?

Pour déterminer quelles sont les cibles communicationnelles visées par les FL, nous distinguons trois cas de figure. Premièrement, nous distinguons les FL dont les valeurs sont des collocatifs non verbaux, comme, par exemple, les FL **Magn** et **S₀Oper₁** dans les liens lexicaux suivants :

$\text{Magn}(\text{gai}\#\text{a}) = \textit{comme un pinson}$;

$\text{S}_0\text{Oper}_1(\textit{enveloppe}\#\text{III}) = \textit{allocation, versement}$ [de ART \sim].

Le deuxième cas de figure représente les FL dont les valeurs sont des collocatifs verbaux, qui, avec le mot-clé (et éventuellement un ou plusieurs de ses actants), permettent la formation d'un énoncé. Il s'agit, par exemple, de la FL **Oper₁**, illustrée ci-dessous :

$\text{Oper}_1(\textit{affection}) = \textit{éprouver, ressentir}$ [ART \sim pour N=Y],

où la valeur de la FL, le mot-clé et ses actants, permettent de construire un énoncé du type : *Pierre éprouve de l'affection pour Marianne.*

Enfin, le troisième cas de figure concerne les FL paradigmatiques, qui associent deux unités lexicales sur le plan paradigmatique, telles que **S₁**, **S_{loc}** et **Syn**, illustrées dans les exemples ci-dessous :

$\text{S}_1(\textit{boire}) = \textit{buveur}$;

$\text{S}_{\text{loc}}(\textit{abeille}) = \textit{ruche, rucher}$;

$\text{Syn}(\textit{bicyclette}) = \textit{vélo}$.

5.5.2.1 FL dont les valeurs sont des collocatifs non verbaux

Pour déterminer la cible communicationnelle d'un syntagme collocationnel non verbal, nous introduisons la notion de dépendance communicative. Celle-ci est définie dans [Mel'čuk, à paraître][chapitre 6] comme suit :

In a configuration of semantemes [...] $(\sigma_1) - (\sigma_2)$, the semantic node (σ_2) is said to depend communicatively on the semantic node (σ_1) in a direct way, if this configuration can be reduced to (σ_1) (rather than to

(σ_2) such that the meaning conveyed is simply reduced (\approx impoverished) but not distorted.

En d'autres termes, on considère que le nœud communicativement dominant d'une représentation sémantique est sa paraphrase minimale. On note que, dans la représentation sémantique, le nœud communicativement dominant est souligné.

Deux types de syntagmes doivent être distingués : les syntagmes collocatifs qui modifient le mot-clé et les syntagmes collocatifs qui le prennent pour actant. Illustrons le premier type de syntagme avec les relations en (112) et (113) ci-dessous :

(112) $\text{Magn}(\textit{acclamation}) = \textit{chaleureuses, enthousiastes}$;

(113) $\text{AntiVer}(\textit{applaudir}\#\text{I}) = \textit{mollement, } \ulcorner \textit{du bout des doigts} \urcorner$.

On dégage respectivement de ces relations les représentations sémantiques suivantes :

'acclamation' $\leftarrow 1$ — 'intense' ;

'applaudir#I' $\leftarrow 1$ — 'non sincèrement'.

Les nœuds 'acclamation' et 'applaudir' sont communicativement dominants dans les représentations sémantiques des relations (112) et (113). Les syntagmes : *acclamation chaleureuse* et *applaudir mollement* peuvent être, en effet, minimalement paraphrasés et réduits à *acclamation* et *applaudir*. Des syntagmes dont 'intense' et 'non sincèrement' sont les sens communicativement dominants sont des syntagmes tels que : *la chaleur de l'acclamation* et *la mollesse des applaudissement*.

On remarque, pour ce premier type de syntagme, que le nœud communicativement dominant correspond au gouverneur syntaxique du syntagme.

Le deuxième type de syntagme concerné sont des syntagmes formés d'une valeur de FL prenant le mot-clé pour actant, par exemple :

(114) $\text{S}_0\text{Oper}_1(\textit{enveloppe}\#\text{III}) = \textit{allocation, versement}$ [de ART \sim];

(115) $\text{Loc}_{\text{in}}(\textit{bateau}) = \ulcorner \textit{à bord de} \urcorner, \textit{dans}$ [ART \sim];

(116) $\text{Instr}(\textit{couteau}) = \textit{avec}$ [ART \sim].

On dégage de la relation (114) la représentation sémantique suivante :

'enveloppe#III' $\leftarrow 2$ — 'allocation'. Les syntagmes formés par les valeurs et le mot-clé

de la FL sont : *allocation d'une enveloppe* et *versement d'une enveloppe*. Le test de la paraphrase minimale permet de réduire les syntagmes *allocation d'une enveloppe* et *versement d'une enveloppe* aux noms *allocation* et *versement*. Ces syntagmes s'opposent aux syntagmes *enveloppe allouée* et *enveloppe versée*, qui peuvent être minimalement paraphrasés par *enveloppe* et dont le nœud correspondant dans la représentation sémantique est communicativement dominant. Dans le cadre de la relation (114), on en déduit donc que 'allocation' et 'versement' constituent les nœuds dominants de la représentation sémantique. Ici encore, on remarque que le nœud communicativement dominant correspond au gouverneur syntaxique du syntagme.

On peut dégager des relations (115) et (116), les représentations sémantiques suivantes :

'situé dans'—2→'bateau'

'à l'aide de'—2→'couteau'

Le test de paraphrase minimale est plus difficilement applicable à ce type de syntagme, car, tout comme dans les exemples précédents, l'emploi de $\text{Loc}_{\text{in}}(\text{L})$ ou de $\text{Instr}(\text{L})$ nécessite la présence de L. Les paraphrases minimales des syntagmes *à bord du bateau* et *avec un couteau* : *à bord de* et *avec* ne sont pas complètes. Toutefois, dans ces syntagmes, la dominance syntaxique de $\lceil \text{à bord de} \rceil$ et *avec* reflète la dominance communicative des nœuds 'situé dans' et 'à l'aide de' dans les deux représentations sémantiques correspondantes. Dans les relations (114), (115) et (116), la cible communicationnelle de la FL est donc sa valeur.

Nous considérons que la cible communicationnelle des FL dont les valeurs sont des collocatifs non verbaux formant avec le mot-clé un syntagme est le gouverneur syntaxique du syntagme, puisqu'il correspond au nœud communicativement dominant de la configuration sémantique correspondante.

5.5.2.2 FL dont les valeurs sont des collocatifs verbaux

Pour les FL dont les valeurs sont des collocatifs verbaux, on remarque que la valeur de la FL, son mot-clé et éventuellement les actants du mot-clé permettent

de construire un énoncé. Par exemple, la FL Oper_{12} retourne une valeur verbale, dont le sujet grammatical est le premier actant syntaxique profond du mot-clé, le premier complément du verbe, le mot-clé lui-même, et le second complément du verbe, le deuxième actant syntaxique profond du mot-clé. Dans la relation ci-dessous :

(117) $\text{Oper}_{12}(\ulcorner \text{coup de fil} \urcorner) = \text{donner, passer} [\text{ART} \sim \text{à N=Y}]$,

le mot-clé de la FL et ses deux actants peuvent former un énoncé du type : *Pierre donne un coup de fil à Paul*. Dans un énoncé, on distingue généralement le rhème, ce qui est déclaré dans l'énoncé, et le thème, ce à propos de quoi on le déclare. L'élément communicativement dominant d'un énoncé correspond à son thème. Or, la notion de thème dans le cadre d'un énoncé ou d'une proposition, correspond à la cible communicationnelle que nous cherchons à déterminer pour les FL de ce type. Ceci revient donc à déterminer quel est, pour chaque FL permettant la formation d'un énoncé, le thème de cet énoncé. Mel'čuk [2001 : 127] considère deux moyens dans les langues pour l'expression du thème :

Languages manifest two major strategies to implement the [Semantic-Theme] syntactically : the [Semantic-Theme] is implemented either as the first Deep-Syntactic Actant [= DSyntA I] of the main verb of the clause (which gives the Surface Syntactic Subject) or as a Deep-Syntactic 'append' element - a Surface Syntactic Prolepsis [...]. In many languages, the [Semantic-Theme] tends to be implemented by the DSyntA I of the main verb, i.e. the Surface-Syntactic Subject.

Nous considérons donc que la cible communicationnelle d'une FL dont les valeurs permettent de former un énoncé avec le mot-clé et éventuellement ses actants, est le sujet grammatical de la valeur, qui représente, par défaut, le thème du message exprimé. Toutefois, le parallèle que nous suggérons entre le thème d'un énoncé et la cible communicationnelle d'une FL ne vaut que pour les cas où un énoncé n'est ni clivé, ni pseudo-clivé, ni focalisé. Les FL ne se situent pas, en effet, à ce niveau d'expression communicationnelle.

À partir des relations (117) plus haut et (118) ci-dessous :

(118) $\text{Oper}_{21}(\ulcorner \text{coup de fil} \urcorner) = \text{recevoir} [\text{ART} \sim \text{de N=X}]$,

il est possible de former les énoncés *Pierre passe un coup de fil à Paul* et *Paul reçoit un coup de fil de Pierre*, dont les thèmes respectifs sont *Pierre* et *Paul*. La cible communicationnelle de Oper_{12} constitue donc le premier actant sémantique de son mot-clé alors que la cible communicationnelle de Oper_{21} constitue le deuxième actant sémantique de son mot-clé.

Selon le procédé qui vient d'être exposé, nous pouvons déterminer que les cibles communicationnelles des FL Func_0 , Func_1 , Fact_0 et Fact_2 , illustrées ci-dessous, correspondent à leur mot-clé, puisqu'il constitue le sujet grammatical des valeurs qu'elles retournent, qui permet l'expression du thème de l'énoncé formé :

$\text{Func}_0(\text{catastrophe}\#\text{I}) = \textit{se produire}$;

$\text{Func}_1(\text{catastrophe}\#\text{I}) = \textit{s'abattre}$ [sur N=X] ;

$\text{Fact}_0(\text{avocat}(1)\#\text{I}) = \textit{exercer}$;

$\text{Fact}_2(\text{avocat}(1)\#\text{I}) = \textit{représenter}$ [N=Y].

5.5.2.3 FL encodant des relations paradigmatiques

Pour ce qui est des FL paradigmatiques, nous distinguons trois cas. Premièrement, les cas où la FL permet l'identification ou la qualification d'un actant sémantique (en correspondance avec l'actant syntaxique profond) du mot-clé. Il s'agit principalement des dérivés nominaux ou adjectivaux actantiels, qui représentent les noms et modificateurs typiques des actants syntaxiques profonds n° *i*, tels que S_1 , S_2 , S_3 , Able_1 , Qual_2 , etc., par exemple :

(119) $\text{Able}_1(\text{bagarre}\#\text{a}) = \textit{bagarreur}_{\text{Adj}}$;

(120) $\text{Qual}_2(\text{adoration}\#\text{a}) = \textit{merveilleux}$;

(121) $\text{S}_3(\text{accusation}\#\text{I.2}) = \textit{motif, } \lceil \textit{chef d'~} \rceil, \textit{crime}\#\text{I.a, délit, méfait}$.

Pour ces types de relations, nous considérons que la cible communicationnelle est l'actant sémantique du mot-clé identifié ou qualifié par la FL. Ainsi, dans la relation (119), la cible communicationnelle est le premier actant sémantique de $\text{BAGARRE}\#\text{a}$, dans la relation (120), la cible communicationnelle est le deuxième actant sémantique de $\text{ADORATION}\#\text{a}$ et dans la relation (121), la cible communi-

cationnelle est le troisième actant sémantique de ACCUSATION#I.2.

Le deuxième cas regroupe les FL paradigmatiques, notamment, les dérivés nominaux circonstanciels ou différentes FLNS, qui identifient un circonstant ou un participant extérieur à la structure actantielle du mot-clé de la FL, comme dans les quatre relations ci-dessous :

(122) Individu qui vérifie \sim des voyageurs dans les trains(*billet*#I.2) = *contrôleur* ;

(123) Individu qui contrôle le déroulement de \sim (*débat*#2) = *animateur* [de ART \sim] ;

(124) S_{instr} (*guérison*#b) = *médicament, remède*#I ;

(125) S_{loc} (*abeille*) = *ruche, rucher*.

Dans les relations (122), (123), (124) et (125), les cibles communicationnelles des FL sont des participants extérieurs aux structures actantielles des lexies BILLET#I.2, DÉBAT#2, GUÉRISON#b et ABEILLE.

Enfin, le troisième type de FL paradigmatique sont les FL représentant les relations telles que la synonymie, l'hyponymie, l'antonymie, la conversion, etc. comme illustrées par les exemples ci-dessous :

Syn (*bicyclette*) = *vélo* ;

$Hypo$ (*chanson*#I.1) = *berceuse* ;

$Anti$ (*ami*#II) = *ennemi*#II ;

$Conv_{21}$ (*craindre*) = *effrayer* [N=X].

Nous considérons que ces relations ont pour cible communicationnelle le mot-clé lui-même. Les dérivés sémantiques ne servent pas, en effet, à identifier un actant du mot-clé ni un participant extérieur, mais fournissent des équivalents, antonymes, conversifs, termes plus génériques, plus spécifiques, etc. visant directement le mot-clé.

5.5.3 Structuration des FL selon leurs cibles communicationnelles

Afin d'organiser les FL selon les différentes cibles communicationnelles qu'elles visent, nous créons huit classes (cf. figure 5.14) correspondant aux cibles potentielles d'une FL. La cible d'une FL peut être son mot-clé : **Cible mot-clé (ou #)**, un des actants sémantiques du mot-clé : **Cible n** (où $n = \{1, 2, 3, 4, 5 \text{ et } 6\}$), un participant extérieur à la structure actantielle du mot-clé : **Cible @**, ou enfin, la valeur de la FL : **Cible V**.

- ▶ **Cible #** : Mot-clé de la FL
- ▶ **Cible @** : Participant extérieur
- ▶ **Cible 1** : Premier actant du mot-clé
- ▶ **Cible 2** : Deuxième actant du mot-clé
- ▶ **Cible 3** : Troisième actant du mot-clé
- ▶ **Cible 4** : Quatrième actant du mot-clé
- ▶ **Cible 5** : Cinquième actant du mot-clé
- ▶ **Cible 6** : Sixième actant du mot-clé
- ▶ **Cible V** : Valeur de la FL

Figure 5.14 – Liste des cibles communicationnelles visées par les FL

Cette répartition des FL est illustrée à travers la lexie RANCUNE dans la figure 5.15.

RANCUNE

- ▶ **CIBLE : MOT-CLÉ (#)**
 - QSyn rancœur, ressentiment
 - Cf amiosité
 - QAnti reconnaissance
 - Magn profonde, féroce
 - AntiMagn légère
 - Magn_{temp} longue, durable, tenace, virulente, indélébile, vieille
 - Ver₂ légitime
 - AntiVer₂ injustifiée
 - AntiVer₁manifestation sourde, cachée, secrète
 - IncepPredPlus s'accroître, grandir
 - IncepPredMinus s'apaiser, s'atténuer, se calmer

- ▶ **CIBLE : PREMIER ACTANT**
 - Magn.A₁ plein [de ~], dévoré [par ART ~]
 - Able₁ porté [à la ~], rancunier
 - AntiAble₁ indulgent
 - Connotation₁Able₁ éléphant#I
 - Oper₁₂ avoir, éprouver, nourrir, ressentir [ART ~ Prep-envers N=Y]
 - IncepOper₁₂ accumuler [ART ~ Prep-envers N=Y]
 - ContOper₁₂₃ garder [~ à N=Y pour N=Z]
 - NonOper₁ être [sans ~]
 - Real₁ satisfaisre [ART ~], céder [à ART ~]
 - Perm₁Manif manifester, montrer [ART ~ Prép-envers N=Y]

- NonPerm₁Manif cacher, contenir, dissimuler, étouffer [ART ~]

- ▶ **CIBLE : DEUXIÈME ACTANT**
 - S_{2/3} objet [de ~]
 - Oper₂₃ s'attirer [ART ~ avec /^r du fait de / par N=Z < en V-pprés=Z>]

- ▶ **CIBLE : TROISIÈME ACTANT**
 - S_{2/3} objet [de ~]
 - Oper₃ engendrer, provoquer, susciter [ART ~]
 - Oper₃₁ inspirer [ART ~ à N=X]
 - De_nouveauOper₃ ranimer, raviver [ART ~]

- ▶ **CIBLE : PARTICIPANT EXTÉRIEUR @**
 - LiquFunc₀ désarmer, dissiper, éteindre [ART ~]
 - CausContFunc₀ alimenter, entretenir, nourrir [ART ~]
 - CausPredPlus accroître, attiser, exciter, renforcer [ART ~]
 - CausPredMinus apaiser, atténuer, calmer [ART ~]

- ▶ **CIBLE : VALEUR DE LA FL**
 - Propt par [~]

Figure 5.15 – Relations lexicales de RANCUNE organisées en fonction des cibles communicationnelles

5.6 Conclusion

À partir du constat que les classifications des FL ne sont que partielles, nous avons postulé une organisation plus riche des FL sous la forme d'un système, dont nous avons détaillé les caractéristiques. Ce système représente l'ensemble des FL selon un rapport différentiel et permet de faire ressortir leurs similitudes et différences, et ce, selon plusieurs points de vue. Les perspectives d'organisation du système correspondent à plusieurs informations que l'on peut dégager à l'observation d'une relation lexicale, soit, son sens, la dimension combinatoire de ses éléments, les parties du discours des lexies en jeu, ainsi que l'élément communicationnellement dominant dans la relation. Pour chacun de ces modules d'organisation, nous avons établi un certain nombre de classes. Elles sont présentées dans les figures 5.5 (p. 188), 5.9 (p. 203) et 5.14 (p. 221) ainsi que dans le tableau 5.I (p. 211). Le système a été implémenté sous la forme d'une table de données MySQL que nous présentons au chapitre suivant (section 6.2).

Ainsi constitué, le système va pouvoir être exploité de plusieurs façons. Premièrement, nous envisageons, dans de futurs travaux, la création d'interfaces de consultation de BLEC épousant les quatre perspectives du système. Deuxièmement, il est possible d'en dériver directement un outil pour guider le processus d'encodage des relations lexicales au moyen de FL. Enfin, il sert de référence pour effectuer des calculs sur les informations lexicographiques pouvant entrer en compte pour la génération automatique d'éléments de fiches lexicographiques. Nous consacrons le chapitre suivant à la présentation de ces deux dernières applications.

Chapitre 6

Applications dérivées du système de fonctions lexicales

6.1 Introduction

Le sixième chapitre de cette thèse est consacré à la présentation d'applications que nous avons dérivées du système de FL. Ces applications entrent dans le cadre d'une nouvelle méthodologie pour l'élaboration du DiCo et, plus largement, de toute BLEC. La méthodologie d'élaboration du DiCo reflète l'évolution de la lexicographie, dans la mesure où il a lui-même incarné différents formats. Créé, dans un premier temps, en tant que version informatisée du DEC [Mel'čuk et al., 1995] (dont il reprend tous les types de données mis à part les définitions), sa structure a la forme d'une base de données conçue avec le logiciel FileMaker. Cependant, son système de requête est peu développé, si bien que le DiCo au format File Maker reste structuré comme un dictionnaire textuel informatisé avec une distinction des différents champs et quelques potentialités de requêtes (cf. figure 6.2 p. 234). Dans une seconde étape, Steinlin et al. [2005] ont procédé à une atomisation des champs du DiCo : un programme de compilation, le DiCompil, transforme la structure linéaire de type FileMaker en une base de données relationnelle MySQL consultable via l'interface DiCouèbe [Jousse et Polguère, 2005]¹. Le DiCo s'affranchit alors de sa structuration par fiches pour devenir une ressource consultable par n'importe quelle donnée lexicographique [Steinlin et al., 2004 : 7]. Aujourd'hui, nous entrons dans la troisième phase de développement du DiCo, au cours de laquelle il subit un

¹La base de données est disponible à l'adresse suivante : <http://olst.ling.umontreal.ca/dicouebe/>.

autre changement important dans sa structuration. Polguère [2006, 2009] souhaite transformer la base de données traditionnelle en un réseau lexical (ou système lexical) et développe actuellement une structuration en graphe.

Nous souhaitons développer des outils pour faciliter la rédaction du DiCo et autres BLEC, fondés sur le calcul d'éléments de fiches lexicographiques et sur les inférences que l'on peut faire en exploitant notre système de FL. L'automatisation devient de plus en plus nécessaire dans le domaine du TAL où plus aucune ressource ou presque n'est élaborée entièrement de façon manuelle. Quelques exemples subsistent comme le Dicovalence [van den Eynde et Mertens, 2006] et le DiCo, mais la majorité est effectuée par extraction automatique à partir de corpus, et ce, aussi bien pour les bases de données morphologiques et syntaxiques comme le Lexique des formes fléchies du français (*Lefff*) [Sagot et al., 2006] que pour les bases de données sémantiques comme le projet *Dixem* (dictionnaires de sèmes) [Valette et al., 2006]. Nous cherchons donc à élaborer une autre méthode pour la rédaction des BLEC, qui n'est pas destinée à remplacer la méthode actuelle, mais qui lui est complémentaire.

Le chapitre est structuré de la façon suivante : nous présentons, dans un premier temps (section 6.2), la façon dont le système de FL a été implémenté. Cette brève présentation nous permet d'introduire, dans un deuxième temps, l'outil d'aide à l'encodage des liens lexicaux au moyen des FL (section 6.3), directement lié à cette représentation du système de FL. Enfin, dans la section 6.4, nous présentons des heuristiques pour l'automatisation de la rédaction de certains champs de fiches lexicographiques.

6.2 Implémentation du système de fonctions lexicales

Nous avons choisi d'implémenter le système de FL sous la forme d'une table de données, qui s'inscrit dans la base de données relationnelle plus générale du DiCo,

développée par Alain Polguère à l'OLST². Cette table contient six champs (cf. figure 6.1) dans lesquels figurent des codes renvoyant aux différentes classes du système de FL.

fonction	pdd	sém	syntp	cible	FL_inverse
Magn	3.1, 4.2, 4.3	4.3.1	Mod.1	#	–
Oper ₁₂	2.1, 2.2, 2.4	1.3	3.7	1	–
QSyn	1.1, 2.3, 3.2, 4.4, 5.5	1.1	–	#	QSyn ⁻¹
S ₃	1.1, 1.3	5.1	–	3	–

Figure 6.1 – Système de FL représenté sous forme de table de données

Le champ **fonction** est généré automatiquement lors de la compilation du DiCo, il contient le nom de la FL.

Le champ **pdd** renvoie aux classes du système de FL représentant les couples de parties du discours associés aux FL.

Dans le champ **sém** sont représentées les classes sémantiques.

Le champ **syntp** contient les codes renvoyant aux classes de valences syntaxiques profondes.

Dans le champ **cible** est indiqué quelle est la cible communicationnelle visée par la relation. Les valeurs du champ sont le mot-clé de la FL : #, les différents actants du mot-cé : 1, 2, 3, 4, 5 et 6, un actant extérieur : @ ou la valeur de la FL : V.

Enfin, le dernier champ **FL_inverse** présente, le cas échéant, la ou les FL inverses associées à une FL donnée³.

Le système de codes est établi en fonction des classes de chaque perspective, élaborées et présentées au chapitre précédent dans les figures et tableaux 5.5 (p. 188), 5.9 (p. 203), 5.I (p. 211) et 5.14 (p. 221). La correspondance entre les classes et les codes est présentée à l'annexe III. Cette codification reflète la structuration hiérarchique du système en classes et sous-classes, et offre ainsi la possibilité d'effectuer

²La table représentant le système de FL est disponible sur le site de l'OLST : <http://www.olst.ling.umontreal.ca>.

³La notion de FL inverse est définie plus loin dans ce chapitre 6.4.6.1 (p. 244).

des requêtes très précises ou, au contraire, plus larges. Par exemple, si l'on interroge la table sur les parties du discours, on peut spécifier que la valeur sera verbale (code pdd 2) sans forcément spécifier ce que sera le mot-clé. Au contraire, si l'on cherche à extraire des FL verbales prenant pour mot-clé un adjectif, on indiquera le code pdd 2.2.

6.3 Assistant à l'encodage des liens lexicaux

Nous pouvons dériver directement, à partir du système de FL et de sa représentation sous forme de codes, un assistant à l'encodage des liens lexicaux. Celui-ci n'est autre qu'une modélisation du processus d'encodage des liens lexicaux. Il consiste à envisager tour à tour différents paramètres associés aux relations lexicales qui ont été modélisés dans le système de FL⁴. On propose à un utilisateur de sélectionner un ensemble de paramètres descriptifs pour un lien lexical. Chaque fois qu'un paramètre est sélectionné, le nombre de FL potentielles diminue, jusqu'à ne conserver que les FL les plus susceptibles de modéliser le lien.

6.3.1 Sélection des différents paramètres

Pour sélectionner les paramètres caractérisant un lien lexical, on procède de façon déclarative, c'est-à-dire qu'on récolte un ensemble d'informations sur le lien, de façon non ordonnée. Illustrons le processus avec le lien entre PISTE#I.2 ('voie'), dont la structure actantielle est la suivante :

~ [UTILISÉE PAR L'être animé X] MENANT AU lieu Y [À PARTIR DU lieu Z]
 et le verbe S'ENGAGER, par exemple dans : *Pierre s'engage dans une piste impraticable.*

– Critères liés aux parties du discours

Quel couple de parties du discours la relation concerne-t-elle ? Il s'agit du couple

$N \rightarrow V$

⁴Notons que Alonso Ramos [2000] a mis en évidence des critères heuristiques pour l'encodage des collocations au moyen de FL.

→ On élimine, dans la table de FL, toutes les FL qui ne correspondent pas potentiellement à ce couple.

– **Critères syntaxiques**

- La valeur entretient-elle une relation actantielle ou attributive avec le mot-clé? Le mot-clé est ici un actant de la valeur.

→ On élimine toutes les FL qui ont, pour patron de dépendance, un rôle de modificateur du mot-clé ou d'un de ses actants.

- Combien d'actants a la valeur y compris le mot-clé? La valeur a deux actants (X et le mot-clé).

→ On élimine toutes les FL qui ont moins ou plus de deux actants.

- Quelle est la position syntaxique profonde du mot-clé dans la relation? Le mot-clé est le deuxième actant de la valeur.

→ On élimine toutes les FL pour lesquelles le mot-clé serait le premier actant (par exemple : Fact₁).

– **Critères communicationnels**

Quelle est la cible communicationnelle de la relation? Il s'agit du premier actant de PISTE#I.2, qui est sujet grammatical du verbe et thème de l'énoncé potentiellement formé.

→ On élimine toutes les FL qui n'ont pas le premier actant du mot-clé pour cible communicationnelle.

– **Critères sémantiques**

Sélectionner une ou plusieurs classes ou sous-classes sémantiques.

On sélectionne la sous-classe : « début » ainsi que la sous-classe « Utilisation ».

→ On élimine les FL représentant les catégories sémantiques non sélectionnées.

Dans le meilleur des cas, l'assistant ne proposera qu'une seule FL pour l'encodage d'un lien donné. Toutefois, ce résultat unique n'étant pas la situation la plus fréquente, nous devons envisager les cas où l'assistant renvoie plus d'une proposition. Deux types d'information permettent à un utilisateur de choisir entre les résultats proposés : les gloses de vulgarisation et les autres paramètres descriptifs. Pour la

présente requête, par exemple, deux FL sont proposées : PreparReal_1 et IncepReal_1 . Les gloses de ces deux FL peuvent aider à trancher entre les deux. PreparReal_1 est vulgarisée par la glose $\rightarrow [\text{X}] \text{ préparer} \sim$ et IncepReal_1 , par l'ensemble de gloses suivantes : $\rightarrow [\text{X}] \text{ commencer à utiliser/occuper/se trouver dans} \sim$. Ce sont les gloses associées à la seconde FL qui semblent le mieux correspondre à la relation qui nous concerne. L'examen des gloses pousse donc en faveur de la FL IncepReal_1 pour représenter la relation.

Dans d'autres cas, ce sont les paramètres descriptifs qui permettent de trancher entre les différentes solutions offertes. En effet, si les FL proposées ont les critères demandés en commun, rien n'indique qu'elles partagent l'ensemble de leurs paramètres descriptifs. Prenons pour exemple les deux FL AntiBonFact_0 et Degrad illustrées ci-dessous :

$\text{Degrad}(\text{maison}\#\text{I.1}) = \text{se délabrer, tomber en ruine} (\rightarrow \sim \text{se dégrade}) ;$

$\text{AntiBonFact}_0(\text{robinet}) = \text{couler, fuir} (\rightarrow \sim \text{fonctionne mal}) .$

Ces deux FL possèdent beaucoup de paramètres communs : elles décrivent des liens entre un nom et un verbe ($\text{N} \rightarrow \text{V}$), sont associées à un patron de structure actantielle à un seul actant, ont le mot-clé pour cible communicationnelle et possèdent une composante sémantique de qualification négative. Ce qui les distingue, en revanche, c'est qu'en plus de cette composante sémantique commune, chacune d'entre elles en possède une autre qui lui est propre :

Degrad : qualificatif négatif + diminution/dégradation ;

AntiBonFact_0 : qualificatif négatif + utilisation/fonctionnement typique.

C'est donc, au moment de la visualisation des résultats, l'absence ou la présence de certains paramètres qui permettra de trancher en faveur d'une FL particulière.

6.3.2 Interrogation sous forme de requêtes SQL

Afin d'exploiter concrètement le système de FL et d'effectuer des requêtes comme celles que nous venons d'exposer, on se réfère à la base de données de FL au format MySQL. Prenons un exemple précis pour illustrer la façon dont est formée une requête SQL : on cherche à encoder la relation entre les lexies $\text{COMPLIMENT}\#\text{A}$ et

DÉBITER. Nous distinguons plusieurs étapes pour illustrer à la fois le langage SQL et la façon dont la sélection de différents critères permet de restreindre le nombre de FL ramenées par la requête.

Dans un premier temps, on utilise les critères sur le sens de la FL et sur les parties du discours de son argument et de ses valeurs : on sélectionne la composante sémantique « Qualificatif négatif » (correspondant au code **4.1.2** du champ **sém**) et on indique que la valeur de la FL est verbale (ce qui correspond au code **2** du champ **pdd**). La requête est donc formulée comme suit :

```
SELECT * FROM fonction WHERE sém like '%4.1.2%' AND pdd like
      '%2%'
```

La requête retourne les onze résultats suivants :

AntiBon.AntiVer.Oper1	AntiBonReal1-manifestation
AntiBon.Oper12	AntiVerFact0/AntiBonFact0
AntiBonFact0	Degrad
AntiBonFact1	IncepPredAntiBon-usual
AntiBonFact3	LiquPredVer
AntiBonInvolv	

Ces FL étant trop nombreuses, on cherche, dans un deuxième temps, à raffiner la requête. On peut alors indiquer, par exemple, que le lien lexical a pour cible communicationnelle le premier actant du mot-clé, qui est le sujet grammatical du verbe DÉBITER (ce qui se traduit par le code : **1** du champ **cible**). La requête devient alors :

```
SELECT * FROM fonction WHERE sém like '%4.1.2%' AND pdd like
      '%2%' AND cible like '1'
```

Ceci restreint le nombre de FL à trois :

AntiBon.AntiVer.Oper1
AntiBon.Oper12
AntiBonReal1-manifestation.

Enfin, une troisième étape précisant la configuration syntaxique de la relation, c'est-à-dire qu'il s'agit d'une relation actantielle à trois actants (soit le code **3** du champ **syntp**), permet de restreindre davantage les résultats de la requête :

```
SELECT * FROM fonction WHERE sém like '%4.1.2%' AND pdd like
    '%2%'
    AND cible like '1' AND syntp like '%3%'
```

La requête ainsi précisée permet d'éliminer deux FL pour ne laisser que la FL AntiBon.Oper12, que l'on sélectionne alors pour représenter la relation :

AntiBon.Oper12(compliment#a) = débiter.

En plus d'aider les lexicographes à encoder les relations lexicales, nous pensons que cet assistant peut être utilisé pour l'enseignement des FL. Les paramètres déterminés pour chaque FL apportent une dimension didactique, dans la mesure où il devient très facile de comparer des FL similaires, et, plus généralement, de visualiser ce qui différencie ou rapproche un ensemble de FL.

6.4 Calculs de données lexicographiques pour l'ébauche automatique de fiches BLEC

6.4.1 Objectif

Il existe deux façons d'enrichir automatiquement une base de données lexicale. On peut utiliser des ressources extérieures (ou exogènes) ou exploiter l'information de façon endogène, c'est-à-dire récupérer toute l'information qui est contenue dans une base de données sans qu'elle y soit explicitée. Pour la stratégie exogène, on peut citer les travaux de Wanner et al. [2006] et Wanner et Bohnet [2007], qui cherchent à identifier et caractériser syntaxiquement et sémantiquement les collocations en corpus de façon automatique. Nous nous intéressons, quant à nous, à l'exploitation endogène des données du DiCo, d'une part, parce que le DiCo contient un grand nombre d'informations non exploitées, d'autre part, parce que ceci va permettre de contribuer à la cohérence d'un système de relations lexicales et, plus largement,

de descriptions « où tout se tient ». Nous pensons qu'en termes d'informations données sur chaque unité lexicale, le DiCo est une des ressources les plus détaillées réalisées sur le français, cependant, le revers de la médaille est qu'il n'est que très peu développé au niveau du nombre d'unités lexicales décrites. Il est, en effet, très long de créer les différentes fiches d'un vocable. C'est la raison pour laquelle il semble nécessaire d'automatiser certaines tâches.

Notre objectif est de mettre à profit le système de FL, constitué et présenté au chapitre précédent, afin d'effectuer des calculs sur certaines informations lexicographiques, qui aboutiront, dans de futurs travaux, à la génération automatique d'ébauches de fiches de dictionnaire. Nous n'entendons pas par *fiche lexicographique*, une fiche traditionnelle de DiCo au format FileMaker mais, plus largement, un ensemble de champs associés à une lexie, et ce, quelque soit le format utilisé. La méthodologie de rédaction que nous souhaitons élaborer consiste à développer la rédaction du DiCo non pas fiche par fiche, mais plutôt champ par champ, voire, pour ce qui est du champ des relations lexicales, FL par FL. Chaque information se verra assigner un indice de fiabilité suivant ceux qui existent déjà pour le DiCo : 0, 1, 2 et 3. De cette façon, un utilisateur n'aura pas uniquement accès aux fiches complètes, mais à chaque information validée (c'est-à-dire au statut 0) pour une fiche, même si celle-ci demeure incomplète. Nous tentons de calculer les différents champs d'une fiche et utilisons certains de ces champs comme des indices pour déterminer le caractère polysémique ou homonymique d'un ensemble de lexies. Nous menons une expérimentation sur toutes les lexies du DiCo, y compris sur celles déjà décrites, ce qui a l'avantage, d'une part, de vérifier si nos inférences sont exactes et, d'autre part, de permettre une révision des fiches dans les cas où certaines erreurs sont révélées. L'expérimentation a été réalisée à l'aide de petits scripts et validée manuellement, cependant, elle n'est pas quantifiée : il s'agit d'un travail préalable proposant un algorithme pour la confection d'un programme informatique ultérieur.

À la suite de la présentation des différents champs (section 6.4.2), nous exposons notre méthode d'enrichissement endogène du DiCo en présentant chacun des

champs concernés et la façon dont on peut calculer leurs valeurs (sections 6.4.3, 6.4.4, 6.4.5 et 6.4.6). Finalement, (section 6.4.7), nous montrons comment il est possible de combiner un certain nombre de critères pour tenter de reconstituer partiellement les différentes fiches de lexies en relation de polysémie ou d’homonymie.

6.4.2 Présentation des différents champs

Pour donner une illustration d’une fiche lexicographique et de ses différents champs, nous prenons l’exemple de la fiche de la lexie ANCRE reproduite à la figure 6.2.

- Le champ **Nom** contient le nom de la lexie, son numéro d’acception et le statut indiquant la validité de la fiche. Dans l’exemple de la lexie ANCRE, on ne trouve pas de numéro d’acception puisqu’une seule lexie est décrite dans le DiCo⁵. La fiche est marquée au statut 0, ce qui signifie qu’elle a été validée.
- Le champ **cg** présente les caractéristiques grammaticales d’une lexie.
- Le champ **Étiq** représente l’étiquette sémantique de la lexie décrite. Il s’agit d’une brève caractérisation du sens de la lexie⁶.
- Le champ **fp** précise la structure actantielle de la lexie, c’est-à-dire le nombre de positions actantielles qu’elle contrôle et, le cas échéant, le typage de celles-ci. Par exemple, la lexie ANCRE possède un actant X de type sémantique ‘bateau’.
- Le champ **nb** contient des notes éventuellement écrites par le lexicographe qui a rédigé la fiche.
- Le champ **tr** présente le régime de la lexie, c’est-à-dire la correspondance entre ses actants sémantiques et syntaxiques et la façon dont les actants syntaxiques sont exprimés en surface.
- Le champ **Syn** contient les FL de type synonymique, soit les FL **Syn**, **QSyn**, **Gener**, **Cf** et **=**.
- Le champ **fl** présente toutes les autres FL.

⁵Il existe d’autres acceptions du vocable ANCRE mais elles relèvent de domaines spécialisés comme l’informatique et ne sont pas représentées dans le DiCo.

⁶Il existe une hiérarchie d’étiquettes sémantiques développée par Alain Polguère et accessible à l’adresse suivante : <http://olst.ling.umontreal.ca/dicopop-laf/hierarchie.php>. Il convient d’utiliser le navigateur Firefox pour la visualiser.

Nom	ANCRE	0	N	V
cg	nom, fém			
Étiq.	instrument			
fp	: ~ DU bateau X			
nb				
tr	X = I = de N, A-poss			
Syn	{Cf} grappin			
fl	<pre> /*De grande taille*/ {Magn} grosse antépos; grande /*[X <Équipage de X> faire usage de ~*/ {Reall--I/Reall[1]--I} jeter, mouiller [l'~] ["Le navire a jeté/mouillé l'ancre."] //mouiller ["Le navire a mouillé."] /*[Équipage de X] causer que X soit retenu par ~*/ {Labreal[1]1--I} mettre [N=X à l'~] //ancrer [N=X] ["Le capitaine a ancré le navire."] /*Fait de causer que X soit retenu par ~*/ {S0Labreal[1]1--I} ancrage, mouillage [de N=X] /*[X] être retenu par ~*/ {Reall--II} être [à l'~], se tenir [à l'~] </pre>			
ex	Le croiseur est arrivé hier à Malte, où il a jeté l'ancre. Les navires retenus en aval et en amont de l'embâcle ont finalement pu lever l'ancre.			
ph	_ancre de salut_ _jeter l'ancre_ _lever l'ancre_			

Figure 6.2 – Fiche DiCo de la lexie ANCRE

- Le champ **exemple** regroupe quelques exemples illustrant l’usage de la lexie en contexte.
- Le champ **ph** contient tous les phrasèmes dans lesquels la lexie figure.

Dans les ébauches de fiche que nous générerons, nous ne cherchons à reproduire que les champs **Nom**, **cg**, **Étiq** et **fl**, et fusionnons le champ **Syn** et le champ **fl**.

6.4.3 Parties du discours

6.4.3.1 Présentation

Le système de FL présenté au chapitre précédent attribue à chaque FL une ou plusieurs classes représentant les parties du discours de son mot-clé et de sa valeur. Il est donc possible, en se référant aux classes associées à une FL, de déterminer automatiquement la partie du discours de sa valeur. Par exemple, **IncepPredPlus** ($\rightarrow \sim$ devient plus intense) est associée au schéma suivant : $N \rightarrow V$. Ses valeurs seront, par conséquent, verbales, par exemple :

$\text{IncepPredPlus}(\text{compétition}\#1) = s'intensifier ;$

$\text{IncepPredPlus}(\text{abcès}\#I) = m\hat{u}rir.$

Une FL peut être associée à plus d’un couple de parties du discours. C’est le cas de S_0 (\rightarrow équivalent nominal), par exemple, qui produit des valeurs uniquement nominales mais qui peut être appliquée à des verbes ou à des adjectifs (et éventuellement à des adverbes) comme on le voit ci-dessous :

$S_0(\text{curieux}\#I.1) = curiosité\#1 ;$

$S_0(\text{admirer}\#I) = admiration.$

D’autres encore peuvent avoir non seulement plusieurs types de mots-clés différents, mais aussi plusieurs types de valeurs, par exemple **Magn** (\rightarrow intense) et **AntiBon** (\rightarrow mauvais) ci-dessous :

$\text{Magn}(\text{autopsie}) = approfondie, m\acute{e}ticuleuse ;$

$\text{Magn}(\text{chanter}(1)\#1a) = fort, \lceil \text{à gorge déployée} \rceil ;$

$\text{AntiBon}(\text{corde}\#I.1) = effilochée, usée ;$

$\text{AntiBon}(\text{chanter}(1)\#1a) = mal, faux.$

On doit, par conséquent, distinguer :

1. les FL pour lesquelles la partie du discours de la valeur est unique : par exemple, *IncepPredPlus* et S_0 ;
2. les FL pour lesquelles la partie du discours de la valeur varie en fonction de celle du mot-clé :
 - elle peut être équivalente à celle du mot-clé : par exemple, *Syn*, *Anti* et *Conv_{ij}*,
 - ou différente : par exemple, *Magn*, *AntiBon*, etc.

Ces différents cas conduisent à des traitements particuliers : dans le premier, la FL seule permet de déterminer la partie du discours d'une valeur ; dans les deux autres, on doit également avoir recours à la partie du discours du mot-clé pour déterminer celle de la valeur.

6.4.3.2 Expérimentation

Le calcul d'informations lexicales sert un double objectif dans la mesure où il permet, en parallèle, un diagnostic des erreurs pour la révision des fiches.

◇ Diagnostic d'erreurs

La recherche des parties du discours constitue un outil pour la vérification de l'encodage des données. Dans le DiCo, lorsqu'un même mot-forme est associé à deux lexies de parties du discours différentes, il reçoit un marqueur venant préciser de quelle partie du discours il s'agit : prenons l'exemple de *vivant*, noté de façon différente dans les deux relations ci-dessous :

$$A_1(vie\#I.1) = vivant_{Adj} (\rightarrow [X] \text{ qui a } \sim);$$

$$S_1(vie\#I.1) = vivant_N (\rightarrow \text{Nom pour } X).$$

La génération automatique des parties du discours des lexies non décrites, fondée sur le système de FL, met en évidence un certain nombre d'oublis et de manques d'homogénéité dans les données. Par exemple, *mourant* dans la relation suivante,

$$A_1ProxOper_1(mort(1)\#I.1) = mourant (\rightarrow [X] \text{ qui est proche de } \sim),$$

aurait dû être noté : $mourant_{Adj}$. De la même façon, $vainqueur$, dans les relations ci-dessous, devrait être noté $vainqueur_N$, etc.

$S_1 Real_2$ issue favorable($compétition\#2$) = $vainqueur$ [de ART \sim] (\rightarrow Nom pour Y qui gagne \sim);

$S_1(couronne\#II)$ = $vainqueur$.

◇ Indice pour la distinction de plusieurs vocables

On note que la présence de deux parties du discours associées à un même mot-forme figurant en tant que valeur de FL constitue un indice de l'existence de deux vocables, qui devront être distingués.

Par exemple, $porteur$ est valeur des FL S_3 (\rightarrow Nom pour Z) et A_1 :

$S_3(chèque)$ = $porteur$ [de ART \sim];

$A_1(espoir\#I)$ = $porteur$ [de (ART) \sim].

Dans le premier cas, il s'agit d'une lexie nominale, dans le second, d'une lexie adjectivale. De même, $malade$ est valeur des FL $Magn.A_1$ (\rightarrow [X] qui éprouve un intense \sim) et S_2 (\rightarrow Nom pour Y) :

$Magn.A_1(dégoût)$ = $malade$ [de \sim];

$S_2(docteur\#I)$ = $malade$.

Dans le premier cas, MALADE est un adjectif, dans le second, un nom.

Le critère de la partie du discours associée à une valeur sera utilisé dans la section 6.4.7 pour déterminer automatiquement le nombre de vocables devant être distingués.

6.4.4 Marques d'usage

6.4.4.1 Présentation

Le DiCo associe un certain nombre de marques d'usage aux lexies décrites : familier ("fam"), vulgaire ("vulg"), péjoratif ("péjor"), vieilli ("vieilli"), vieillissant ("vieillissant"), littéraire ("litt"), soutenu ("soutenu"), officiel ("offic"), formel ("formel"), spécialisé ("spéc"), religieux ("relig"), rural ("rural"), ironique ("iron"), enfantin ("enfantin"), affectueux ("affectueux"), néologique ("néol"), anglicisme ("angl"), québécoisme

(“Québ”), francisme (ou relatif au français de France) (“ffr”). En s’appuyant sur ces notations, il est possible de déterminer automatiquement les marques d’usage des valeurs de FL pour lesquelles on veut créer une fiche lexicographique. Pour chaque valeur, on pourra créer automatiquement un champ « marque d’usage » à l’intérieur de la fiche qui lui est dédiée. Par exemple, pour la lexie \lceil VIEILLE FILLE \rceil , on pourra exploiter les informations suivantes :

$$\text{QSyn}(\text{célibataire}) = \text{“péjor” } \lceil \text{vieille fille} \rceil,$$

et déduire que le champ « marque d’usage » de sa fiche doit comporter la mention : “péjor”. Il est également possible de prendre en compte les marques d’usage lorsque l’on décrit automatiquement certains liens lexicaux (cf section 6.4.6). Ainsi, à partir de la relation suivante :

$$\text{QSyn}(\text{colle}\#\text{II nom, fém, “fam”}) = \text{problème},$$

on pourra déduire :

$$\text{QSyn}(\text{problème}) = \text{“fam” colle}\#\text{II}.$$

6.4.4.2 Expérimentation

Une fois de plus, l’expérimentation menée a montré qu’outre le fait de remplir automatiquement ce champ dans les fiches de DiCo, l’analyse des marques d’usage des valeurs sert deux autres tâches : repérer des éventuelles erreurs ou incohérences dans l’étiquetage et distinguer des lexies en relation de polysémie ou d’homonymie. Devant plusieurs marques associées à un même mot-forme, le lexicographe devra donc déterminer s’il s’agit d’une lacune ou, au contraire, d’un indice d’une pluralité d’acceptions.

◇ Diagnostic d’erreurs ou oublis

La valeur *bibine* est marquée comme usage familier dans un cas et reste non marquée dans un autre :

$$\text{AntiBon}(\text{bière}(1)\#\text{a}) = \text{“fam” bibine};$$

$$\text{AntiBon}(\text{boisson}\#\text{1}) = \text{bibine}.$$

De la même façon, on remarque que *disert* est, dans les trois relations ci-dessous, tour à tour non marqué, soutenu et vieilli :

$\text{Ver/Bon}(\textit{avocat}(1)\#\text{I}) = \textit{disert} (\rightarrow \text{Bon}) ;$

$A_1(\textit{éloquence}) = \text{“soutenu” } \textit{disert} ;$

$\text{QA}_1(\textit{verve}) = \text{“vieilli” } \textit{disert}.$

gallinacé en tant que valeur de la même FL *Gener* (\rightarrow Terme générique) est marqué spécialisé et non marqué :

$\text{Gener}(\textit{coq}) = \textit{gallinacé} ;$

$\text{Gener}(\textit{poule}\#\text{I}) = \text{“spéc” } \textit{gallinacé}.$

Nous sommes clairement ici en présence de lacunes dans l’encodage des données du DiCo. Cette expérience aura permis une correction importante des marques d’usage.

◇ **Indice pour la distinction de plusieurs vocables**

Mis à part les lacunes dans la description des liens, la présence de deux marques d’usage distinctes pour une même valeur se révèle un indicateur de polysémie du vocable potentiel à décrire ou d’homonymie entre lexies. Ces indices permettent de postuler l’existence de deux (cf. tab 6.I) ou trois (cf. tab 6.II) lexies distinctes pour un même vocable ou pour des vocables distincts.

La présence ou l’absence d’une marque d’usage n’est en aucun cas un critère suffisant pour déterminer le nombre de lexies distinctes dans la mesure où plusieurs lexies peuvent avoir la même marque d’usage ou être non marquées. Par exemple, *expirer* non marqué peut aussi bien vouloir dire ‘prendre fin’ que ‘recracher de l’air’, etc.

<i>boîte</i>	“non marqué” contenant	“fam” entreprise
<i>cuisiner</i>	“non marqué” faire la cuisine	“fam” interroger qqn
<i>chienne</i>	“non marqué” femelle du chien	“vulg” mauvais individu
<i>amalgame</i>	“non marqué” ensemble incohérent	“spéc” plombage
<i>bagne</i>	“non marqué” endroit où on purge une peine	“iron” lieu où on est forcé de faire un travail pénible
<i>expirer</i>	“non marqué” prendre fin/contraire d’inspirer	“litt” mourir
<i>puce</i>	“non marqué” petit insecte	“affectueux” surnom donné à une fillette
<i>s’éteindre</i>	“non marqué” cesser de fonctionner	“soutenu” mourir
<i>délivrer</i>	“non marqué” libérer	“officiel” procurer
<i>poulet</i>	“non marqué” petit de la poule	“vieilli” billet, message
<i>minet</i>	“enfantin” chat	“fam” homme léger
<i>répulsif</i>	“spéc” relatif à la force de répulsion	“soutenu” qui crée du dégoût

Tableau 6.I – Distinction de deux lexies à l’aide des marques d’usage

	“fam et péjor”	“enfantin”	“affectueux”
<i>cocotte</i>	femme légère	poule	surnom affectueux donné à une fille / femme
	“non marqué”	“offic”	“spéc”
<i>train</i>	véhicule pour le transport	ensemble de directives	ensemble des pneus d’un véhicule
	“non marqué”	“fam”	“vulg”
<i>crever</i>	percer	éprouver un sentiment intense	mourir

Tableau 6.II – Distinction de trois lexies à l’aide des marques d’usage

6.4.5 Étiquettes sémantiques

6.4.5.1 Présentation

Chaque lexie du DiCo est sémantiquement caractérisée par une étiquette sémantique. Il semble donc *a priori* possible, à partir d’une lexie décrite, d’inférer l’étiquette sémantique de ses synonymes et de ses équivalents sémantiques verbaux, nominaux, adjectivaux ou adverbiaux. Une telle généralisation peut également être faite à partir des genres prochains des définitions lexicographiques de n’importe

quel dictionnaire ou base de données lexicale.

◇ **Déduction de l'étiquette sémantique via les FL d'équivalence**

On peut reconstituer les étiquettes sémantiques des valeurs des FL d'équivalence : Syn, QSyn, Cf et =. Prenons les trois exemples suivants :

Syn(*vaisselle*#2 : lavage) = *plonge* ;

QSyn(*combat*#I.1 : confrontation physique) = *duel* ;

Cf(*ciment* : matériau de construction) = *chaux*.

Nous pouvons postuler que les trois valeurs *plonge*, *duel* et *chaux* ont les étiquettes sémantiques suivantes :

plonge : lavage ;

duel : confrontation physique ;

chaux : matériau de construction.

Cependant, la « transmission » des étiquettes via ces FL doit se faire avec prudence. Par exemple, si l'on peut postuler que les lexies CHAT et MINET sont synonymes, la lexie MINET n'hérite pas entièrement de l'étiquette de CHAT, qui est double : **animal de compagnie** ou **espèce animale correspondante**. En effet, il est relativement acceptable de dire qu'un minet est un animal de compagnie, mais absurde de dire qu'il s'agit d'une espèce animale. Il semble, par conséquent, plus prudent d'assigner un indice de confiance moins fort à ces données automatiquement générées (cf. Polguère [2009 : 45]). On attribuera le statut 0 aux étiquettes inférées par la FL =, le statut 1 pour celles inférées avec Syn et le statut 2 pour les étiquettes générées par les FL QSyn et Cf.

◇ **Déduction de l'étiquette sémantique via les FL de dérivation syntaxique**

À partir des équivalents verbaux, nominaux, adjectivaux ou adverbiaux, on peut également postuler une similitude entre étiquettes sémantiques. En effet, la hiérarchie d'étiquettes prévoit des formulations en fonction des parties du discours des lexies auxquelles elles s'appliquent. Par exemple, dans l'extrait de la hiérarchie

d'étiquettes reproduit à la figure 6.3, on s'aperçoit que la formulation verbale pour son expressif est émettre un son expressif.

- ▼ Son expressif
 - ◻ émettre un son expressif(V_0)
 - ▶ cri
 - ▶ ensemble de sons expressifs

Figure 6.3 – Extrait de la hiérarchie d'étiquettes sémantiques

En se référant aux équivalences prévues dans la hiérarchie, on déduira pour les valeurs des relations suivantes :

$V_0(\text{aboiment} : \text{cri d'animal}) = \text{aboyer}$,

$S_0(\text{crier}\#I.1 : \text{émettre un son expressif}) = \text{cri}\#I.1$,

$A_0(\text{ami}\#I.1 : \text{individu qui a une certaine relation interpersonnelle avec un autre}) = \text{ami}_{Adj}$,

les étiquettes sémantiques de leurs mots-clés reformulées pour correspondre à la partie du discours adéquate :

aboyer : émettre un cri d'animal,

cri#I.1 : son expressif,

ami_{Adj} : qui a une certaine relation interpersonnelle avec un autre.

6.4.5.2 Expérimentation

L'expérimentation menée pour reconstituer les étiquettes sémantiques des lexies non décrites a, d'une part, mis en valeur certaines lacunes et, d'autre part, apporté des indices pour la distinction de plusieurs lexies.

◇ Diagnostic d'erreurs

Le remplissage automatique révèle un certain nombre d'erreurs parmi les étiquettes déjà présentes dans le DiCo. Il permet notamment de repérer des étiquettes sémantiques trop générales. Par exemple, à partir de la relation ci-dessous, on a pu

reconstituer, via la lexie BÉTON, l'étiquette sémantique de la lexie CIMENT#I : matériau :

$$\text{QSyn}(\text{béton matériau}) = \text{ciment}\#I.$$

Or, la lexie CIMENT#I est décrite dans le DiCo par l'étiquette sémantique **matériau de construction**, ce qui met en valeur un manque d'homogénéité puisque BÉTON devrait avoir l'étiquette **matériau de construction**.

Par ailleurs, l'étiquette sémantique de la lexie *combat*#II.2 est reconstituée comme **action sociale** alors que son étiquette est **lutte sociale** (qui est l'étiquette fille de **action sociale**). Parfois, l'écart peut être plus grand qu'un lien direct mère-fille : la lexie COMPÉTITION#2, par exemple, est étiquetée **événement** dans le DiCo alors que la reconstitution par le biais d'une FL d'équivalence lui attribue l'étiquette **épreuve sportive**, qui semble plus adéquate.

◇ **Indice de l'existence de plusieurs lexies**

Là encore, la présence d'étiquettes différentes pour une valeur d'une ou plusieurs FL est un indice d'une pluralité de lexies. Par exemple, l'expérimentation de génération d'étiquettes sémantiques via les FL d'équivalence a permis de dégager deux étiquettes pour la valeur *cinoche* :

cinoche : établissement de spectacle ;

cinoche : comportement.

cocotte est associée à trois étiquettes :

cocotte : femme qui a un certain comportement ;

cocotte : le locuteur dit (cette étiquette est employée pour les énoncés directs) ;

cocotte : oiseau d'élevage ou espèce animale.

Parfois les différences entre étiquettes sont très fines, par exemple :

yogourt : aliment préparé ;

yogourt : portion d'aliment préparé ;

ce qui permet, dans certains cas, de rendre compte de la polysémie avec un bon degré de subtilité.

6.4.6 Calculs pour l'encodage des relations lexicales

Le système de FL réalisé au cours de cette thèse permet non seulement de formuler des règles pour la description automatique des champs du DiCo, comme les parties du discours ou l'étiquette sémantique, mais également de fournir des généralisations pour la génération et l'encodage automatique de FL. Nous explorons, dans cette section, une méthode fondée sur la propriété de réversibilité partagée par certaines FL, propriété que nous définissons ci-dessous (section 6.4.6.1). Nous montrons, par la suite (section 6.4.6.2), comment cette propriété peut être exploitée pour le calcul de relations lexicales.

6.4.6.1 Réversibilité de certaines fonctions lexicales

Les relations représentées par les FL sont orientées du mot-clé vers la valeur. La relation entre *précipitations* et *fortes*, par exemple, va être encodée par la FL **Magn** de la façon suivante : $\text{Magn}(\textit{précipitations}) = \textit{fortes}$. La relation est ici orientée du nom vers l'adjectif et non l'inverse. Cependant, certaines FL sont réversibles, c'est-à-dire que le mot-clé de l'une devient la valeur de l'autre et réciproquement. Par exemple, à partir de $\text{S}_0(\textit{aboyer}) = \textit{aboiement}$, on peut inférer $\text{V}_0(\textit{aboiement}) = \textit{aboyer}$ et inversement. C'est cette observation qui a conduit Grimes à rechercher des FL inverses comme nous les avons présentées dans la figure 5.2 (p. 164) et à créer de nouvelles FL pour former des couples de FL inverses.

Nous entendons par réversibilité, la capacité pour une FL à avoir une FL inverse. Les paires de FL inverses sont notées $f^{-1} = g$, par exemple : $\text{V}_0^{-1} = \text{S}_0$.

Notons que la propriété de réversibilité ne concerne pas toutes les FL, de plus, les FL concernées n'obéissent pas toutes au même schéma d'inférence. Nous distinguons différents cas de figure que nous répartissons en fonction des types d'inférences qu'ils sous-entendent.

◇ Inférence simple

Le premier type d'inférence concerne les cas où $f^{-1} = g$ sans que l'inverse soit possible : $g^{-1} \neq f$. On peut postuler une inférence selon la règle suivante : à partir

de la relation $f(L_1) = L_2$, on peut inférer la relation $g(L_2) = L_1$ mais non l'inverse : *si $g(L_2) = L_1$ alors $f(L_1) = L_2$. Nous formalisons cette règle de la façon suivante : $f(L_1) = L_2 \Rightarrow g(L_2) = L_1$.

Prenons l'exemple de Son^{-1} et QS_1 : à partir de la relation : $Son(\textit{cheval}) = \textit{hennir}$, on peut inférer la relation suivante : $QS_1(\textit{hennir}) = \textit{cheval}$. Toutefois, à partir de $QS_1(\textit{abdiquer}) = \textit{souverain}$, on ne peut déduire $*Son(\textit{souverain}) = \textit{abdiquer}$, dans la mesure où la FL QS_1 peut s'appliquer à un grand ensemble de verbes qui ne sont pas pour autant des verbes de son typique. Les paires de FL concernées sont présentées dans le tableau 6.III.

Relation	FL	Exemple
Verbe de son typique Premier actant	$Son^{-1} = QS_1$	$Son(\textit{grenouille}) = \textit{coasser}$ $QS_1(\textit{coasser}) = \textit{grenouille}$
Nom de son typique Premier actant	$S_0Son^{-1} = QS_1$	$S_0Son(\textit{cheval}) = \textit{hennissement}$ $QS_1(\textit{hennissement}) = \textit{cheval}$
Nom de dirigeant Deuxième actant	$Cap^{-1} = S_2$	$Cap(\textit{équipage}) = \textit{commandant}$ $S_2(\textit{commandant}) = \textit{équipage}$
Nom d'équipe Deuxième actant	$Equip^{-1} = S_2$	$Equip(\textit{navire}) = \textit{équipage}$ $S_2(\textit{équipage}) = \textit{navire}$
Nom d'instrument Verbe d'utilisation	$S_{instr}^{-1} = Labreal_{12}$	$Labreal_{12}(\textit{chiffon}\#a) = \textit{essuyer}$ [N=Y avec ART ~] $S_{instr}^{-1}(\textit{essuyer}) = \textit{chiffon}$

Tableau 6.III – Paires de FL à inférence simple

◇ Inférence double

Il existe des paires de FL qui sont inverses l'une de l'autre, c'est-à-dire pour lesquelles $f^{-1} = g$ et $g^{-1} = f$. Dans ce cas, à partir d'une relation $f(L_1) = L_2$, on peut postuler la relation inverse $g(L_2) = L_1$ et vice-versa, c'est-à-dire qu'il est également possible, à partir de la relation $g(L_2) = L_1$, d'inférer la relation $f(L_1) = L_2$. Par exemple, à partir de : $S_0(\textit{craindre}) = \textit{crainte}$, on peut inférer $V_0(\textit{crainte}) = \textit{craindre}$, et vice-versa. f et g peuvent ou non être identiques. Si elles sont identiques, on parlera d'un cas de FL réfléchie, dans le cas contraire, on dira qu'il s'agit d'une paire de FL inverses. Pour représenter la double implication, nous adoptons

la notation suivante : $f(\mathbf{L}_1) = \mathbf{L}_2 \equiv g(\mathbf{L}_2) = \mathbf{L}_1$.

• **FL réfléchie**

Les FL réfléchies représentent les différentes formes d'équivalence et d'opposition sémantiques. Elles sont détaillées dans le tableau 6.IV.

Relation	FL	Exemple
Synonymie	$\text{Syn}^{-1} = \text{Syn}$	$\text{Syn}(\textit{frigidaire}) = \textit{réfrigérateur}$
Quasi-synonymie	$\text{QSyn}^{-1} = \text{QSyn}$	$\text{QSyn}(\textit{bataille}\#\text{II.1}) = \textit{combat}\#\text{II.1}$
Relation confère	$\text{Cf}^{-1} = \text{Cf}$	$\text{Cf}(\textit{aboitement}\#\text{I}) = \textit{grognement}\#\text{I}$
Antonymie	$\text{Anti}^{-1} = \text{Anti}$	$\text{Anti}(\textit{ami}\#\text{II}) = \textit{ennemi}\#\text{II}$
Quasi-antonymie	$\text{QAnti}^{-1} = \text{QAnti}$	$\text{QAnti}(\textit{compliment}\#\text{a}) = \textit{réprimande}$
Conversion	$\text{Conv}_{21}^{-1} = \text{Conv}_{21}$	$\text{Conv}_{21}(\textit{craindre}) = \textit{effrayer}$ [N=X]
Relation contrastive	$\text{Contr}^{-1} = \text{Contr}$	$\text{Contr}(\textit{cigarette}\#\text{1}) = \textit{pipe}\#\text{I}$

Tableau 6.IV – FL réfléchies

• **Paire de FL inverses**

Nous avons relevé quelques paires de FL inverses dans le DiCo. Elles sont présentées dans le tableau 6.V.

Nous avons également repéré quelques relations potentiellement réversibles, mais dont l'une n'est pas encodée par une FL. Il s'agit de paires comprenant les dérivés actantiels adjectivaux (A_1 , A_2) et les dérivés adjectivaux potentiels (Able_1 , Able_2). Les FL Able_1 et Able_2 n'ont pas de FL inverses recensées alors que, comme on le voit à travers les exemples ci-dessous, il serait légitime de les créer.

$\text{Able}_2(\textit{estimer}\#\text{II.2}) = \textit{estimable}$ et $\text{Able}_2(\textit{estime}) = \textit{estimable}$ mais
 $?(\textit{estimable}) = \textit{estimer}\#\text{II.2}$ et $?(\textit{estimable}) = \textit{estime}$.

$\text{Able}_1(\textit{haïr}) = \textit{haineux}$ mais
 $?(\textit{haineux}) = \textit{haïr}$.

On observe le même phénomène pour les dérivés actantiels adjectivaux :

Relation	FL	Exemple
Hollonymie Méronymie	Hollo ⁻¹ = Mero Mero ⁻¹ = Hollo	Hollo(<i>ballast</i>) = <i>bateau</i> Mero(<i>bateau</i>) = <i>ballast</i>
Terme générique Hyponymie	Gener ⁻¹ = Hypo Hypo ⁻¹ = Gener	Gener(<i>avertissement</i> #I.2b) = <i>sanction</i> Hypo(<i>sanction</i>) = <i>avertissement</i> #I.2b
Ingrédient typique de ~ Produit dérivé de ~	Compos ⁻¹ = Produit dérivé Produit dérivé ⁻¹ = Compos	Compos(<i>clafoutis</i>) = <i>cerise</i> Produit dérivé(<i>cerise</i>) = <i>clafoutis</i>
Équivalent féminin Équivalent masculin	Fem ⁻¹ = Masc Masc ⁻¹ = Fem	Fem(<i>coq</i>) = <i>poule</i> #I Masc(<i>poule</i> #I) = <i>coq</i>
Unité minimale Ensemble	Sing ⁻¹ = Mult Mult ⁻¹ = Sing	Sing(<i>dialogue</i> #1b) = <i>réplique</i> Mult(<i>réplique</i>) = <i>dialogue</i> #1b
Nominalisation Verbalisation	S ₀ ⁻¹ = V ₀ V ₀ ⁻¹ = S ₀	S ₀ (<i>abdiquer</i> #I) = <i>abdication</i> #I V ₀ (<i>abdication</i> #I) = <i>abdiquer</i> #I
Quasi-nominalisation Quasi-verbalisation	QS ₀ ⁻¹ = QV ₀ QV ₀ ⁻¹ = QS ₀	QS ₀ (<i>hair</i>) = <i>haine</i> QV ₀ (<i>haine</i>) = <i>hair</i>
Adjectivalisation Verbalisation	A ₀ ⁻¹ = V ₀ V ₀ ⁻¹ = A ₀	A ₀ (<i>se suicider</i>) = <i>suicidaire</i> V ₀ (<i>suicidaire</i>) = <i>se suicider</i>
Adjectivalisation Nominalisation	A ₀ ⁻¹ = S ₀ S ₀ ⁻¹ = A ₀	A ₀ (<i>fleuve</i> #I) = <i>fluvial</i> S ₀ (<i>fluvial</i>) = <i>fleuve</i> #I
Fait que X soit ~ tel que X est ~	S ₀ Pred ⁻¹ = A ₁ A ₁ ⁻¹ = S ₀ Pred	S ₀ Pred(<i>célibataire</i>) = <i>célibat</i> A ₁ (<i>célibat</i>) = <i>célibataire</i>

Tableau 6.V – Paires de FL inverses

$A_1(\textit{consternation}) = \textit{consterné}$ mais

$?(consterné) = \textit{consternation}$.

$A_2(\textit{consternation}) = \textit{consternant}$ mais

$?(consternant) = \textit{consternation}$.

6.4.6.2 Calculs de relations lexicales à partir des FL inverses

En s'appuyant sur les deux types d'inférences qui viennent d'être présentés, nous pouvons calculer un certain nombre de relations lexicales. Par exemple, à partir des six relations ci-dessous, il est possible d'inférer six autres relations réversibles (notées à droite des premières) :

(126) $\text{Syn}(\textit{abeille}) = \ulcorner \textit{mouche à miel} \urcorner \equiv \text{Syn}(\ulcorner \textit{mouche à miel} \urcorner) = \textit{abeille}$;

(127) $\text{Conv}_{21}(\textit{craindre}) = \textit{effrayer} [N=X] \equiv \text{Conv}_{21}(\textit{effrayer}) = \textit{craindre} [N=X]$;

- (128) $S_0(\textit{homicide}_{\text{Adj}}) = \textit{homicide}_{\text{N}} \equiv A_0(\textit{homicide}_{\text{N}}) = \textit{homicide}_{\text{Adj}}$;
 (129) $S_{\text{instr}}(\textit{fumer}\#2) = \textit{fume-cigarette} \Rightarrow \text{Labreal}_{12}(\textit{fume-cigarette}) = \textit{fumer}\#2$;
 (130) $\text{Cap}(\textit{assemblée}\#2) = \textit{président} \Rightarrow S_2(\textit{président}) = \textit{assemblée}\#2$;
 (131) $\text{Son}(\textit{abeille}) = \textit{bourdonner} \Rightarrow S_1(\textit{bourdonner}) = \textit{abeille}$.

◇ Surgénération de relations lexicales

Les règles de réversibilité des FL s'appuient sur une observation empirique des relations encodées dans le DiCo. Par conséquent, on doit observer une certaine précaution vis-à-vis des liens lexicaux générés. Ces derniers sont, en effet, parfois très généraux et il n'est pas toujours pertinent de les faire figurer dans une fiche lexicographique. Par exemple, la relation générée en (130) est valide, cependant, si l'on applique la règle de réversibilité $\text{Cap}(L_1) = L_2 \Rightarrow S_2(L_2) = L_1$ à l'ensemble du lexique, on risque d'obtenir un grand nombre de valeurs pour la lexie PRÉSIDENT :

$S_2(\textit{président}) = \textit{assemblée}\#2$; *conférence\#I.1, parti, association, conseil, république, séance, entreprise, université, etc.*

Or, il n'est peut-être pas pertinent d'inclure l'ensemble de ces valeurs dans une fiche dédiée à PRÉSIDENT. Cependant, si l'on envisage la mise en relation de ces données dans un cadre plus large que la seule application lexicographique, c'est-à-dire comme formant un réseau de connexions explicites des éléments du lexique, ce qui est considéré comme du bruit pour constituer des fiches de dictionnaire ne pose plus problème. Par conséquent, lors de l'étape de validation des fiches, le lexicographe devra retenir les valeurs qu'il juge les plus pertinentes et écarter les autres sans pour autant rejeter la règle qui a permis de les générer.

◇ Diagnostic des erreurs d'encodage

Les règles de réversibilité se révèlent utiles pour le diagnostic des erreurs d'encodage. Prenons un exemple avec les FL S_0 et A_0 , qui correspondent aux équivalents nominaux et adjectivaux. On doit entendre par équivalent adjectival un adjectif de type relationnel paraphrasable par « relatif à » comme dans les paires de lexies FAMILLE/FAMILIAL, ENFANT/ENFANTIN, CHAT/FÉLIN, etc. Cette FL ne doit pas être

confondue avec A_1 qui modélise un modificateur adjectival du premier actant d'une lexie, comme dans les paires ADMIRATION/ADMIRATIF, ÉLOQUENCE/ÉLOQUENT, ZÈLE/ZÉLÉ, etc. Nous avons relevé, dans le DiCo, les relations suivantes que nous marquons comme erronées à l'aide d'un astérisque :

$$*S_0(\textit{curieux}\#\text{I.1}) = \textit{curiosité}\#1;$$

$$*S_0(\textit{fier}\#\text{I}) = \textit{fierté}\#\text{I.1};$$

$$*S_0(\textit{gai}\#\text{a}) = \textit{gaîté}\#\text{a};$$

$$*S_0(\textit{têtu}) = \textit{entêtement}.$$

Ces erreurs ont été mises en valeur par l'application de la règle $S_0^{-1} = A_0$, qui a généré les relations inverses :

$$*A_0(\textit{curiosité}\#1) = \textit{curieux}\#\text{I.1};$$

$$*A_0(\textit{fierté}\#\text{I.1}) = \textit{fier}\#\text{I};$$

$$*A_0(\textit{gaîté}\#\text{a}) = \textit{gai}\#\text{a};$$

$$*A_0(\textit{entêtement}) = \textit{têtu}.$$

Ces relations étaient elles-mêmes erronées, puisque les valeurs ne correspondent pas aux équivalents adjectivaux des mots-clés, mais aux modificateurs de leurs premiers actants ('tel que X a de la curiosité#1', etc.), c'est-à-dire, à la FL A_1 . L'encodage initial adéquat n'est pas S_0 (\rightarrow Équivalent nominal), mais $S_0\text{Pred}$ (\rightarrow Fait que X soit \sim). On peut alors déclencher la règle $A_1^{-1} = S_0\text{Pred}$ et générer les descriptions suivantes :

$$S_0\text{Pred}(\textit{curieux}\#\text{I.1}) = \textit{curiosité}\#1 \equiv A_1(\textit{curiosité}\#1) = \textit{curieux}\#\text{I.1};$$

$$S_0\text{Pred}(\textit{fier}\#\text{I}) = \textit{fierté}\#\text{I.1} \equiv A_1(\textit{fierté}\#\text{I.1}) = \textit{fier}\#\text{I};$$

$$S_0\text{Pred}(\textit{gai}\#\text{a}) = \textit{gaîté}\#\text{a} \equiv A_1(\textit{gaîté}\#\text{a}) = \textit{gai}\#\text{a};$$

$$S_0\text{Pred}(\textit{têtu}) = \textit{entêtement} \equiv A_1(\textit{entêtement}) = \textit{têtu}.$$

6.4.7 Reconstitution de fiches lexicographiques

L'analyse des relations lexicales encodées par des FL nous permet, grâce, entre autres, au système de FL, de dégager un certain nombre d'indices signalant la présence de plusieurs lexies distinctes. Rappelons que deux lexies sont en relation de

co-polysémie si elles sont reliées par un élément de sens appelé pont sémantique⁷. Elles figurent alors sous un même vocable polysémique. En revanche, on parle de lexies homonymiques lorsqu'on ne peut trouver d'élément les rapprochant sur le plan sémantique. Les lexies homonymiques appartiennent à des vocables différents.

Voyons à présent la méthodologie que nous allons suivre. À partir de l'analyse d'une relation lexicale, nous reproduisons une fiche sous forme d'une structure de traits⁸. Chaque relation nous permet de recomposer une fiche comprenant quelques éléments. Nous essayons, dans une deuxième étape, de déterminer si, parmi l'ensemble des fiches obtenues, certaines sont liées à la même lexie et devraient, le cas échéant, être fusionnées en une seule et même fiche. Pour cela, nous tentons de fusionner ou d'unifier les différentes fiches en envisageant les critères suivants :

1. la partie du discours ;
2. la marque d'usage ;
3. l'étiquette sémantique.

La partie du discours nous permet d'office de distinguer la présence de vocables différents, par exemple, $MORT_N$ ou $MORT_{Adj}$, dans la mesure où un vocable ne peut contenir deux lexies ne partageant pas la même partie du discours. Les deux autres critères, en revanche, ne permettent pas de déterminer si les lexies sont en relation de co-polysémie ou d'homonymie.

Prenons l'exemple des deux relations suivantes dont *cinoche* est la valeur et à partir desquelles nous allons reconstituer deux fiches :

$Syn(cinéma\#I.3, \text{nom masc, établissement de spectacle}) = \text{"fam"}$

cinoche ;

$Syn(cinéma\#II, \text{nom masc, comportement}) = \text{"fam" } cinoche.$

⁷Un pont sémantique entre les lexies L1 et L2 est la composante non triviale commune à leurs définitions, qui exprime formellement le lien sémantique direct ressenti entre les deux [Mel'čuk et al., 1988 : 28].

⁸Notons qu'il peut s'agir de n'importe quel format de type prédicat-argument. Nous optons ici pour une structure de traits car elle offre une meilleure visualisation des données.

Dans la première relation, la FL Syn nous permet d’inférer que la valeur *cinoche* est nominale, qu’elle possède une étiquette sémantique similaire à celle de la lexie CINÉMA#I.3 (soit, **établissement de spectacle**) et qu’elle est reliée à cette dernière par la FL Syn. Par ailleurs, on peut également déterminer que la lexie est d’usage familier.

Dans la seconde relation, nous pouvons également inférer, par le biais de la FL Syn, que la valeur *cinoche* est nominale, qu’elle possède une étiquette sémantique similaire à celle de CINÉMA#II, c’est-à-dire **comportement**, qu’elle est reliée à cette lexie par la FL Syn. Nous pouvons également déterminer la marque d’usage familier. Avec ces données, il nous est donc possible de reconstituer les fiches de deux lexies potentielles CINOCHÉ sous forme de structure de traits comme ci-dessous :

$$\left[\begin{array}{l} \mathbf{nom} : \text{CINOCHÉ} \\ \mathbf{pdd} : \text{nom} \\ \mathbf{usage} : \text{familier} \\ \mathbf{Étiq} : \text{établissement de spectacle} \\ \mathbf{fl} : [\text{Syn} : \langle \text{cinéma}\#\text{I.3} \rangle] \end{array} \right] \left[\begin{array}{l} \mathbf{nom} : \text{CINOCHÉ} \\ \mathbf{pdd} : \text{nom} \\ \mathbf{usage} : \text{familier} \\ \mathbf{Étiq} : \text{comportement} \\ \mathbf{fl} : [\text{Syn} : \langle \text{cinéma}\#\text{II} \rangle] \end{array} \right]$$

On observe que les champs **pdd** et **usage** sont équivalents dans les deux fiches. Toutefois, les champs **Étiq** sont différents. Il n’est donc pas possible d’unifier les deux structures de traits. On peut alors postuler l’existence de deux lexies : CINOCHÉ_{*i*} et CINOCHÉ_{*ii*}. Notons que l’identification des lexies est tout à fait arbitraire et ne sert qu’à les différencier : *i* ne prévaut en aucune manière sur *ii*.

Poursuivons cette méthodologie avec la valeur *cuisiner* présente dans les deux relations suivantes :

$$\begin{aligned} V_0(\textit{interrogatoire nom, masc fait de questionner}) &= \textit{“fam” cuisiner} \\ &[\text{N=Y}]; \\ \text{Real}_@(\textit{plaque}\#\text{III nom, fém accessoire de cuisson}) &= \textit{cuisiner} [\text{sur} \\ &\text{ART } \sim]. \end{aligned}$$

L'analyse des FL et des autres éléments nous permet de reconstituer les deux fiches suivantes :

$\left[\begin{array}{l} \mathbf{nom} : \text{CUISINER} \\ \mathbf{pdd} : \text{verbe} \\ \mathbf{usage} : \text{familier} \\ \mathbf{\acute{E}tiq} : V_0(\text{fait de questionner}) \\ \mathbf{fl} : [S_0 : <\textit{interrogatoire}>] \end{array} \right]$	$\left[\begin{array}{l} \mathbf{nom} : \text{CUISINER} \\ \mathbf{pdd} : \text{verbe} \\ \mathbf{usage} : \text{non marqué} \\ \mathbf{\acute{E}tiq} : - \\ \mathbf{fl} : - \end{array} \right]$
---	---

Là encore, les fiches ne s'unifient pas car les champs **usage** sont différents. On peut donc postuler que l'on est en présence de deux lexies CUISINER_i et CUISINER_{ii}.

Nous prenons un dernier exemple avec la valeur *crever*, présente dans les quinze relations suivantes :

- (132) FinFunc₀(*abcès*#I nom, masc affection et excroissance cutanée) = *crever* ;
- (133) FinFunc₀(*abcès*#II nom, masc situation conflictuelle) = *crever* ;
- (134) FinFact₀-paroi(*pneu* nom, masc accessoire pour véhicules) = *crever* ;
- (135) Liqu_{1,2}Func₀(*abcès*#II nom, masc situation conflictuelle) = *crever* [l'~];
- (136) LiquFunc₀(*abcès*#I nom, masc affection et excroissance cutanée) = *crever* [ART ~];
- (137) LiquFact₀(*canalisation* nom, fém artefact servant de conduit) = *crever* [ART ~];
- (138) LiquFact₀-paroi(*pneu* nom, masc accessoire pour véhicules) = *crever* [ART ~];
- (139) LiquFact₀(*tuyau*#I.1 nom, masc artefact servant de conduit) = *crever* [ART ~];
- (140) Oper₁(*faim*#I.2 nom, fém, pas de pl, déf situation = "fam" *crever* [de ~];
- (141) Magn.Oper₁(*dépit* nom, masc sentiment négatif) = "fam" *crever* [de ~];
- (142) Magn.Oper₁(*effroi* nom, masc émotion négative) = "fam" *crever* [de ~];
- (143) Magn.Oper₁(*faim*#I.1a nom, fém, pas de pl sensation) = "fam" *crever* [de ~];

- (144) Magn.Oper₁(honte# I.1 nom, fém sentiment négatif) = “fam” *crever* [de ~];
- (145) QSyn(*mourir*#I verbe cesser de vivre) = “vulg” *crever* ;
- (146) QSyn(*mourir*#III verbe éprouver un sentiment) = “fam” *crever*.

Pour chacune de ces relations, on reproduit une structure de traits. Parmi les quinze relations ci-dessus certaines produisent des structures de traits identiques : (132), (133), (134), (135), (136), (137), (138) et (139) ; (140), (141), (142), (143) et (144). Nous ne reproduisons donc dans la figure 6.4 que les structures différentes, soit quatre au total.

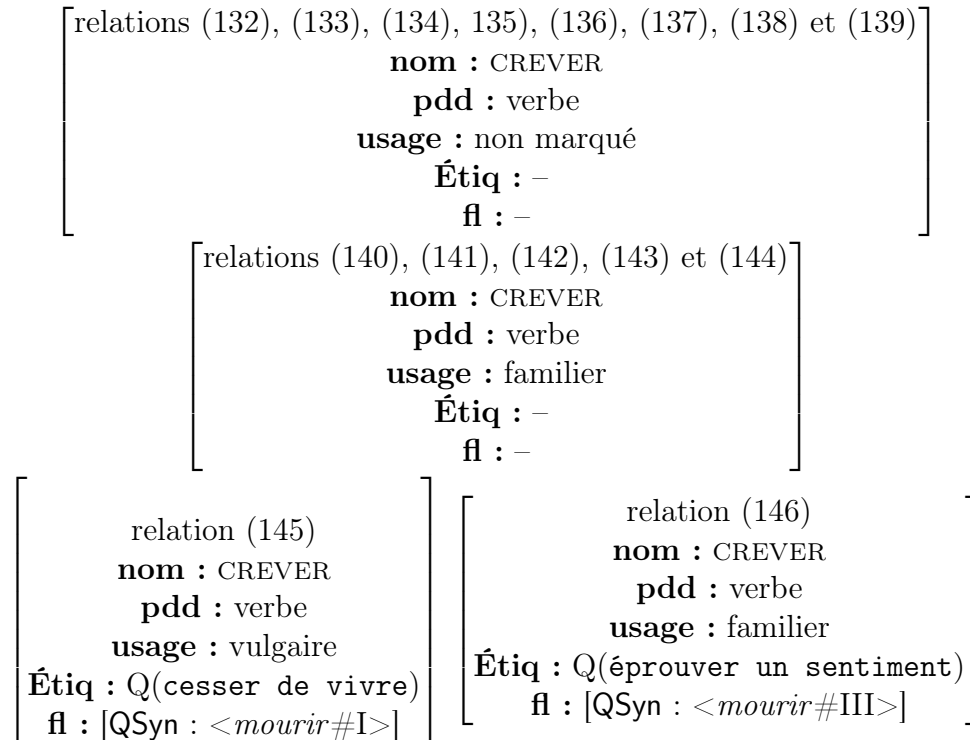


Figure 6.4 – Structures de traits constituées à partir des relations dont CREVER est la valeur

Si deux structures de traits sont identiques, rien n’indique pour autant qu’elles ne représentent qu’une seule lexie. Il est tout à fait possible d’avoir affaire à deux lexies possédant des caractéristiques communes. Ainsi les lexies CREVER transitif

(par exemple : *On a crevé ce pneu.*) et CREVER intransitif (par exemple : *Ce pneu a crevé.*) ne sont pas différenciées et sont représentées par la même structure. À cette étape-ci du travail de description, les structures de traits identiques sont unifiées en une seule et les erreurs potentiellement occasionnées seront corrigées par un lexicographe au moment de la validation des fiches.

On peut maintenant unifier la structure de la relation (146) avec celles des relations (140), (141), (142), (143) et (144). En effet, les champs **pdd** et **usage** sont similaires. Le contenu des champs **Étiq** et **fl** de la première structure figurera dans la structure finale unifiée. Les quinze structures de traits initiales ont été réduites à trois. L'ensemble des lexies CREVER n'est pas entièrement reconstitué mais l'on distingue au moins clairement trois lexies potentielles distinctes représentées dans la figure 6.5.

	$\left[\begin{array}{l} \text{CREVER}_i \\ \mathbf{pdd} : \text{verbe} \\ \mathbf{usage} : \text{non marqué} \\ \mathbf{Étiq} : - \\ \mathbf{FL} : - \end{array} \right]$	
$\left[\begin{array}{l} \text{CREVER}_{ii} \\ \mathbf{pdd} : \text{verbe} \\ \mathbf{usage} : \text{familier} \\ \mathbf{Étiq} : \text{éprouver un sentiment} \\ \mathbf{FL} : \text{QSyn} : \text{mourir}\#\text{III} \end{array} \right]$		$\left[\begin{array}{l} \text{CREVER}_{iii} \\ \mathbf{pdd} : \text{verbe} \\ \mathbf{usage} : \text{vulgaire} \\ \mathbf{Étiq} : \text{cesser de vivre} \\ \mathbf{FL} : \text{QSyn} : \text{mourir}\#\text{I} \end{array} \right]$

Figure 6.5 – Fiches de trois lexies potentielles CREVER

Comme nous le voyons à travers ces exemples, le calcul que nous venons d'effectuer ne permet pas de reconstituer de façon exhaustive les fiches de lexies du DiCo. Il n'a pas non plus la prétention de reconstituer la polysémie d'un vocable. Il permet toutefois de faciliter de façon non négligeable le travail du lexicographe et de récupérer les données internes du DiCo jusqu'alors non exploitées.

6.5 Conclusion

Nous avons présenté, dans ce chapitre, l'implémentation du système de FL sous la forme d'une table de données. En utilisant ce format et les potentialités de requêtes qu'il sous-entend, nous en avons dérivé un assistant à l'encodage des liens lexicaux. Cet assistant consiste à poser à un utilisateur un certain nombre de questions destinées à le guider dans la sélection des différents paramètres descriptifs qui caractérisent un lien lexical. Une fois ces paramètres identifiés, il suffit d'interroger la base de données sous forme d'une requête SQL et de recueillir une ou plusieurs FL candidates. Nous développerons, dans nos futurs travaux, une interface plus conviviale pour cet assistant.

Le système de FL a également été exploité dans le but d'élaborer une méthode endogène pour l'enrichissement du DiCo (ou autre BLEC). Cette méthode permet, en même temps que de compléter automatiquement certains champs du DiCo, de mettre en évidence les erreurs et d'en proposer un diagnostic. Cette application a également l'avantage de rendre cohérent et solidaire l'ensemble des informations contenues dans le DiCo. Les heuristiques présentées pourront facilement être implémentées pour constituer un programme informatique et mener l'expérience sur l'ensemble des données du DiCo, et ce, à chacune de ses modifications.

Chapitre 7

Conclusion générale

7.1 Bilan

Au cours de cette thèse, nous avons réalisé un système d'organisation des relations lexicales fondé sur le formalisme des fonctions lexicales de la Théorie Sens-Texte. Les relations lexicales et les liens lexicaux qu'elles permettent de regrouper constituent une part très importante du lexique des langues. Bien que cet aspect des langues soit encore très peu étudié en comparaison avec la grammaire, la modélisation des liens lexicaux est pourtant d'une importance primordiale pour la lexicologie et, plus largement, pour la modélisation des langues naturelles.

Plusieurs volets ont été traités dans la thèse. Premièrement, nous avons effectué un travail sur les FLNS, qui se situent en marge du système que forment les FL standard. Nous avons discuté de leur statut en tant que FL et avons proposé deux méthodologies pour le repérage des analogies entre liens lexicaux et la régularisation de leur encodage.

- La première méthodologie s'appuie sur les liens entre la dérivation morphologique et les FL pour repérer et même, dans certains cas, prévoir certaines régularités. Par exemple, les suffixes *-logie* et *-tique* ont permis de mettre en évidence la relation 'étude de \sim ', qui s'avère relativement récurrente dans le lexique, et d'homogénéiser sa description.
- La seconde méthodologie consiste à dégager l'attribut d'une lexie ou, en d'autres termes, la composante sémantique de sa définition concernée par un collocatif donné. Par exemple, pour la lexie CHAPEAU, les FLNS en (147a), (147b), (147c) concernent respectivement les attributs ou composantes sémantiques 'fonction', 'position' et 'forme' que l'on peut rattacher à la définition de la lexie.

- (147) a. Équipé pour que les passagers y dorment = \sim -*couchette(s)*, \sim -*lit(s)* ;
 b. Qui est le premier dans un ensemble de W. = *premier, de tête* ;
 c. Qui comporte deux niveaux pour les passagers = *à impériale*.

Le repérage de ces attributs permet de normaliser l'encodage des FLNS selon le modèle suivant : on identifie l'attribut concerné puis on lui attribue une valeur. Les relations précédentes peuvent maintenant être écrites à l'aide des formules en (147a'), (147b'), (147c').

- (147') a. Fonction : dormir = \sim -*couchette(s)*, \sim -*lit(s)* ;
 b. Position : premier = *premier, de tête* ;
 c. Forme : à deux niveaux = *à impériale*.

La révision de l'encodage des FLNS nominales et adjectivales nous a amenée à réduire l'ensemble de FLNS en regroupant des liens lexicaux qui étaient auparavant encodés sous des formules distinctes. Ce nouvel encodage se révèle davantage compositionnel et permet une meilleure recherche dans le DiCo (ou les BLEC). L'encodage proposé pour les FLNS adjectivales permet même d'envisager leur utilisation dans une application de TAL (comme la traduction automatique ou la génération de textes).

Le second volet traité dans la thèse est la structuration de l'ensemble des relations lexicales dans une classification multidimensionnelle, permettant d'atteindre une relation par des requêtes de nature très diverses. Le modèle d'organisation que nous avons proposé est une représentation des FL sous forme d'un système mettant en avant leurs différences et similitudes de façon différentielle. Nous avons dégagé quatre dimensions d'organisation du système : la première organise les FL en classes sémantiques, la seconde les regroupe en fonction des propriétés combinatoires qu'elles partagent, la troisième a trait aux parties du discours des lexies en jeu dans une relation et enfin, la dernière organise les FL en fonction des cibles communicationnelles visées par les FL.

L'implémentation du système de FL a été réalisée sous la forme d'une table de données MySQL, qui s'intègre à la base de données générale du DiCo (cf. chapitre

6). Cette implémentation nous permet d'exploiter le système pour réaliser deux applications. Premièrement, nous avons créé un assistant à l'encodage des liens lexicaux, qui, à partir de la sélection de paramètres descriptifs d'un lien lexical, identifie le ou les FL les plus adaptées pour le représenter. Deuxièmement, nous avons présenté une méthode servant à récupérer et mettre à profit les informations lexicales contenues dans une BLEC pour proposer une description automatique de certains champs de fiches lexicographiques.

Plus de 2 200 FL (tous statuts confondus) recensées dans le DiCo ont été décrites en fonction des différentes perspectives du système, ce qui en fait une solide base de référence pour l'encodage des liens lexicaux.

Le système ainsi constitué a une utilité fondamentale :

- pour la compréhension de la structure du système des FL ;
- pour la maintenance de la base de données lexicale à laquelle il s'applique, puisqu'il permet la recherche d'erreurs et d'incohérences d'encodage ;
- pour le développement futur de la base de données lexicale à travers l'automatisation de certaines descriptions et la consultation d'un outil d'aide à l'encodage des liens lexicaux ;
- pour la popularisation de l'accès au BLEC et leur reconnaissance en tant qu'outil d'enseignement du lexique.

Enfin, le troisième volet traité dans la thèse porte sur le traitement de la granularité dans la description des liens lexicaux. Nous pensons qu'il est important de décrire et de regrouper les liens lexicaux d'une langue selon différents niveaux de granularité. Si l'on se place au niveau de granularité des FLNS, beaucoup de liens lexicaux sont décrits de manière isolée. En revanche, si l'on relâche le niveau de granularité, on voit apparaître des similitudes entre des liens lexicaux qu'il devient possible de représenter sous un même encodage. Notre système de FL permet de contourner les obstacles imposés par l'encodage traditionnel et ses degrés de granularité figés, et d'accéder à tous les liens lexicaux en fonction de certains paramètres descriptifs, indépendamment de leur description formelle dans l'encodage traditionnel.

7.2 Perspectives futures

Nous envisageons d’explorer deux voies complémentaires dans nos recherches futures. La première concerne le développement de l’encodage des FL, la seconde a trait au développement logiciel des outils de consultation et d’automatisation du DiCo, et, plus largement, de toute BLEC.

Nous avons mentionné à plusieurs reprises, au cours de cette thèse, la nécessité de créer une grammaire de formation des FL semi-standard. Rappelons que celles-ci sont formées d’une formule de FL standard et d’une notation venant préciser une composante sémantique additionnelle (cf. exemples ci-dessous) :

(148) **pour frapper qqn.** $\text{Labreal}_{1@}(\text{ceinture}\#\text{I.1}) = \text{battre}$ [N avec ART \sim];

(149) **temporairement** $\text{LiquFunc}_0(\text{conversation}\#\text{a}) = \text{interrompre}$ [ART \sim].

Les FL semi-standard forment un ensemble très hétérogène : les notations utilisées ne sont pas rédigées selon un patron défini et ne respectent aucune norme d’écriture. Notre objectif est d’établir un patron d’encodage qui encode deux types d’information : d’une part, l’identification du type de composante sur laquelle on apporte une précision (par exemple, la manière, la finalité, la conséquence, la durée, les participants, etc.), d’autre part, l’information en tant que telle, le tout respectant une syntaxe prédéfinie. Nous avons déjà proposé un tel modèle pour les FLNS adjectivales. Si l’on suit ce modèle, les FL en (148) et (149) pourraient être écrites comme en (148’) et (149’) :

(148’) **Finalité : frapper qqn.** $_ \text{Labreal}_{1@}(\text{ceinture}\#\text{I.1}) = \text{battre}$ [N avec ART \sim];

(149’) **Durée : temporairement** $_ \text{LiquFunc}_0(\text{conversation}\#\text{a}) = \text{interrompre}$ [ART \sim].

Ce travail de normalisation permettrait de mieux gérer et organiser l’ensemble des FL semi-standard ainsi que de mettre en valeur le lien qui les unit aux définitions des lexies auxquelles elles s’appliquent.

Pour ce qui est des développements logiciels, nous souhaitons mettre l’accent sur l’avancement et la popularisation du DiCo. Nous pensons qu’il est indispensable que le DiCo intègre un processus d’automatisation s’il veut être reconnu comme une ressource lexicographique de référence et non comme une simple illustration du

modèle théorique de la LEC. Nous envisageons la réalisation de trois projets pour amorcer l’automatisation et assister les lexicographes dans la rédaction d’articles de DiCo. Rappelons une fois de plus que ces projets peuvent aussi bien s’appliquer à n’importe quelle BLEC.

- Le premier projet consiste à poursuivre la méthode endogène d’enrichissement du DiCo. D’autres pistes restent à explorer concernant, notamment, la détermination du nombre d’actants d’une lexie, la transmission des étiquettes sémantiques et des typages d’actants par le biais des FL ainsi que les héritages de collocations parmi les lexies d’un même champ sémantique. Cette dernière question a été traitée par Mel’čuk et Wanner [1996] à partir d’observations faites sur le champ sémantique des émotions en allemand. Par exemple, les noms d’émotions sont susceptibles d’être associés à la FL *Oper*₁ et d’avoir la valeur *éprouver*. Barrios Rodríguez [2008] a approfondi la question de l’héritage lexical en proposant une génération de valeurs par analogie et en définissant des domaines pour chaque type de FL. À partir de la définition de traits sémantiques liés aux unités lexicales constituant le domaine d’une FL, il est possible de prévoir, d’une part, qu’une lexie relève du domaine d’une FL donnée, d’autre part, la ou les valeurs qui lui seront associées si on lui applique la FL en question. Nous envisageons d’appliquer une telle méthode aux fiches développées pour le français dans le DiCo en menant l’expérience sur des champs sémantiques bien définis. Ceci nous permettra de prédire à la fois des FL et des valeurs typiquement associées à des lexies de champs sémantiques particuliers.
- Nous aimerions, par ailleurs, exploiter des ressources exogènes pour récupérer et identifier un grand nombre de collocations et de dérivés sémantiques. Tout comme le montrent les travaux de Wanner (entre autre Wanner et al. [2006]), il est possible, en extrayant des collocations à partir de corpus, ou en utilisant celles déjà identifiées dans des dictionnaires de collocations, d’émettre des hypothèses sur les FL adéquates pour les représenter. La plupart des outils d’extraction de collocations comme Les Voisins de Le Monde [Bourigault et Galy, 2005] identi-

fient les patrons syntaxiques d'une lexie et de ses collocatifs, ce qui réduit déjà considérablement les FL potentielles pour une relation.

- Nous souhaitons également perfectionner l'assistant à l'encodage de FL présenté au sixième chapitre afin d'en faire une application plus interactive que les simples requêtes au format SQL. Rappelons que cet assistant a pour objectif, à partir de la sélection de plusieurs paramètres décrivant une relation lexicale, de déterminer la ou les FL les mieux adaptées pour l'encoder.

Parallèlement à ces travaux pour l'aide à la rédaction de BLEC, nous envisageons la création d'interfaces de consultation. Nous avons déjà évoqué ce travail au cours du cinquième chapitre de la thèse ainsi que dans Jousse et al. [2006] et Jousse [2007b]. Il s'agit d'offrir aux utilisateurs de BLEC plusieurs types d'accès aux relations lexicales proposées pour une unité lexicale donnée. À partir du système de FL et de ses différentes perspectives d'organisation, il est possible de développer directement quatre interfaces proposant :

1. un accès onomasiologique permettant de trouver le bon collocatif ou dérivé sémantique à partir de son sens ;
2. un accès syntaxique, organisant les collocatifs et dérivés sémantiques en fonction de leur configuration syntaxique avec la lexie vedette ;
3. un accès portant sur les parties du discours des lexies reliées à une lexie vedette ;
4. un accès portant sur la cible communicationnelle visée par la relation. Par exemple, pour la lexie 「COUP DE FIL」, le collocatif verbal *passer* met l'accent sur le premier actant de 「COUP DE FIL」 alors que *recevoir* met l'accent sur son deuxième actant.

Nous espérons que notre travail constituera un moteur pour le développement de nouvelles structurations parmi les ressources lexicales élaborées dans le cadre de la LEC, et qu'il permettra le développement d'applications exploitant toutes les potentialités que ce modèle théorique peut offrir en matière d'apprentissage du lexique.

Bibliographie

- Jean Aitchison. *Words in the Mind : An Introduction to the Mental Lexicon*. Basil Blackwell, Oxford, 3^e édition, 2003.
- Margarita Alonso Ramos. *Las funcionales léxicas en el modelo lexicográfico de I. Mel'čuk*. Thèse de doctorat, Universidad nacional de educacion a distancia, Madrid, 1993.
- Margarita Alonso Ramos. Critères heuristiques pour l'encodage des collocations au moyen des fonctions lexicales. Dans *Actes d'Euralex 2000*, volume 2, pages 222–232, Université de Stuttgart, 2000.
- Margarita Alonso Ramos. Elaboración del Diccionario de colocaciones en español y sus aplicaciones. Dans Paz Bataner et Janet de Cesaris, dir., *De Lexicographia. Actas del I Simposio internacional de Lexicografía*, pages 149–162, Barcelona, 2004. IULA-Edicions Petició.
- Margarita Alonso Ramos. Glosas para un Diccionario de colocaciones del español. Dans Margarita Alonso Ramos, dir., *Diccionarios y Fraseología, Anejo de la Revista de Lexicografía*, pages 59–88. Servicio de Publicaciones de la Universidad de La Coruña, 2006.
- Margarita Alonso Ramos. Synthesis of Support Verb Constructions. Dans Leo Wanner, dir., *Selected Lexical and Grammatical Issues in the Meaning-Text Theory*, pages 97–137. John Benjamins, Amsterdam/Philadelphie, 2007.
- Margarita Alonso Ramos, Owen Rambow et Leo Wanner. Using Semantically Annotated Corpora to Build Collocation Resources. Dans *Proceedings of the LREC 2008*, Marrakesh, 2008.
- Margarita Alonso Ramos et Agnès Tutin. A Classification and Description of Lexical Functions for the Analysis of their Combinations. Dans Leo Wanner, dir., *Lexical Functions in Lexicography and Natural Language Processing*. John Benjamins, Amsterdam/Philadelphia, 1996.

- Joel Altman et Alain Polguère. La BDéf : base de définitions dérivée du Dictionnaire Explicatif et Combinatoire. Dans *Actes de la première conférence internationale de Théorie Sens-Texte (MTT2003)*, pages 43–54, Paris, Juin 2003.
- Juri Apresjan. *Systematic lexicography*. Oxford University Press, Oxford, 2000.
- Juri Apresjan, Igor Boguslavsky, Leonid Iomdin et Leonid Tsinman. Lexical Functions as a Tool of Etap-3 (première version du texte). Dans *Proceedings of MTT 2003*, Paris, 2003.
- Jury Apresjan, Igor Boguslavsky, Leonid Iomdin et Leonid Tsinman. Lexical functions in NLP : Possible Uses. Dans *Computational Linguistics for the New Millennium : Divergence or Synergy*, pages 1–11, Heidelberg, 2000.
- Beth T. Atkins et Alain Duval. *Robert & Collins Dictionnaire Français Anglais, Anglais Français*. Paris : Le Robert/Glasgow : Collins, 1978.
- Lucie Barque. FrameNet et le traitement de la polysémie. Journée d'étude de l'ATALA, 13 mai 2006.
- María Auxiliadora Barrios Rodríguez. *El dominio de la funciones léxicas en el marco de la teoría sentido-texto*. Thèse de doctorat, Universidad Complutense de Madrid, 2008.
- Pascale Bélanger. Exploration des liens de condensation pour le résumé des textes grâce à l'application des formalismes de la théorie Sens-Texte. Mémoire de maîtrise, Département de linguistique et traduction, Université de Montréal, 2003.
- Pascale Bélanger et Richard Kittredge. Paraphrasing with Space Constraints : Linguistic Operations in Journal Abstracting. Dans *Proceedings of MTT 2005*, Moscou, 2005.
- Morton Benson, Evelyn Benson et Robert Ilson. *The BBI Dictionary of English Word Combinations*. John Benjamins, Amsterdam/Philadelphia, 1997.

- Jean Binon, Serge Verlinde, Jan van Dick et Ann Bertels. *Dictionnaire d'Apprentissage du Français des Affaires. Dictionnaire de compréhension et de production de la langue des affaires*. Didier, Paris, 2000.
- Igor Boguslavsky, Maria Auxialora Barrios Rodriguez et Pavel Diachenko. Callex-esp : a Software System for Learning Spanish Lexicon and Collocations. *Current Developments in Technology-Assisted Education*, pages 22–26, 2006.
- Bernd Bohnet, Andreas Langjahr et Leo Wanner. A Development Environment for an MTT-based sentence generator. Dans *Proceedings of the First International Conference on Natural Language Generation*, pages 260–263, Mitzpe Ramon, Israel, 2000.
- Henri Bonnard. *Code Du Français Courant 2^{nde} 1^{ere} et Terminale - Grammaire*. Magnard, Paris, 1983.
- Ignacio Bosque. *Redes. Diccionario combinatorio del español contemporáneo : las palabras en su contexto*. Boadilla del Monte, Madrid, 2004.
- Samia Bouchaddakh. La définition dans les dictionnaires bilingues : problèmes de polysémie et d'équivalences interlangues. Dans Elisenda Bernal et Janet de Cesaris, dir., *Actes d'Euralex 2008*, pages 807–812, Barcelone, 2008.
- Didier Bourigault et Edith Galy. Les Voisins de Le Monde : Une base lexicale distributionnelle du français construite à partir d'un gros corpus. Dans *4^e Journées de Linguistique de corpus*, Lorient, septembre 2005.
- Didier Bourigault et Monique Slodzian. Pour une terminologie textuelle. *Terminologies nouvelles*, pages 29–32, 1999.
- Myriam Bouveret. Le terme : une dénomination au sens réglé. Dans *Actes du colloque TIA 97*, pages 115–127, 1997.
- Myriam Bouveret. Fonctions lexicales pour le typage des relations syntagmatiques et paradigmatisques en bioindustries, une approche lexicographique du terme.

Processing of Term in Specialized Dictionaries : New Models and Techniques, special issue of Terminology, 12(2):235–259, 2006.

Myriam Bouveret et Charles J. Fillmore. Matching Verbo-nominal Constructions in FrameNet with Lexical Functions in MTT. Dans Elisenda Bernal et Janet De Cesaris, dir., *Actes d'Euralex 2008*, pages 297–308, Barcelone, 2008.

Simon Charest, Éric Brunelle, Jean Fontaine et Bertrand Pelletier. Élaboration automatique d'un dictionnaire de cooccurrences grand public. Dans *Actes de TALN 2007*, pages 283–292, 2007.

José Coch. Overview of Alethgen. Dans *Proceedings of the Eighth International Workshop on Natural Language Generation (INLG-1996) (Demonstrations and Posters)*, pages 25–28, 1996.

Danielle Corbin. *Morphologie dérivationnelle et structuration du lexique*. Niemeyer, Tübingen, 1987.

David A. Cruse. *Lexical Semantics*. Cambridge University Press, Cambridge (UK), 1986.

Jeanne Dancette et Christophe Réthoré. *Dictionnaire analytique de la distribution*. Presses universitaires de Montréal, 2000.

Laurence Danlos et Benoît Sagot. Constructions pronominales dans Dicovalence et le lexique-grammaire — Intégration dans le Lefff. *Linguisticæ Investigationes*, 2(32):293–304, 2009.

Gilles-Maurice De Schryver. Lexicographers' Dream in the Electronic-Dictionary Age. *International Journal of Lexicography*, 16(2):143–199, 2003.

Pavel Diachenko. Lexical Functions in Learning the Lexicon, Current Developments. *Technology-Assisted Education*, pages 538–542, 2006.

Jean Dubois et Françoise Dubois Charlier. *La dérivation suffixale en français*. Nathan, Paris, 1999.

- Karine Duvignau. *La métaphore, berceau et enfance de la langue. La métaphore verbale comme approximation sémantique par analogie dans les textes scientifiques et les productions enfantines (2-4 ans)*. Thèse de doctorat, Université Toulouse 2, 2002.
- K. O. Erastov. Primery slovarnyx opisaniy [Some sample lexicographic descriptions]. *Mavsinnyj perevod i prikladnaja lingvistika [Automatic Translation and Applied Linguistics]*, 11:36–59, 1969.
- Martha Evens, James Vandendorpe et Yih-Chen Wang. Lexical-Semantic Relations in Information Retrieval. Dans Stephanie Williams, dir., *Humans and Machines*, pages 73–100. Ablex, Norwood, New Jersey, 1985.
- Martha W. Evens, Bonnie E. Litowitz, Judith A. Markowitz, Raoul N. Smith et Oswald Werner. *Lexical-semantic relations : a comparative survey*. Linguistic Research, Inc., Edmonton, Alberta, 1980.
- Pamela Faber et Maribel Tercedor Sánchez. Codifying conceptual information in descriptive terminology management. *Meta*, XLVI, 1:192–204, 2001.
- Christiane Fellbaum, dir. *WordNet : an Electronic Lexical Database*. MIT Press, Cambridge, 1998.
- Christiane Fellbaum, Grabowski et Shari Landes. Performance and Confidence in a Semantic Annotation Task. Dans Christine Fellbaum, dir., *WordNet : an Electronic Database*, chapitre 9, pages 217–238. MIT Press, 1998.
- Charles J. Fillmore. Scenes-and-frames Semantics. *Linguistic Structure Processing*, pages 55–82, 1977.
- Charles J. Fillmore. Frame semantics. *Linguistics in the Morning Calm*, pages 111–137, 1982.
- Charles J. Fillmore, Christopher R. Johnson et Miriam R.L. Petruck. Background to FrameNet. *International Journal of Lexicography — Special Issue on FrameNet*, 16 (3):235–250, 2003.

- Thierry Fontenelle. Dictionnaires électroniques et relations lexicales : une comparaison entre quelques programmes européens. *Revue Française de Linguistique Appliquée*, II-1:65–77, 1997a.
- Thierry Fontenelle. *Turning a Bilingual Dictionary into a Lexical Semantic Database*. Max Niemeyer, Tübingen, 1997b.
- William Frawley. News forms of Specialized Dictionaries. *International Journal of Lexicography*, 1(3):89–213, 1998.
- François Gaudin. *Socioterminologie. Une approche sociolinguistique de la terminologie*. Champs linguistiques. Duculot, Louvain-la-Neuve, 2003.
- Bruno Gaume et Karine Duvignau. Pour une ergonomie cognitive des dictionnaires électroniques. *DN - Fouille de textes*, 8:157–181, 2004.
- Eli Goldberg, Norbert Driedger et Richard Kittredge. Using Natural Language Processing to Produce Weather Forecasts. Dans *IEEE Expert*, 9(2), pages 45–53, 1994.
- Joseph Grimes. Inverse Lexical Functions. Dans James Steele, dir., *Meaning Text Theory : Linguistics, Lexicography and Implications*, pages 350–364. Ottawa University Press, Ottawa, 1990.
- Joseph Grimes. Lexical functions across languages. Dans Leo Wanner, dir., *Lexical functions in lexicography and natural language processing*, Studies in Language Companion, pages 103–13. Benjamins, Amsterdam, 1996.
- Bénédicte Grizolle. Classification des fonctions lexicales non standard du Dico. Rapport de stage - OLST, Université de Montréal, 2003.
- Maurice Gross. *Méthodes en syntaxe. Le régime des constructions complétives*. Hermann, Paris, 1975.
- Yvan Gut, Puteri Rashida Megat Ramli, Zaharin Yusoff, Chuah Choy Kim, Salina A. Samat, Christian Boitet, Nicolas Nédobejkine et Mathieu Lafourcade.

- Kamus Perancis-Melayu Dewan, dictionnaire Français-Malais.* Dewan Bahasa Dan Pustaka, Kuala Lumpur, 1996.
- Zellig S. Harris. The Elementary transformations. Dans *Transformations and Discourse Analysis Papers 54*, Université de Pennsylvanie, Philadelphie, 1964.
- Franz Joseph Hausmann. Tout est idiomatique dans les langues. Dans M. Martins-Baltar, dir., *La Locution, entre langue et usages*, pages 277–290. ENS Éditions, Fontenay-aux-Roses, 1997.
- Gabriel Antoine Joseph Hécart. *Dictionnaire rouchi-français.* Lemaître, Valenciennes, 3^e édition, 1834.
- Douglas J. Herrmann, Roger Chaffin, M. P. Daniel et R. S. Wool. The Role of Elements of Relation Definition in Antonym and Synonym Comprehension. *Zeitschrift für Psychologie*, 194-2:133–153, 1986.
- Lidja Iordanskaja, Myunghee Kim, Richard Kittredge, Benoit Lavoie et Alain Polguère. Generation of Extended Bilingual Statistical Reports. Dans *Proceedings of the 14th International Conference on Computational Linguistics (COLING-1992)*, vol. 3, pages 1019–1023, 1992.
- Lidja Iordanskaja, Richard Kittredge et Alain Polguère. Implementing a meaning-text model for language generation. Dans *Proceedings of the 12th International Conference on Computational Linguistics (COLING-88)*, pages 22–27, Budapest, 1988.
- Lidja Iordanskaja et Alain Polguère. Hooking up Syntagmatic Lexical Functions to Lexicographic Definitions. Dans *Proceedings of the Second International Conference on the Meaning Text Theory*, pages 176–186, Moscou, 2005.
- Anne-Laure Jousse. Dérivation morphologique de termes, analyse en corpus spécialisé et modélisation au moyen des fonctions lexicales. Mémoire de maîtrise, Université du Maine, Le Mans, 2002.

- Anne-Laure Jousse. Normalisation des fonctions lexicales non standard. Mémoire de maîtrise, Université Paris 7- Denis Diderot, 2003.
- Anne-Laure Jousse. Extension de l'encodage formel des fonctions lexicales dans le cadre de la lexicologie explicative et combinatoire. Dans *Actes de TALN 2007*, pages 469–478, Toulouse, 5-8 juin 2007a.
- Anne-Laure Jousse. Organisation des fonctions lexicales pour un meilleur accès à l'information dans le DiCo. Dans Mathieu Loiseau et al., dir., *Autour des langues et du langage : perspective pluridisciplinaire. Papiers sélectionnés du Colloque International des Étudiants Chercheurs en Didactique des Langues et en Linguistique, Grenoble, juillet 2006*. P.U.G., 2007b.
- Anne-Laure Jousse et Myriam Bouveret. Lexical Functions to Represent Derivational Relations in Specialized Dictionaries. *Terminology*, 9(1):71–98, 2003.
- Anne-Laure Jousse et Alain Polguère. Le DiCo et sa version DiCouèbe. Document descriptif et manuel d'utilisation - OLST, Université de Montréal, Janvier 2005.
- Anne-Laure Jousse, Alain Polguère et Ophélie Tremblay. Du dictionnaire au site lexical pour l'enseignement/apprentissage du vocabulaire. Dans Francis Grossmann et Sylvie Plane, dir., *Lexique et production verbale : vers une meilleure intégration des apprentissages lexicaux*, Éducation et didactiques. Presses Universitaires du Septentrion, Villeneuve d'Ascq, 2006.
- Anne-Laure Jousse et Ophélie Tremblay. Le dictionnaire électronique, un outil précieux pour enseigner le lexique. *Québec Français*, 141, 2006.
- Sylvain Kahane. The Meaning Text Theory. Dans *Dependency and Valency, Handbooks of Linguistics and Communication Sciences*, volume 25 : 1-2, pages 32. De Gruyter, Berlin/NY, 2003a.
- Sylvain Kahane. Une 'blessure profonde' dans le dictionnaire explicatif et combinatoire : Sur le lien entre la définition lexicographique et les fonctions lexicales. Dans Francis Grossmann et Agnès Tutin, dir., *Les collocations : analyse et traitement*,

Travaux et recherches en linguistique appliquée, Revue Française de Linguistique Appliquée, pages 61–73, Amsterdam, 2003b. De Werelt.

Sylvain Kahane. Les unités minimales de la syntaxe et de la sémantique : le cas du français. Dans *1^{er} Congrès Mondial de Linguistique Française*, Paris, 2008.

Sylvain Kahane et Alain Polguère. Formal foundation of lexical functions. Dans *Workshop proceedings : COLLOCATION : Computational Extraction, Analysis and Exploitation*, pages 8–15, Toulouse, Juillet 2001. 39th Annual Meeting and 10th Conference of the European Chapter of the Association for Computational Linguistics.

Karin Kipper, Anna Korhonen, Neville Ryant et Martha Palmer. Extensive classifications of english verbs. Dans *Proceedings of the 12th EURALEX International Congress*, Turin, Italie, Septembre 2006.

Annette Klosa, Ulrich Schnörch et Petra Storjohann. Elexiko — a lexical and lexicological, corpus-based hypertext information system at the institut für deutsche sprachemannheim. Dans *Proceedings of Euralex 2006*, 2006.

Shari Landes, Claudia Leacock et Randee I. Teng. Building Semantic Concordances. Dans Christiane Fellbaum, dir., *WordNet : an Electronic Database*, chapitre 8, pages 199–216. MIT Press, 1998.

François Lareau. La synthèse automatique de paraphrases comme outil de vérification des dictionnaires et grammaires de type sens-texte. Mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal, 2002.

Benoit Lavoie et Owen Rambow. A fast and portable realizer for text generation systems. Dans *Proceedings of the Fifth Conference on Applied Natural Language Processing (ANLP97)*, pages 265–268, 1997.

Chantal Lemay, Marie-Claude L’Homme et Patrick Drouin. Two Methods for Extracting “Specific” Single-word Terms from Specialized Corpora : Experimenta-

- tion and Evaluation. *International Journal of Corpus Linguistics*, 10(2):227–255, 2005.
- Beth Levin. *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press, Chicago et Londres, 1993.
- Marie-Claude L’Homme. Fonctions lexicales pour représenter les relations sémantiques entre termes. *Traitement Automatique de la Langue*, 43(1):19–41, 2002.
- Marie-Claude L’Homme. A Lexico-semantic Approach to the Structuring of Terminology. Dans *Computerm 2004 (dans le cadre de Coling 2004)*, 2004.
- Marie-Claude L’Homme. Conception d’un dictionnaire fondamental de l’informatique et de l’internet : sélection des entrées. *Le langage et l’homme*, 40(1):137–154, 2005.
- Marie-Claude L’Homme. DicoInfo : le dictionnaire fondamental de l’informatique et de l’internet. OLST - Université de Montréal, janvier 2007.
- Birte Lönneker-Rodman et Collin F. Baker. The FrameNet Model and its Applications. *Natural Language Engineering*, 15(3):415–453, juillet 2009.
- Mathieu Mangeot. Projet Papillon : intégration de dictionnaires existants et gestion des contributions. Dans *Actes de JST 2002*, volume 1/1, pages 64–65, National Olympic Memorial Youth Center, Tokyo, Japon, 17-19 novembre 2002.
- Mathieu Mangeot. Papillon Project : Retrospective and Perspectives. Dans Pierre Zweigenbaum, dir., *Proceedings of Acquiring and Representing Multilingual, Specialized Lexicons : the Case of Biomedicine*, Gênes, Italie, 22 Mai 2006. LREC workshop 2006.
- Mathieu Mangeot et Kyoko Kuroda. Divergences interlinguistiques dans le dictionnaire multilingue Papillon. Dans *Actes de MTT 2003 (First International Conference on Meaning-Text Theory)*, pages 33–42, 16-18 juin 2003.

- Colin McIntosh, Ben Francis et Richard Poole, dir. *Oxford Collocations Dictionary for Students of English*. Oxford University Press, Oxford, 2002.
- Igor Mel'čuk. Lexical Functions in Lexicographic Description. Dans *Proceedings of the Eight Annual Meeting of the Berkeley Linguistics Society*, pages 427–444, Berkeley : Department of Slavic Languages and Literatures, University of California, 1982.
- Igor Mel'čuk. *Cours de morphologie générale. Vol. 2 : Significations morphologiques*. Montréal : Les Presses de l'Université de Montréal; Paris : CNRS Éditions, 1994a.
- Igor Mel'čuk. Fonctions lexicales dans le traitement du langage naturel. Dans André Clas et Pierrette Bouillon, dir., *TA-TAO : Recherches de pointe et applications immédiates*, pages 193–219. AUPELF-UREF/FMA, Beyrouth/Montréal, 1994b.
- Igor Mel'čuk. Lexical functions : A tool for the description of lexical relations in the lexicon. Dans Leo Wanner, dir., *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–102. Benjamins, Amsterdam/Philadelphie, 1996.
- Igor Mel'čuk. Vers une linguistique Sens-Texte. Leçon inaugurale. Collège de France, Paris, 10 janvier 1997.
- Igor Mel'čuk. *Communicative Organization in Natural Language*. John Benjamins, Amsterdam/Philadelphie, 2001.
- Igor Mel'čuk. Parties du discours et locution. *Bulletin de la société linguistique de Paris*, 101(1):29–65, 2006.
- Igor Mel'čuk. Phraséologie dans la langue et dans le dictionnaire. Dans *Repères et Applications (VI), XXIV^e Journées Pédagogiques sur l'Enseignement du Français en Espagne*, 2008.
- Igor Mel'čuk. *Semantics : From Meaning to Text*. John Benjamins, Amsterdam/Philadelphie., à paraître.

- Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Louise Dagenais, Léo Elnitsky, Lidja Iordanskaja, Marie-Noëlle Lefebvre et Suzanne Mantha. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II*. Les Presses de l'Université de Montréal, Montréal, 1988.
- Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Léo Elnitsky, Lidja Iordanskaja et Adèle Lessard. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*. Les Presses de l'Université de Montréal, Montréal, 1984.
- Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Lidja Iordanskaja et Suzanne Mantha. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III*. Les Presses de l'Université de Montréal, Montréal, 1992.
- Igor Mel'čuk, Nadia Arbatchewsky-Jumarie, Lidja Iordanskaja, Suzanne Mantha et Alain Polguère. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques IV*. Les Presses de l'Université de Montréal, Montréal, 1999.
- Igor Mel'čuk, André Clas et Alain Polguère. *Introduction à la Lexicologie Explicative et Combinatoire*. Duculot, Louvain-la-Neuve, 1995.
- Igor Mel'čuk et Alain Polguère. Dérivations sémantiques et collocations dans le Dico/LAF. *Langue française*, 150(2):66–83, juin 2006.
- Igor Mel'čuk et Alain Polguère. *Lexique Actif du Français*. De Boeck/Larcier, Bruxelles, 2007.
- Igor Mel'čuk et Leo Wanner. Lexical Functions and Lexical Inheritance for Emotion Lexemes in German. Dans Leo Wanner, dir., *Lexical Functions in Lexicography and Natural Language Processing*, pages 209–278. John Benjamins, Amsterdam/Philadelphie, 1996.
- Igor Mel'čuk et Aleksandr K. Zholkovsky. Toward a Functioning "Meaning-Text" Model of Language. *Linguistics*, 57:10–47, 1970.

- Igor Mel'čuk et Aleksandr K. Zholkovsky. *The Explanatory Combinatorial Dictionary of Modern Russian*. Wiener Slawistischer Almanach, Viena, 1984.
- Igor Mel'čuk et Aleksandr K. Zholkovsky. The Explanatory Combinatorial Dictionary. Dans Martha W. Evens, dir., *Relational models of the lexicon : representing knowledge in semantic networks*, pages 41–74. Cambridge University Press, Cambridge, 1988.
- Ingrid Meyer et John Steele. The Presentation of an Entry and Super-entry in an Explanatory Combinatorial Dictionary of English. Dans James Steele, dir., *Meaning-Text Theory, Linguistics, Lexicography and Implications*, pages 62–94. University of Ottawa Press, Ottawa, 1990.
- Jasmina Miličević. *Modélisation sémantique, lexicale et syntaxique de la paraphrase langagière*. Thèse de doctorat, Département de linguistique et de traduction, Université de Montréal, 2003.
- Jasmina Miličević. *La paraphrase. Modélisation de la paraphrase langagière*. Peter Lang, Bern, 2007.
- Jasmina Miličević. Schéma de régime : le pont entre le lexique et la grammaire. *Langages*, 4(176):94–116, 2009.
- M. Lynne Murphy. *Semantic Relations and the Lexicon : Antonymy, Synonymy, and Other Paradigms*. Cambridge University Press, Cambridge, 2003.
- Alexis Nasr. *Un système de reformulation automatique de phrases fondé sur la Théorie Sens-Texte : application aux langues contrôlées*. Thèse de doctorat, Université Paris 7, 1996.
- Étienne Van Tien Nguyen. *Unité lexicale et morphologie en chinois mandarin. Vers l'élaboration d'un Dictionnaire Explicatif et Combinatoire du chinois*. Thèse de doctorat, Département de linguistique et de traduction, Université de Montréal, 2006.

- Brigitte Orliac. *Automatisation du repérage et de l'encodage des collocations en langue de spécialité*. Thèse de doctorat, Université de Montréal, 2004.
- Jacqueline Picoche. *Précis de lexicologie française*. Nathan, Paris, 1977.
- Jacqueline Picoche. *Structures sémantiques du lexique français*. Nathan, Paris, 1986.
- Jacqueline Picoche. *Didactique du vocabulaire français*. Nathan, Paris, 1993.
- Jacqueline Picoche et Jean-Claude Rolland. *Dictionnaire du Français Usuel*. De Boeck/Duculot, Bruxelles, 2002.
- Sabine Ploux et Bernard Victorri. Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. *Traitement Automatique des Langues*, 39/1:161–182, 1998.
- Alain Polguère. La Théorie Sens-Texte. *Dialangue*, 8-9:9–30, 1998.
- Alain Polguère. Towards a Theoretically-motivated General Public Dictionary of Semantic Derivations and Collocations for French. Dans *Actes d'Euralex 2000*, pages 517–527, Stuttgart, 2000.
- Alain Polguère. Collocations et fonctions lexicales : pour un modèle d'apprentissage. Dans Francis Grossmann et Agnès Tutin, dir., *Les Collocations. Analyse et traitement*, volume E :1, pages 117–133. De Werelt, Amsterdam, 2003.
- Alain Polguère. Étiquetage sémantique des lexies dans la base de donnée DiCo. *Traitement Automatique des Langues*, 44(2):39–68, 2003.
- Alain Polguère. Structural Properties of Lexical Systems : Monolingual and Multilingual Perspectives. Dans *Proceedings of the Workshop on Multilingual Language Resources and Interoperability (COLING/ACL 2006)*, pages 50–59, Sydney, 2006.
- Alain Polguère. Lessons from the *Lexique Actif du Français*. Dans *Proceedings of the Third International Conference on the Meaning Text Theory (MTT 2007)*, 2007.

- Alain Polguère. Lexical Function Standardness. Dans *Selected Topics in Meaning Text Theory, in honour of Igor Mel'čuk*, pages 43–92. John Benjamins, 2007.
- Alain Polguère. Lexical Systems : Graph Models of Natural Language Lexicons. *Language Resources and Evaluation*, 43(1):41–55, mars 2009. Springer.
- Stefan Popovic. Paraphrasage des liens de fonctions lexicales. Mémoire de maîtrise, Département de linguistique et de traduction, Université de Montréal, 2004.
- Bernard Pottier. *Sémantique générale*. Presses Universitaires de France, Paris, 1992.
- James Pustejovsky. *The Generative Lexicon*. MIT Press, Cambridge (Mass.), 1995.
- François Rastier. Représentation du contenu lexical et formalismes de l'intelligence artificielle. *Langages*, 87:79–102, 1987.
- François Rastier. Ontologie(s). *Revue des sciences et technologies de l'information, Série : Revue d'Intelligence artificielle*, 18, n°1:15–40, 2004.
- Stephen D. Richardson, William B. Dolan et Lucy Vanderwende. MindNet : acquiring and structuring semantic information from text. Dans *Proceedings of the 17th international conference on Computational linguistics - Volume 2*, pages 1098 – 1102, 1998.
- Eleanor H. Rosh. Natural categories. *Cognitive Psychology*, 4:328–350, 1973.
- Benoît Sagot. The Lefff, a freely available, accurate and large-coverage lexicon for French. Dans *Actes de LREC 2010*, La Valette, Malte, 2010.
- Benoît Sagot, Lionel Clement, Eric Villemonte de La Clergerie et Pierre Boullier. The Lefff 2 syntactic lexicon for French : architecture, acquisition, use. Dans *LREC'06*, 2006.
- Benoît Sagot et Darja Fišer. Building a free french wordnet from multilingual resources. Dans *Actes de Ontolex*, Marrakech, Maroc, 2008.

- Thierry Selva, Serge Verlinde et Jean Binon. Le DAFLES, un nouveau dictionnaire électronique pour apprenants du français. Dans *Actes de Euralex 2002*, pages 199–208, Copenhague, 2002.
- Monique Slodzian. WordNet : what about its linguistic relevancy? Dans *Proceedings of the EKAW2002 workshop on « Ontologies and texts »*, Juan-les-Pins (France), Octobre 2000.
- John F. Sowa. Semantic networks. Dans Stuart C. Shapiro, dir., *Encyclopedia of Artificial Intelligence*. Wiley, New York, 2^e édition, 1987.
- James Steele. A Lexical Entry for an Explanatory Combinatorial Dictionary of English (hope II.1). *Dictionaries*, 8:1–54, 1986.
- James Steele et Ingrid Meyer. Lexical Functions in an Explanatory and Combinatorial Dictionary : Kinds, Descriptions, and English Examples. Dans James Steele, dir., *Meaning Text Theory : Linguistics, Lexicography and Implications*, pages 41–61. Ottawa University Press, Ottawa, 1990.
- Jacques Steinlin, Sylvain Kahane et Alain Polguère. Compiling a “classical” explanatory combinatorial lexicographic description into a relational database. Dans *Proceedings of the Second International Conference on the Meaning Text Theory*, pages 477–485, Moscou, 2005.
- Jacques Steinlin, Sylvain Kahane, Alain Polguère et Adil El Ghali. De l’article lexicographique à la modélisation objet du dictionnaire et des liens lexicaux. Dans *Actes de Euralex 2004*, pages 177–186, Lorient, 2004.
- Mathieu Valette, Alexander Estacio-Moreno, Etienne Petitjean et Evelyne Jacquy. Éléments pour la génération de classes sémantiques à partir de définitions lexicographiques. Pour une approche sémique du sens. Dans Piet Mertens, Cédric Fairon, Anne Dister et Patrick Watrin, dir., *Verbum ex machina, Actes de la 13^e conférence sur le traitement automatique des langues naturelles (TALN 06)*, volume 1 de *Cahiers du CENTAL*, pages 357–366. UCL Presses Universitaires de Louvain, 2006.

- Karel van den Eynde et Piet Mertens. Le dictionnaire de valence Dicovalence. Manuel d'utilisation - Université de Louvain, 17 novembre 2006.
- Serge Verlinde, Thierry Selva et Jean Binon. The Base Lexicale du Français (BLF) : A Multifunctional Online Database for Learners of French. Dans *Actes d'Euralex 2006*, pages 471–483, Turin, septembre 2006.
- Bernard Victorri et Catherine Fuchs. *La Polysémie : construction dynamique du sens*. Hermès, Paris, 1996.
- Ellen M. Voorhees. Using WordNet for Text Retrieval. Dans Christiane Fellbaum, dir., *WordNet : an Electronic Database*, chapitre 12, pages 285–304. MIT Press, 1998.
- Piek Vossen, dir. *EuroWordNet : a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, Dordrecht, 1998.
- Piek Vossen. Ontologies. Dans Ruslan Mitkov, dir., *The Oxford Handbook of Computational Linguistics*. Oxford University Press, Oxford, 2003.
- Leo Wanner et Margarita Alonso Ramos. What Type of Entity is a Lexical Function? Dans Jury Apresjan et Leo Iomdin, dir., *East West Encounter : Second International Conference in Meaning-Text Theory*, pages 518–528, Moscou, 2005. Slavic Culture Language Publishing House.
- Leo Wanner, Margarita Alonso Ramos et Antonia Martí. Enriching the Spanish EuroWordNet by Collocations. Dans *Proceedings of the LREC 2004*, Lisbonne, 2004.
- Leo Wanner et Bernd Bohnet. Towards Automatic Recognition of Lexical Function Instances. Dans Leo Wanner, dir., *Selected Lexical and Grammatical Issues in the Meaning-Text Theory*, pages 235–265. John Benjamins, Amsterdam/Philadelphie, 2007.

- Leo Wanner, Bernd Bohnet et Mark Giereth. What Is Beyond Collocations? Insights from Machine Learning Experiments. Dans *Actes d'Euralex 2006*, Turin, septembre 2006.
- Anna Wierzbicka. *Semantics : Primes and Universals*. Oxford University Press, Oxford, 1996.
- Morton E. Winston, Roger Chaffin et Douglas Herrmann. A taxonomy of part-whole relations. *Cognitive Science*, 11-4:417–444, 1987.
- Aleksandr K. Zholkovsky. *Sintaksis somali*. Nauka, Moskva, 1971.
- Aleksandr K. Zholkovsky et Igor Mel'čuk. O vozmožnom metode i instrumentax semantičeskogo sinteza. *Naučno-tehničeskaya informacija*, 5:23–28, 1965.
- Aleksandr K. Zholkovsky et Igor Mel'čuk. O sisteme semantičeskogo sinteza. I. stroenie slovarja. *Naučno-tehničeskaya informacija*, 11:48–55, 1966.
- Aleksandr K. Zholkovsky et Igor Mel'čuk. O semantičeskom sinteze. *Problemy kibernetiki*, 19:177–238, 1967.
- Aleksandr K. Zholkovsky et Igor Mel'čuk. Sur la synthèse sémantique. *T. A. Informations*, 2:1–85, 1970.
- Pierre Zweigenbaum, Bruno Bachimont, Jacques Bouaud, Jean Charlet et Jean-François Boisvieux. Le rôle du lexique sémantique et de l'ontologie dans le traitement automatique de la langue médicale. Dans Pierre Le Beux et Anita Burgun, dir., *Actes du Colloque CRISTAL'S*, Saint-Malo, 1996.

Index

- A**
- Assistant à l'encodage des liens lexicaux
230
- Attribut de description 144, 149
- C**
- Cible communicationnelle 216
- Classes sémantiques 180
- Classification des FL 159–161, 163
- Collocation 19, 30
- Base de la 30
- Collocatif 30
- Composantes sémantiques liées aux FL
174, 176
- Condition d'universalité 75
- Condition de cardinalité 74, 115
- Condition de diversité 75, 115
- D**
- Dérivation morphologique 133
- Dérivé morphologique 133
- Dérivatème 135
- Marqueur morphologique 139
- Dérivation sémantique 28
- Dérivé sémantique 29
- Diagnostic d'erreurs dans le DiCo.. 239,
241, 245, 251
- E**
- Élément de cadre 147
- Encodage algébrique 106
- Encodage explicite 106, 170, 198
- Encodage traditionnel 80
- Typologie de 86
- Encodage vulgarisé 79
- Étiquette sémantique 26, 243
- F**
- Fiche lexicographique 234
- Rédaction automatique de 234
- Reconstitution automatique de 252
- FL inverse 122, 162
- FL localement standard 78
- FL non standard (FLNS) 114
- Normalisation des 117
- Standardisation des 117
- FL réversibles 247
- Double implication 248
- Simple implication 247
- FL semi-standard 113
- FL standard 113

FL standard combinées	84		
Configuration de FL	85		
FL complexes	84		
FL standard simples	84		
		G	
Grammaire de FL	170		
Granularité			
des FL de l'encodage traditionnel	92		
du système des relations lexicales	20,		
99			
		I	
Indice de polysémie ou homonymie	240,		
242, 246			
		L	
Lien lexical	16		
Paramètres descriptifs de	230		
		M	
Marques d'usage	240		
Modèle lexical relationnel	9		
Modèle Sens-Texte	196		
		N	
Nombre de FL	92		
		P	
Parties du discours	238		
de surface	210		
des locutions	211		
profondes	209		
Patrons de dépendance syntaxique .	200		
		R	
		Relation lexicale	16, 247
		Encodage automatique de	247
		Relation paradigmatic	17, 31, 203
		Relation syntagmatic	17, 31
		Requêtes SQL	232
		Ressource lexicale relationnelle	9, 23, 25
		S	
		Système de FL	157, 172
		Implémentation du	228
		Perspective combinatoire	195
		Perspective communicationnelle	216
		Perspective sémantique	173
		Perspective sur les parties du discours	208
		Propriétés du	172
		Système lexical	228

Annexe I

Liste des FL citées

Cette annexe présente les différentes FL citées dans la thèse. Nous les répartissons dans deux sections : la première regroupe les FL standard et semi-standard (ces dernières figurent sous les FL standard auxquelles elles se rattachent) ; la seconde présente les FL non standard. Dans chacune des deux sections, les FL sont énumérées par ordre alphabétique. Chaque FL est accompagnée d'une glose explicative (mise à part les FL non standard qui sont suffisamment explicites) et d'un exemple. Les FL (standard, semi-standard ou non standard) marquées d'un petit cercle (o) sont celles que nous avons proposées dans la thèse ; celles marquées d'un losange (◊) ont été proposées par d'autres auteurs, notamment Erastov [1969], Fontenelle [1997b] et Grizolle [2003] ; enfin, celles marquées d'un astérisque (*) sont celles que nous proposons d'éliminer. Bien que les FL aient, dans la plupart des cas, plusieurs valeurs, nous n'en reproduisons qu'une par exemple afin d'alléger la présentation.

Fonctions lexicales standard et semi-standard

$A_0(\text{chat}) = \text{félin} \rightarrow \text{Équivalent adjectival}$	$\text{AntiBonFact}_0(\text{robinet}) = \text{couler} \rightarrow \sim \text{fonctionne mal}$
$A_1(\text{abandon}\#I.2b) = \text{abandonné} \rightarrow [Y] \text{ qui est dans un état de } \sim$	$\text{AntiBon.Oper}_{12}(\text{compliment}\#a) = \text{débiter} [\text{ART } \sim] \rightarrow [X] \text{ dire } \sim \text{ de façon méca-}$ nique
$A_1\text{ProxOper}_1(\text{mort}(1)\#I.1) = \text{mourant} \rightarrow [X] \text{ qui est proche de } \sim$	$\text{AntiBon.Real}_{1\oplus}(\text{blaxon}\#a) = \text{balancer} [\text{ART } \text{coup de } \sim \text{ à } N] \rightarrow [X] \text{ utiliser } \sim \text{ en}$ manifestant de l'agressivité envers N
$A_2(\text{déception}) = \text{décevant} \rightarrow [Y] \text{ à cause duquel } X \text{ éprouve } \sim$	$\text{AntiBonS}_1(\text{théâtre}\#I.1) = \text{cabotin} \rightarrow \text{Nom pour } X \text{ que l'on désapprouve}$
$A_2\text{AntiReal}_1(\text{frigo}) = \text{vide} \rightarrow \text{Qui n'est pas rempli}$	$\text{AntiBonS}_2(\text{bateau}) = \text{passager clandestin} \rightarrow \text{Individu } Y \text{ qui utilise illégalement } \sim$
$A_2\text{Manif}(\text{haine}) = \text{chargé} [\text{de } \sim] \rightarrow [\text{Geste ou parole de } X] \text{ qui est une manifestation}$ de \sim	$\text{AntiMagn}(\text{aboiement}) = \text{faible} \rightarrow \text{Peu intense}$
$A_2\text{NonReal}_1(\text{banc}(1)) = \text{inoccupé} \rightarrow \text{Qui n'est pas utilisé}$	$\text{AntiMagn}^{\text{échange}}(\text{conversation}\#a) = \text{norme} \rightarrow \text{Durant laquelle } X \text{ et } Y \text{ sont peu actifs}$
$\text{Able}_1(\text{crainte}) = \text{craintif} \rightarrow [X] \text{ qui tend à éprouver } \sim$	$\text{AntiMagn}^{\text{plaisir}}(\text{conversation}\#a) = \text{houleuse} \rightarrow \text{Peu plaisante}$
$\text{Able}_2(\text{estimer}\#II.2) = \text{estimable} \rightarrow [Y] \text{ qui tend à être } \sim$	$\text{AntiMagn}^{\text{temp}}(\text{conversation}\#a) = \text{brève} \rightarrow \text{Qui dure peu de temps}$
$\text{Adv}_0(\text{absolu}) = \text{absolument} \rightarrow \text{Équivalent adverbial}$	$\text{AntiReal}_1(\text{paupière}) = \text{ouvrir} [\text{ART } \sim] \rightarrow [\text{Individu } X] \text{ relever } \sim$ complètement $\text{AntiReal}_1(\text{armoire}) = \text{vider} [\text{ART } \sim] \rightarrow [X] \text{ faire en sorte que } \sim \text{ soit}$ non utilisée
$\text{Adv}_0\text{NonManif}(\text{admirer}\#I) = \text{secrètement} \rightarrow \text{Sans le manifester}$	$\text{AntiReal}_2(\text{ultimatum}) = \text{refuser} [\text{ART } \sim] \rightarrow [Y] \text{ ne pas obéir à } \sim$
$\text{Adv}_1(\text{gratitude}) = \text{avec } [\sim] \rightarrow [X \text{ faire qqch.}] \text{ en éprouvant } \sim$	$\text{AntiVer}(\text{applaudir}) = \text{mollement} \rightarrow \text{De façon non sincère}$
$\text{Adv}_1\text{Real}_1(\text{flèche}\#III) = \text{dans le sens} [\text{de } \text{ART } \sim] \rightarrow [X \text{ se déplacer}] \text{ en utilisant } \sim$	$\text{AntiVerbut}(\text{conversation}\#a) = \text{stérile} \rightarrow \text{Non productive}$
$\diamond \text{Antecedent}(\text{shut}[\text{event}]) = \text{open}$	$\text{AntiVer.Magn}(\text{enthousiasme}\#1) = \text{excessif} \rightarrow \text{Non fondé et intense}$
$\text{Anti}(\text{joyeux}) = \text{triste} \rightarrow \text{Antonyme}$	$\diamond \text{Augm}(\text{dom. 'house'}) = \text{domisce} \rightarrow \text{Augmentatif}$
$\text{AntiBon}(\text{aboiement}) = \text{intempestifs} \rightarrow \text{Indésirables}$	$\text{Bon}(\text{aspiration}) = \text{noble} \rightarrow \text{Bon/Bien}$
$\text{AntiBon}^{\text{clairé}}(\text{conversation}\#a) = \text{décousue} \rightarrow \text{Peu claire}$	$\text{Bon}^{\text{contenu}}(\text{conversation}\#a) = \text{intelligente} \rightarrow \text{Intéressante}$
$\text{AntiBon}^{\text{contenu}}(\text{conversation}\#a) = \text{inintéressante} \rightarrow \text{Peu intéressante}$	$\text{Bon.CausPredPlus}(\text{vie}\#I.2) = \text{enrichir} [\text{ART } \sim] \rightarrow [\text{Qqn/Qqch.}] \text{ rendre } \sim \text{ de } X \text{ plus}$ de bas niveau
$\text{AntiBon.AntiMagn}(\text{bagage}\#I.1) = \text{maigres} \rightarrow \text{Peu volumineux et tout juste suffisants}$	difficile
$\text{AntiBon.AntiMagn.Oper}_1(\text{conversation}\#b) = \text{avoir peu} [\text{de } \sim] \rightarrow [X] \text{ faire preuve de}$ \sim de bas niveau	$\text{BonS}_{1/2}(\text{débat}\#2) = \text{débattre} \rightarrow \text{Nom pour } X/Y \text{ qui a du talent pour } \sim$
$\text{AntiBon.AntiVer.Oper}_1(\text{conversation}\#a) = \text{accaparer} [\text{ART } \sim] \rightarrow [X] \text{ ne pas laisser } Y$ parler suffisamment durant \sim	$\text{Cap}(\text{assemblée}\#2) = \text{président} [\text{de } \text{ART } \sim] \rightarrow \text{Chef de } \sim$
$\text{AntiBon.CausPredPlus}(\text{vie}\#I.2) = \text{comploter} [\text{ART } \sim] \rightarrow [\text{Qqn/Qqch.}] \text{ rendre } \sim \text{ de}$ $X \text{ plus difficile}$	$\text{CausContFunc}_0(\text{débat}\#1) = \text{alimenter} [\text{ART } \sim] \rightarrow [\text{Qqch.}] \text{ causer que } \sim \text{ continue}$

CausDe_nouveauFact₀(*évier*) = *déboucher* [ART ∼] → [Qqn] rendre ∼ de nouveau fonctionnel
 CausDe_nouveauFunc₀(*silence*#I.2a) = *ramener* [ART ∼] → [Qqn/Qqch.] causer de nouveau ∼
 CausFact₀(*imagination*) = *fire* (*sb's imagination*) → Déclencher ∼
 CausFunc₀(*bouquet*#I.1) = *assembler* [ART ∼] → [Qqn] fabriquer ∼
 Caus₁Fact₀(*verrou*#I) = *pousser* [ART ∼] → [X] causer que ∼ ferme Y plusieurs Caus₂Func₀(*témoignage*#I.1) = *collecter* [ART ∼s] → [Y] faire en sorte d'obtenir plusieurs ∼
 Caus₂Func₁(*hilarité*) = *déclencher* [ART ∼ Loc-in N=X] → [Y] causer ∼ de X
 CausOper₁(*silence*) = *réduire* [qqn. au ∼] → [Qqn/Qqch.] causer ∼ en échange de qqch. CausOper₁(*silence*) = *acheter* [le ∼ de qqn.] X et Y étant plutôt des enfants, Caus₁Oper₁(*ami*#I.1) = *faire* [∼ avec N=Y]
 CausPredMinus(*fardeau*#II) = *alléger* [ART ∼] → [Qqn/Qqch.] rendre moins important ∼
 Centri(*débat*#I) = *centre* → Aspect de Z le plus important dans ∼
 Cf(*aboiement*#I) = *grognement* → Confère
 ◦Compos(*assiette*) = *porcelaine* → Matériau dont est fait ∼
 ◦Consequent (*open*[state]) = *shut*
 Contr(*cigarette*#I) = *pépe* → Contrastif
 Conv₂₁(*craindre*) = *effrayer* [N=X] → Conversif
 ◦Cum(*citizen*) = *compatriot* → Appartenant au même groupe
 ContOper₁(*espérance*#I) = *garder* [ART ∼] → [X] continuer à avoir ∼
 Culm(*haine*) = *paroxysme* → Maximum d'intensité de ∼
 Degrad(*maison*#I.1) = *se délabrer* → ∼ se dégrade
 De_nouveauFunc₀(*hostilité*) = *reprandre* [ART ∼] → ∼ commencent de nouveau
 De_nouveauOper₁(*accusation*#I.1) = *réitérer* [ART ∼] → [X] émettre de nouveau ∼
 ◦Dimin(*dom*¹*house.*) = *domik* → Diminutif
 Equip(*navire*) = *équipage* → Nom pour X

Essayer de IncepReal₁(*fauteuil*) = *brigner* [ART ∼] → [X] essayer d'occuper ∼
 Essayer de Oper₁₂(*victoire*#I) = *disputer* [ART ∼ à N=Y] → [X] essayer d'obtenir ∼
 essayer de Oper₂(*augmentation*#2) = *demander* [ART ∼] → [Y] essayer d'obtenir ∼
 Fact₀(*cri*#I.1) = *parvenir* [à N] → ∼ est entendu par N
 Fact₀(*pouls*#I) = *battre* → ∼ est en activité
 Fact₀boire(*chat*) = *laper*
 Fact₀chasser(*chat*) = *chasser*
 Fact₀gratter(*chat*) = *faire ses griffes*
 Fact₀nouvement(*chat*) = *Γ faire le gros dos*¹
 Fact₀nettoyage(*chat*) = *faire sa toilette*
 Fact₁(*abat-jour*) = *coiffer* [N=X] → ∼ est sur N
 Fact₂(*charrette*) = *transporter* [N=Y] → ∼ transporte Y en cour Fact₂(*avocat*(I)#I) = *plaider* → ∼ travaille pour Y en cour
 Fact₂₁(*piste*#II) = *mener* [N=Y à N=X] → ∼ est suivie avec succès
 ◦Fas(*house*) = *front* → Partie avant de ∼
 FinFact₀(*brake*) = *fail* → ∼ arrête de fonctionner
 FinFunc₀(*abcès*#II) = *crever* → ∼ disparaît
 FinOper₁(*abîme*#II.1) = *sortir* [de ART ∼] → [X] cesser d'être dans ∼
 Func₀(*commémorage*#a) = *circuler* → ∼ ont lieu
 Gener(*coq*) = *gallinacé* → Terme générique
 ◦Hollo(*plaque*#I.3) = *terre* → Hollonyme
 ◦Hypo(*chanson*#I.1) = *berceuse* → Hyponyme
 IncepFunc₁(*bourgeon*) = *apparaître* [sur N=X] → ∼ apparaît sur X
 IncepFunc₂(*fardeau*#II) = *tomber* [sur (les épaules de) N=Y] → ∼ commence à se faire sentir sur Y
 IncepOper₁(*extase*#I) = *tomber* [en ∼] → [X] commencer à éprouver ∼
 IncepPredMinus(*fièvre*#I) = *tomber* → ∼ devient moins intense
 IncepPredDifférent(*vie*#I.2) = *changer* → ∼ commence à être différent
 IncepPredPlus(*abcès*) = *mûrir* → Commencer à devenir plus important

IncepReal₁(gare#a) = entrer [en ~] → [X] arriver dans ~
 Instr(couteau) = au [~] → [X faire qqch.] en utilisant ~
 Involv(plaque#I.1) = recouvrir [N] → ~ se trouve sur N
 Labor₂₁(émot) = remplir [N=X d'~] → [Y] être la cause de ~
 ◇ Lim(territoire) = frontière → Limite
 Labreal₁₂(bague#2) = munir [N=Y de ART ~] → [X] mettre ~ à Y
 ouvrir Labreal₁₂(clé#I.1) = ouvrir [N=Y2 avec ART ~] → [X] utiliser ~ pour ouvrir
 Y2
 fermer Labreal₁₂(clé#I.1) = fermer [N=Y2 à ~] → [X] utiliser ~ pour fermer Y
 pour frapper qq. Labreal₁@(ceinture#I.1) = battre []
 LiquFact₀(tuyau#I.1) = crever [ART ~] → [Qqn/Qqch.] rendre non fonctionnel ~
 LiquFact₀-air(pneu) = dégonfler [ART ~] → [Qqn] causer que ~ se vide de son air
 LiquFact₀-paroi(pneu) = crever [ART ~] → [Qqn/Qqch.] causer que ~ subisse un
 déchirement de paroi
 LiquFunc₀(espoir#I) = enterrer [ART ~] → [Qqn] faire cesser ~
 (temporairement) LiquFunc₀(conversation#a) = interrompre [ART ~]
 Liqu_{1,2}Func₀(abcès#II) = crever [l'~] → [X et Y] causer que ~ disparaisse
 ◇ Liquidative (mistake) = erase → Liquidatif
 LiquOper₁(pleurs) = sécher [~s] → [Qqn/Qqch.] faire cesser ~ de X
 ◇ Loc(work) = at [~] → [X] se trouvant dans ~
 Loc_{ab}(balcon) = de [~] → À partir de ~
 Loc_{in}(bataille#I.1) = durant [ART ~] → Au moment de ~
 ◇ LocLim(country) = within the borders [of ART ~] → Aux limites de ~
 Magn(pluie) = torrentielle → Intense
 Magn_{echange}(conversation#a) = enflammée → Durant laquelle X et Y sont très
 actifs
 Magn_{plaisir}(conversation#a) = amicale → Plaisante
 Magn_{temp}(conversation#a) = longue → Qui dure longtemps
 Magn_{quant}(coup de feu#I) = nourris → Nombreux

Magn.A₁(dégout) = malade [de ~] → [X] qui éprouve ~ intense
 MagnFact₀(brake) = grip → ~ fonctionne de façon intense
 Magn.Fact₁(angoisse) = dévorer [N=X] → ~ intense a un effet sur X
 Magn.IncepLabor₃₁(admiration) = saisir [N=X d'~] → [Z] commencer à être l'objet
 de ~
 Magn.IncepOper₁(abattement(1)) = sombrer [dans ART ~] → [X] commencer à être
 dans un état de ~
 Magn+Involv(barbe) = dévorer [ART joues/visage de N=X] → [~] envahir N
 Magn.Labor₂₁(enthousiasme#1) = remplir [N=X de ~] → [Y] être la cause de ~
 intense
 Manif(pleurs) = couler → ~ se manifestent
 ◇ Matur(fruit) = to ripen → Commencer à être prêt
 ◇ Mero(cerise) = queue → Méronyme
 ◇ Metr(inflation) = rate → Mesure
 ◇ Mobil(water) = to flow, to run → Être en mouvement de façon prototypique
 Mult(abeille) = colonie → Ensemble de ~
 Nocer(moustique) = piquer → ~ fait une action jugée négative
 Non(désir#1) = désintéret → Absence de ~
 NonAdv₁(amimosité) = sans [~] → [X faire qqch.] sans éprouver/manifeste ~
 NonFact₀(usine) = chômer → ~ ne fonctionne pas
 NonOper₁(tact) = manquer [de ~] → [X] ne pas manifester ~
 NonOper₁actual(imagination#2) = être [à court] [de ~]
 NonPermFact₀(zone) = boucler [ART ~] → [Qqn] rendre ~ non accessible
 NonReal₂(chance#2) = [laisser passer] [ART ~] → [Y] ne pas saisir ~
 Oper₁(retraite#I.1) = battre [en ~] → [X] effectuer ~
 Oper₁₂(avertissement#I.1) = émettre [ART ~ à N=Y] → [X] communiquer ~
 Oper₁₂₃(défaite) = encaisser [ART ~ [face à [N=Y dans N=Z] → [X] subir ~
 Oper₂(cadeau#I) = recevoir [ART ~] → [Y] recevoir ~
 Oper₂₁([coup de fil]) = recevoir [ART ~ de N=X] → [Y] recevoir ~ de X

- Oper₂₃(*rancune*) = *s'attirer* [ART ~ par N=Z] → [Y] être visé par ~ de X à cause de Z
- Oper₃(*rancune*) = *engendrer* [ART ~] → [Z] être la cause de ~ de X
- ◊Part(*brake* / *frein*) = *drag* → Partie de ~
- ◊Pel(*bean*) = *pod* → Quelque chose qui couvre ~
- Prepar₁(*meurtre*) = *comploter* [ART ~] → [X] préparer ~
- ◊Prim(*city*) = // *capital* → ~ le plus important
- Propt(*animosité*) = *par* [ART] → [X faire qqch.] du fait de ~
- QAnti(*compliment*#a) = *réprimande* → Quasi-antonyme
- QS₀(*ému*#1) = *émotion* → Quasi-équivalent nominal
- QS₁(*abâiquer*#1) = *monarque* → Type particulier de X
- QSyn(*bataille*#II.1) = *combat* → Quasi-synonyme
- Qual₂(*haine*) = *odieux* → [Y] envers qui on tend à éprouver ~
- Real₁(*tambour*#I.1) = *battre* [ART ~] → [X] utiliser ~
- Real₁^I(*colle*#III) = *donner* [ART ~ à N=Y pour N=Z] → [X] imposer ~ à Y
- Real₁^{usual}(*cinéma*#I.3) = *fréquenter* [ART ~]
- ~ étant parure Real₁(*couronne*#I.2) = *porter* [ART ~]
- En tant que carrière Real₁(*théâtre*#I.1) = *faire* [du ~]
- destinée à fermer un passage Y Real₁(*chaîne*#I) = *tendre* [ART ~]
- pour éclairer Real₁(*allumette*) = *brûler* [ART ~]
- pour subsistance Real₁(*plume*#III.a) = *vivre* [de A-poss=X ~]
- rapidement Real₁(*bière*(1)#b) = *siffler* [ART ~]
- Real₁₂(*bague*#2) = *poser* [ART ~ à N=Y] → [X] mettre ~ à Y
- Real₁₃(*avocat*(1)#1) = *passer* [ART ~ dans Z] → [X] utiliser ~ pour nettoyer Z
- Real₂(*avocat*(1)#1) = *consulter* [ART ~] → [Y] être le client de ~
- Real₂^I(*accusation*#I.2) = *se défendre* [de ART ~]
- Real₂^{II}(*accusation*#I.2) = *se blanchir* [de ART ~]
- Real₃(*mort*(2)) = *enterrer* [ART ~] → [Qqt] s'occuper de ~
- Real₃^I(*crime*#I.a) = *instruire* [ART ~]
- Real₃^{II}(*crime*#I.a) = *élucider* [ART ~]
- Real₃^{III}(*crime*#I.a) = *punir* [ART ~]
- dans le sens du courant Real₃(*fleuve*#I) = *descendre* [ART ~]
- dans le sens inverse du courant Real₃(*fleuve*#I) = *remonter* [ART ~]
- Result(*to buy*) = *to have* → Action résultant de ~
- S₀(*admirer*#1) = *admiration* → Équivalent nominal
- S₀IncepPredPlus(*mécontentement*) = *montée* [de (ART) ~] → Intensification de ~
- S₀Pred(*célibataire*) = *célibat* → Situation de ~
- S₀Real₁(*désir*#1) = *accomplissement* [de ART ~] → Satisfaction de ~
- S₀SingReal₁(*chiffon*#a) = *coup* [de ~] → Instance d'utilisation de ~
- S₀Son(*cheval*) = *hennissement* → Nom de son typique
- S₁(*conduire*) = *conducteur* → Nom pour X
- S_{1/2}(*bataille*#II.1) = *adversaire* → Nom pour X/Y
- S₁CausFunc₀(*farine*) = *meunier* → Celui qui fabrique ~
- S₁CausPredIllég₁AntiVer(*billet*#II) = *contrefacteur* → Individu qui fabrique ~ illé-gaux
- S₁Real₂ issue favorable(*compétition*#2) = *vainqueur* → Nom pour Y qui gagne ~
- S₂(*balai*) = *poussière* → Nom pour Y
- S₃(*chèque*) = *porteur* → Nom pour Z
- ◊S-INIT(*wake up*) = *sleep* → État antérieur
- S_{instr}(*fumer*#2) = *fume-cigarette* → Dispositif aidant à ~
- S_{instr}Real₁(*frein*#I) = *pédale* → Instrument permettant d'actionner ~
- S_{loc}(*vendre*) = *marché* → Endroit où se déroule ~
- S_{med}(*excuse*#II) = *lettre* [d'~] → Moyen de réaliser ~
- S_{mod}(*vie*#I.2) = *genre* → Façon dont se passe ~
- S_{res}(*labeur*) = *fruit* → Résultat de ~
- Sing(*dialogue*#1b) = *dialogue* → Unité de ~
- Son(*grenouille*) = *coasser* → Produire un son typique
- ◊State(*erase*) = *mistake*

Sympt₁(*crainte*) = *frémir* [de ~] → [X] manifester ~
Syn(*frigidaire*) = *réfrigérateur* → Synonyme
V₀(*funérailles*) = *enterrer* [ART ~] → Équivalent verbal
◊Vacant(*seat*) = *free* → Vacant
Ver(*ami*) = *fidèle* → Conforme à son objectif/fonction
Verbut(*conversation*#a) = *constructive*

Vercontenu(*conversation*#a) = *franche*
Vermanière(*accuser*#I.1) = *ouvertement*
Vermotif(*accuser*#I.1) = 「à raison」
Ver/Bon(*avocat*(1)#I) = *disert*
VerFact₀(*montre*) = *être exacte* → ~ fonctionne bien

Fonctions lexicales non standard

- ~ des pirates(*dirapeau*) = *noir*
- *~ femelle(*mouton*) = *brébis*
- [Individu X] abaisser et relever rapidement \sim (*paupière*) = *battre* [des ~s]
- *Affections de \sim (*estomac*#I) = *crampe*
- *Aliment préparé dérivé de \sim (*lait*#b) = *fromage*
- o Altération de \sim (*estomac*#I) = *crampe*
- *Animal dont provient \sim (*agneau*#c) = *agneau*
- *Animal qui se nourrit de \sim (*grain*(1)#I.1) = *granivore*
- Appareil servant à mesurer l'intensité avec laquelle X \sim (*applaudir*#I) = *applaudir*-*mètre*
- *Arbre qui donne \sim (*cerise*) = *cerisier*
- Artefact dans lequel on transporte \sim (*cigarette*#1) = *étui*
- o Avant(*compiler*) = *précompiler*
- o Axe(*histoire*) = *trame* [de ART ~]
- *Boisson alcoolisée faite avec \sim (*cerise*) = *cherry*
- *Bouton/furoncle sur \sim (*paupière*) = *compère-loriot*
- *Catégorisé comme étant le plus économique pour les passagers(*wagon*) = *de seconde* (*classe*)
- *Catégorisé comme offrant le meilleur confort aux passagers(*wagon*) = *de première* (*classe*)
- *Ce à partir de quoi on fait \sim (*tirage*(1)#2a) = *pellicule*
- *Ce avec quoi on fabrique \sim (*cuir*) = *peau*
- Chair de \sim utilisée comme aliment(*mouton*#I.a) = *mouton*
- Commis par Adam et Eve(*péché*) = *originel*
- *Coordonnées officielles de \sim (*maison*#I.1) = *adresse*
- *Dans lequel il est interdit de fumer(*wagon*) = *non-fumeur*
- *Dans lequel il est permis de fumer(*wagon*) = *fumeurs*
- *De couleur blanche(*barbe*) = *blanche*
- *De couleur grise(*barbe*) = *grisonnante*
- *Degré de gonflement de \sim (*pneu*) = *pression*
- *Degré de réduction de Y dans un \sim (*plan*(2)#I) = *échelle*
- *Dépôt qui se forme au fond d'un contenant de \sim (*bière*(1)#a) = *lie*
- *De sexe féminin(*avocat*) = *avocate*
- *Dimension des balles que l'on utilise avec \sim (*fusil*(1)#a) = *calibre*
- *Discipline qui étudie \sim (*médicament*) = *pharmacologie*
- *Discipline scientifique qui étudie les phénomènes liés à \sim (*vie*#I.1) = *sciences de la* \sim
- o Document associé(*congrès*) = *actes* [de ART ~]
- Document obtenu par X à un G. comme trace d'une transaction(*guichet*#2) = *reçu*
- *Document officiel où les informations sur la \sim de X sont enregistrées(*naissance*#I) = *acte* [de ~]
- Document qui décrit le déroulement du C. (*concert*) = *programme* [de ART ~]
- Dont la raison est l'amour(*crime*#I.b) = *passionnel*
- *Dont le contenu Z est important(*conversation*#a) = *grave*
- Dont le contenu Z est peu <n'est pas> important(*conversation*#a) = *légère, frivole, futile, insignifiante*
- *Drapeau de \sim (*pirate*#1) = *drapeau noir*
- Durant laquelle X et Y interviennent comme ils veulent sans que \sim ait une structure définie(*conversation*#a) = \lceil à bâtons rompus \rceil
- Durant laquelle X et Y sont seuls(*conversation*#a) = \lceil en tête-à-tête \rceil , *privée*
- Durant laquelle X et Y traitent en détail un sujet donné(*conversation*#a) = *suivie*
- *Durci, renforcé par une armature métallique(*ciment*#I) = *armé*

- *Élément du corps de \sim (*abeille*) = *pattes*
- Endroit où l'on range \sim (*balai*) = *placard*
- En regardant à travers l'orifice de \sim (*serrure*) = *par le trou de la* \sim
- *Enregistrement officiel de \sim (*bateau*) = *immatriculation*
- En s'asphyxiant avec du gaz(*se suicider*) = *au gaz*
- *Entité géologique dont fait partie \sim (*plaque*#I.3) = *terre*
- *Équipé d'un bar(*wagon*) = \sim -*restaurant*
- *Équipé pour que les passagers y dorment(*wagon*) = \sim -*lit*
- *Équipé pour que les passagers y mangent(*wagon*) = \sim -*restaurant*
- *Espèce de \sim (*canard*) = *colvert*
 - oÉtude de \sim (*signe*#III) = *astrologie*
 - *Étude des \sim (*signe*#III) = *astrologie*
- Étui dans lequel on range \sim (*poignard*) = *étui*
- *Fait pour être utilisé par une seule personne(*drap*) = *simple*
- *Fait pour être utilisé simultanément par deux personnes(*drap*) = *double*
 - oFem(*coq*) = *poule*
 - *Femelle du \sim (*canard*) = *cane*
 - *Femelle(*chien*) = *chienne*
- *Foie de \sim produit par gavage(*canard*#I.b) = Γ *foie gras* Υ
- Fourrure/peau de \sim utilisée comme matériau(*mouton*#I.a) = *mouton*
- *Friandise contenant beaucoup de \sim (*sucré*#a) = *sucrerie*
- Fumée en forme de rond que X produit en fumant \sim (*cigarette*#1) = *rond de fumée*
- *Gâteau fait avec \sim (*cerise*) = *clafoutis*
- Habitant de \sim (*terre*#III) = *terrien*
- Individu qui fabrique \sim (*charrette*#a) = *charron*
- Individu qui s'intéresse à \sim (*curiosité*#II) = *amateur* [de \sim s]
- *Ingrédient avec lequel on fait \sim (*yahourt*) = *lait fermenté*
- *Ingrédient utilisé pour faire \sim (*bière*) = *blé*
- *Jeune \sim (*coq*) = *coquelle*
- *Larve de \sim (*grenouille*) = *têtard*
- *Longueur moyenne de \sim dans une population de X(*vie*#I.1) = *espérance* [de \sim]
- *Mâle de \sim (*poule*) = *coq*
- *Mâle(*abeille*) = *faux bourdon*
 - oMarque(*pneu*) = *empreinte*
- *Marque laissée dans le sol par un \sim (*pneu*) = *empreinte*
 - oMasc(*jument*) = *cheval*
- *Matériau dont on fait des \sim (*bijou*) = *pierre*
- *Matériau dont sont faits les \sim (*pneu*) = *caoutchouc*
 - *Matériau pour \sim (*corde*) = *lin*
- *Matériau précieux dont peut être fait la \sim (*vaisselle*) = *argent*
- *Matériau typique dont sont faites les \sim (*assiette*) = *porcelaine*
- *Mélange alimentaire que l'on cuit pour faire \sim (*gâteau*) = *pâte*
 - oMetr*capacité*(*restaurant*) = *couvert*
 - o Metr*courant*(*robinet*) = *débit*
 - o Metr*dimension*(*bateau*) = *calaison*
 - o Metr*durée*(*vie*#I.1) = *espérance*
 - o Metr*qualité*(*conversation*#b) = *niveau*
 - o Metr*résultat*(*partie*(2)) = *score*
 - o Metr*vitesse*(*bateau*) = *neud*
- *Mesure de la taille de \sim (*bateau*) = *calaison*
- *Mesure de \sim (*pouls*#I) = *pulsations*
- *Mesure du courant de \sim (*fleuve*#I) = *débit*
- *Mesure du courant de \sim (*rivière*#I) = *débit*
- *Mesure de \sim (*poids*) = *kilogramme*
- Mousse qui se forme sur une \sim quand on la prépare(*bière*(1)#b) = *mousse*
- *Nom d'une association qui regroupe les \sim de Y(*ami*#II) = *Les amis*
- *Nom de \sim ayant eu lieu à l'endroit N(*bataille*#I.1) = [bataille] *de N*
- *Paramètres selon lesquels on évalue \sim dans la société(*vie*#I.2) = *conditions* [de \sim]

- Parasite du C. (*chien*#I) = *puce*
 Parasite(*arbre*) = *gui*
 *Parcours suivi par \sim (*fleuve*#I) = *cours*
 *Partie de \sim que l'on tient(*couvert*) = *poignée*
 *Partie de \sim (*chèque*) = *talon*
 o Petit de \sim (*chat*) = *chaton*
 *Pièce d'habillement de \sim à la cour(*avocat*#I) = *épitoge*
 *Plat à base de \sim (*agneau*#b) = *blanquette* [de \sim]
 Plat à base de viande de \sim (*coq*) = *au vin*
 Plutôt haut et sans ampleur(*timbre*#2) = *grêle*
 *Points obtenus par X/Y dans \sim (*partie*(2)) = *score*
 *Pour les troupes d'une armée en guerre(*théâtre*#I.1) = *aux armées*
 Produite à l'étranger(*bière*) = *d'importation*
 Provoqué volontairement à des fins thérapeutiques(*abcès*) = *de fixation*
 *Publication résultant de \sim (*congrès*) = *actes* [de ART \sim]
 Pur(*timbre*#2) = *argentin*
 * [Qqch.] qui a la consistance <le goût> de \sim (*farine*) = *farineux*
 [Qqch.] qui a la forme de \sim (*cloche*) = *en*
 * [Qqch.] qui a l'aspect ou la consistance de \sim (*lait*#b) = *laiteux*
 [Qqch.] qui a le goût de \sim (*sucré*#a) = *sucré*
 * [Qqch.] qui évoque le pelage de \sim (*tigre*) = *tigré*
 [Qqch.] qui fait penser à \sim (*mort*(1)#I.1) = *funèbre*
 *Qualité de \sim de X(*conversation*#b) = *niveau* [de \sim]
 *Quantité de marchandise que \sim peut transporter(*bateau*) = *tonnage*
 *Que X porte pour chasser(*chapeau*) = *de chasse*
 *Que X porte pour se protéger du soleil(*chapeau*) = *de soleil*
 *Qui a des parois et un toit pour abriter Y(*wagon*) = *couvert*
 Qui a des plis comme décoration(*jupe*) = *à plis*
 *Qui a été utilisée(*assiette*#a) = *sale*
- *Qui a la rage(*chien*#I) = *enragé*
 *Qui a l'aspect <la consistance> de \sim (*cire*) = *cireux*
 Qui a lieu devant les tribunaux(*bataille*#II.1) = *juridique*
 *Qui a lieu entre les chefs de deux partis(*bataille*#II.1) = *de chefs*
 *Qui a lieu le matin(*concert*) = *matinée*
 *Qui a lieu malgré la volonté de X(*célibat*) = *forcé*
 *Qui a trait à la religion(*juron*#1) = *blasphématoire*
 *Qui a une forme évasée(*barbe*) = \lceil en *éventail* \lceil
 *Qui a une forme pointue(*chapeau*) = *pointu*
 *Qui a une forme ronde(*chapeau*) = *rond*
 *Qui comporte deux niveaux pour les passagers(*wagon*) = *à impériale*
 Qui descend jusqu'aux genoux de X(*jupe*) = *au genoux*
 Qui descend jusqu'aux mollets de X(*jupe*) = *à mi-mollet*
 *Qui est à la Num-e position dans ensemble de W.(*wagon*) = *Num-e*
 *Qui est causé par un incendie(*dégât*) = \lceil du *feu* \lceil
 Qui est commis par des soldats pendant une guerre(*crime*) = *de guerre*
 *Qui est fabriqué par cuisson(*jambon*) = *braisé*
 *Qui est fait de feutre(*chapeau*) = \ll de *feutre*
 *Qui est fait de paille(*chapeau*) = *de paille*
 *Qui est fait, pour ce qui est des ingrédients, à base de farine brute(*pain*#I) = *complet*
 *Qui est fait, pour ce qui est des ingrédients, sans sel(*pain*#I) = *sans sel*
 *Qui est fait, pour ce qui est du mode de fabrication, de façon traditionnelle ou qui en a l'apparence(*pain*#I) = \lceil à *l'ancienne* \lceil
 *Qui est le dernier dans un ensemble(*wagon*) = *de queue*
 *Qui est le premier dans un ensemble de W.(*wagon*) = *de tête*
 Qui est liée à des irrégularités commises par X(*faillite*) = *frauduleuse*
 *Qui est orné de plumes(*chapeau*) = *à plumes*
 *Qui est plutôt fréquenté par des familles(*restaurant*) = *familial*
 Qui est un symbole de respect pour le mort N Real₁(*couronne*#I.2) = *déposer*

- Qui évoque \sim (farine) = *farineux*
- * Qui fait penser à \sim (éléphant#1) = *pachydermique*
- * Qui fonctionne grâce à un mécanisme(*montre*) = *mécanique*
- * Qui n'a ni paroi ni toit(*wagon*) = à *plateforme, découvert, plat*
- Qui ne vit pas à l'état sauvage(*canard*) = *d'élevage*
- * Qui nuit à la réputation de Y(*peine*#II) = *infamante*
- * Qui possède un carillon(*horloge*) = à *carillon*
- * Qui provient d'un mélange de races(*chien*#I) = *bâtard*
- * Qui provient d'un tonneau, et non d'une bouteille(*bière*(1)#b) = *pression*
- Qui se nourrit de \sim (viande) = *carnivore*
- Qui sert à contrôler la vitesse d'un deux-roues(*poignée*#I) = *d'accélérateur*
- Qui sert à ouvrir une fenêtre(*poignée*#I) = *de fenêtre*
- Qui sert à ouvrir une porte(*poignée*#I) = *de porte*
- Qui sert pour toutes les portes d'un bâtiment(*clé*#I.1) = *passé*
- * Qui se trouve en-dessous d'où l'on est(*étage*#1) = *du dessous*
- * Qui suit ou précède le W. en question(*wagon*) = *voisin-adj*
- * Qui tue l'être Y(*coup de feu*#I) = *mortel*
- * Qui vise le grand public(*cinéma*#I.1) = *commercial*
- * Qui vit à l'état sauvage(*canard*#I.a) = *sauvage*
- Rappelant le métal qu'on frappe(*timbre*#2) = *métallique*
- * Rapport rédigé par X pour rendre compte de \sim (*autopsie*) = *rapport* [de \sim]
- Réciprocité(*connexion*) = *interconnexion*
- Si Étude de \sim (*crime*#I.a) = *criminologue*
- * Science qui étudie \sim (*poisson*#I.a) = *ichtyologie*
- * Science qui étudie ce qui a rapport à \sim (*crime*#I.a) = *criminologie*
- * Séquelle de \sim (*plaisie*#I) = *cicatrice*
- * ~ femelle(*mouton*) = *brebis*
- * ~ mâle(*mouton*) = *bélier*
- * Stratégie mise au point pour \sim (*bataille*#I.1) = *plan* [de \sim]
- * Substance alimentaire avec laquelle on fait \sim (*pain*) = *farine*
- Tâche blanche sur l'O.(*ongle*) = *albugo*
- * Taille de \sim (*enveloppe*#II) = *format*
- * Tel que Z est la France(*drapeau*) = *tricolore*
- * Telle qu'elle est sous la terre(*canalisation*) = *souterraine*
- * Telle qu'elle n'est pas professionnelle(*compétition*#2) = *amicale*
- [Qqch./Qqn.] tendre à montrer que \sim est justifiée(*accusation*#I.1) = *corroborer*
- [ART \sim]
- * Titre officiel de \sim (*avocat*#I) = *Maître*
- * Trace laissée à la surface de l'eau par \sim en déplacement(*navire*) = *sillage* [de ART \sim]
- Trait distinctif de \sim (*tigre*) = *pelage rayé*
- * Trajectoire de \sim (*comète*#I) = *course* [de ART \sim]
- Tranche de \sim enduite de N(*pain*#I) = *tartine*
- * Transcription de ce qui est dit pendant un \sim (*interrogatoire*) = *procès verbal* [de ART \sim]
- * Type de \sim (*gâteau*) = *clafoutis*
- * Unité de mesure de la capacité en Y de \sim (*clinique*#I) = *lit*
- * Unité de mesure pour la vitesse de \sim (*navire*) = *nœud*
- * Unité de mesure(*assemblée*#3) = *siège*
- Un seul(*utilisateur*) = *mono-utilisateur*
- Utilisé de façon traditionnelle et non professionnelle(*remède*#I) = \ulcorner *de grand-mère* \urcorner
- * Valable pour un an(*enveloppe*#III) = *annuelle*
- * Variété de \sim (*cerise*) = *griotte*
- * Vêtements de \sim (*facteur*(1)) = *uniforme*
- Vêtement typique(*curé*) = *soutane*
- * Vitesse d'écoulement de Y par \sim (*robinet*) = *débit*

Annexe II

Applications réalisées avec les FL

On distingue trois types d'applications de la LEC et des FL. Bien que ces dernières aient été conçues dans une optique de traduction automatique et, plus largement, de traitement automatique des langues, ce n'est pas le seul cadre d'application dans lequel elles entrent en jeu. Le modèle lexicologique de la LEC est un modèle lexical relationnel et décompositionnel. Son utilisation naturelle est donc la description des propriétés sémantiques et relationnelles des unités lexicales. La LEC a donné lieu à plusieurs types de ressources, que nous présentons dans cette section : des ressources générales monolingues (section II.1), bilingues ou multilingues (section II.2) ainsi qu'à des ressources spécialisées monolingues ou bilingues (section II.3). Par ailleurs, les FL peuvent constituer un bon cadre de référence pour l'enseignement du lexique. C'est là une application majeure qui ne semble pas avoir été prévue au moment de leur élaboration, mais pour laquelle elles se révèlent un outil novateur et efficace. De nombreux travaux ont déjà été réalisés dans cette voie. Nous présentons le LAF, Callex, le Dice, le Dafles et Alphalex dans la section II.4. Enfin, la finalité première des FL étant de participer à des tâches de traitement automatique, nous présentons, dans la section II.5, les recherches qui ont été menées en traduction automatique, génération de texte et production de résumés automatiques.

II.1 Ressources lexicales « générales »

Le Dictionnaire Explicatif et Combinatoire (DEC)

Le *Dictionnaire Explicatif et Combinatoire* (= DEC) a été développé pour le français [Mel'čuk et al., 1984, 1988, 1992, 1999], et pour le russe [Mel'čuk et Zholkovsky,

1984]. Quelques articles de dictionnaire ont été réalisés pour l'anglais [Meyer et Steele, 1990], le chinois [Nguyen, 2006], l'arabe [Bouchaddakh, 2008], le dene suline [Holden, à paraître] et le wolof [Bondeelle, à paraître]. Nous ne présentons ici que le DEC réalisé sur le français, c'est-à-dire le *Dictionnaire Explicatif et Combinatoire du Français Contemporain*. Il en existe quatre volumes décrivant 508 vocables. Le DEC est très souvent qualifié de dictionnaire de recherche formel. Il met directement en application les principes théoriques de la LEC. Pour chaque unité lexicale décrite, le DEC fournit trois types majeurs d'information. Nous prendrons, pour les illustrer, l'exemple de la lexie BOURRASQUE#II.1. Premièrement, le DEC fournit des informations de type sémantique sur les unités lexicales décrites par le biais d'une définition analytique constituée de composantes sémantiques élémentaires :

Bourrasque de X / sur Y = Violente tempête#II.1 soudaine et de courte durée causée par X qui perturbe le calme de Y [comme si c'était une bourrasque#I.1 qui affecte Y]

Le DEC fournit également des informations de type syntaxique en représentant la structure actantielle d'une unité lexicale et la façon dont ces actants sont exprimés. Un tableau de régime (cf. figure II.1) spécifiant les correspondances entre actants sémantiques et syntaxiques profonds, la façon dont les actants syntaxiques profonds peuvent être exprimés en surface ainsi que les contraintes de compatibilité dans l'expression des actants.

Régime

X = 1	Y = 2
1. de N	1. sur N
2. A	

- 1) C1 + C2 : impossible
 C1 : *une bourrasque d'inflation <inflationniste, boursière>*
 C2 : *une bourrasque sur le système monétaire européen*
 :

Figure II.1 – Tableau de régime de BOURRASQUE#II.1 dans le DEC

Enfin, le DEC présente l'ensemble des relations lexico-sémantiques, c'est-à-dire les dérivations sémantiques et les collocations d'une unité lexicale. Ces relations sont modélisées au moyen des FL (cf. figure II.2).

Fonctions lexicales

Syn _C	:	<i>tempête</i> #II.1
Syn _∩	:	<i>tourbillon</i> #II.2 , <i>ouragan</i> #II.1
Adv ₁ Culm	:	<i>au plus fort</i> [de ART ~]
Magn	:	<i>forte</i> antépos
Adv ₂	:	<i>dans</i> [ART ~]
Oper ₂	:	<i>connaître, essayer, subir</i> [ART ~]; <i>être pris</i> [dans ART ~]
ProxFunc ₀	:	<i>s'annoncer</i>
Func ₂	:	<i>souffler</i> [sur N]
ProxFunc ₂	:	<i>attendre</i> #III.2 [N]
essayer de	:	<i>faire face</i> [à ART ~]; <i>résister, tenir tête</i> [à ART ~]
Liqu ₂ Real ₂	:	

Exemples

La direction de cette entreprise a rapidement tiré des leçons de cette bourrasque boursière. Face à la bourrasque essuyée par la gauche, le ministre n'a pu retourner la situation en sa faveur. En une nuit, le 16 décembre 1992, jour de forte bourrasque sur le système monétaire européen, sa fortune s'est enrichie de un milliard de dollars...

Figure II.2 – Fonctions lexicales et exemples pour BOURRASQUE#II.1 dans le DEC

Le DiCo

Le DiCo a fait l'objet d'une présentation au chapitre 2 (section 2.2.1, p. 26) de cette thèse. Nous renvoyons donc le lecteur à la section concernée.

II.2 Ressources lexicales bilingues et multilingues

La Théorie Sens-Texte (et son module lexical, la LEC) est une théorie indépendante d'une langue particulière. Il est donc possible d'utiliser la même structure de représentation et les mêmes outils pour décrire des langues différentes. C'est là l'objectif premier des FL qui, rappelons-le, ont été formalisées pour servir les besoins de la traduction automatique. Les FL standard sont posées comme universelles. Nous présentons deux ressources bilingues et multilingues utilisant les FL : une base de données conçue à partir du Robert & Collins et le projet Papillon. Notons également que Bouchaddakh a développé un échantillon de base de données lexicale bilingue français-arabe.

La base de données lexicales du Robert & Collins en fonctions lexicales

Fontenelle [1997b] a réalisé un travail de conversion du dictionnaire bilingue Robert & Collins [Atkins et Duval, 1978] vers une base de données lexico-sémantique. À l'instar de la majorité des dictionnaires bilingues, le Robert & Collins met l'accent sur la combinatoire des unités lexicales et fournit un grand nombre de collocations. L'objectif principal est la construction d'un réseau sémantique de l'information syntagmatique et paradigmaticque contenue dans le dictionnaire bilingue. Fontenelle constate que beaucoup de dictionnaires se révèlent très peu performants pour mettre en valeur et différencier les collocations et choisit d'utiliser le formalisme des FL pour les encoder et les organiser. Ce faisant, il effectue une première application de masse des FL. Tester les FL sur l'ensemble du Robert & Collins représente une tâche considérable jusqu'alors jamais réalisée qui a permis de confronter ce mode de représentation à un ensemble de données et d'en vérifier la validité pratique.

The general conclusions I draw concerning the imprecise nature of some functions and the evaluation of the material would never have been possible, however, if only a fragment of the dictionary had been analyzed in detail. Exploiting the whole dictionary and focussing on such a large

section of the lexicon had never been attempted and the potential applications, whether purely linguistic or pedagogical, could only have been hinted at if I had decided to focus on a few dozen items and their semantic environment instead of taking into account the whole metalinguistic apparatus of the dictionary. [Fontenelle, 1997b : 102]

L'assignation manuelle des FL aux collocations est une tâche fastidieuse, qui bien heureusement peut être facilitée par certains types de métalangage mettant en valeur des régularités. C'est le cas des micro-définitions, qui dans le Robert & Collins, aident l'utilisateur à choisir la traduction appropriée. Certaines de ces micro-définitions sont récurrentes et correspondent directement à des FL. Ainsi, il est possible, par exemple, d'inférer la FL *Mult* (→ ensemble régulier) à partir des micro-définitions *body of*, *group of*, *set of*, *series of*, *list of*, *cluster of* et *number of* ; de même que la FL *Sing* (→ unité régulière) peut être inférée des formules *piece of*, *member of*, *flash of* et *act of* [Fontenelle, 1997b : 149-153]. À la lumière de ces analogies, les FL peuvent donc être vues comme des définitions très succinctes des relations lexicales.

La structure de la base de données, conçue par Jacques Jansen, dispose chaque type d'information dans des tables séparées. Elle permet une grande potentialité de requêtes : on a la possibilité soit d'interroger la base de données de façon classique, par exemple, faire une requête sur l'unité *brake* pour obtenir les résultats suivants :

$\text{part}(\textit{brake} / \textit{frein}) = \textit{drag}$ (sabot <m> de frein)
 $\text{degrad}(\textit{brake} / \textit{frein}) = \textit{drag}$ (frotter)
 $\text{finfact0}(\textit{brake} / \textit{frein}) = \textit{fail}$ (lâcher)
 $\text{magnfact0}(\textit{brake} / \textit{frein}) = \textit{grip}$ (mordre)
 ...

soit d'interroger la base selon des requêtes plus précises : la base, le collocatif, l'objet-typique, une FL, etc. Il est ainsi possible d'extraire les verbes prenant l'unité lexicale *law* comme objet direct et donc le sens correspond à la FL *Liqu* (→ causer la fin de qqch). Cette requête aura pour résultat les relations suivantes :

abolish (vt) : \sim law \sim => abroger (loi,liqu)
annul (vt) : \sim law \sim => abroger (loi,liqu)
do away with (vt fus) : \sim law \sim => supprimer (loi,liqu)
repeal (vt) : \sim law \sim => abroger (loi,liqu)
rescind (vt) : \sim law \sim => abroger (loi,liqu)
revoque (vt) : \sim law \sim => rapporter (loi,liqu)

Fontenelle [1997b : 254-271] montre qu'une telle base de données peut avoir des applications dans le domaine de l'apprentissage des langues assisté par ordinateur. L'utilisation des FL a effectivement donné lieu à de nombreuses applications dans ce domaine. Nous ne détaillons pas les propositions suggérées par Fontenelle, mais renvoyons le lecteur à la section II.4 consacrée au développement d'outils pédagogiques pour l'apprentissage du lexique.

Le projet Papillon

Nous présentons une seconde ressource lexicale utilisant les FL : le projet Papillon, développé par Mathieu Mangeot et Gilles Sérasset¹. Il s'agit d'un projet collaboratif visant le développement d'une base lexicale multilingue comprenant entre autres l'allemand, l'anglais, le français, le japonais, le malais, le lao, le thaï, le vietnamien et le chinois. Il existe peu de dictionnaires (pour humains et machines) entre le français et de nombreuses langues asiatiques. L'objectif, selon Mangeot [2002 : 64], est donc de réduire les coûts de création de cette connaissance lexicale en adoptant une méthode inspirée des projets « open-source » et fonctionnant grâce à des contributeurs bénévoles.

Ce dictionnaire étant multilingue, sa macrostructure se présente sous la forme d'une structure pivot : chaque entrée des différents dictionnaires monolingues est reliée à ses traductions dans les autres langues via un lien interlingue dans la structure pivot. La structure pivot est fondée sur la notion d'acceptations interlingues appelées *axies* [Mangeot et Kuroda, 2003 : 34]. Mangeot [2006] donne l'exemple du vocable

¹Le projet papillon est présenté à l'adresse suivante : <http://www.papillon-dictionary.org/>.

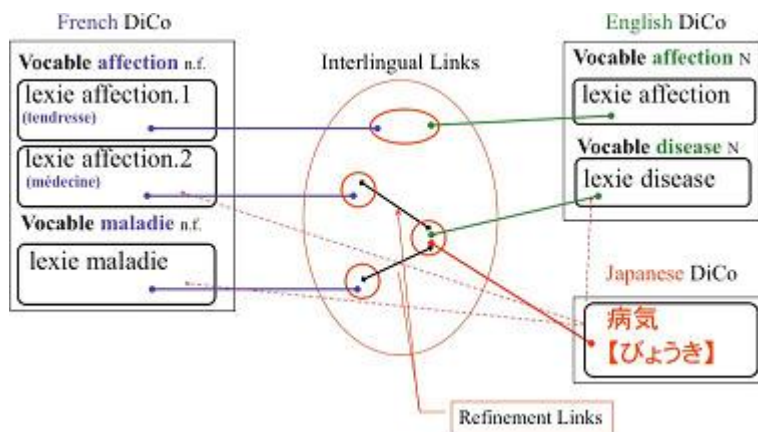


Figure II.3 – Macrostructure du dictionnaire Papillon

AFFECTION en français qui comporte deux sens : ‘tendresse’ et ‘maladie’. Ce vocable est donc séparé en deux lexies dans le dictionnaire monolingue français et relié à deux axes interlingues dans le dictionnaire pivot, comme l’illustre la figure II.3.

Sa microstructure, quant à elle, repose sur les principes de la LEC et est relativement conforme à celle du DiCo : c’est-à-dire que pour chaque unité lexicale sont donnés une étiquette sémantique, une structure argumentale, le régime de la lexie ainsi que les liens paradigmatiques et syntagmatiques privilégiés encodés au moyen de FL. Beaucoup de données sont récupérées à partir de ressources existantes et sont intégrées dans le dictionnaire Papillon. On a ainsi la possibilité de naviguer dans les dictionnaires monolingues ou de trouver les traductions possibles dans plusieurs articles. Par exemple, pour le vocable RÊVER, plusieurs articles de dictionnaire sont proposés dont celui de la base ELRA ² et celui du dictionnaire trilingue FeM [Gut et al., 1996] que nous reproduisons dans la figure II.4 (p. xxxvi). Le projet Papillon est un projet ambitieux qui repose en grande partie sur le principe de la collaboration multiple. Pour l’instant peu de données sont présentées et la structure globale d’interaction interlangue n’est pas encore très développée.

²<http://www.icp.inpg.fr/ELRA>

ELRA : (dictionnaire allemand - anglais - espagnol - français - italien)

rêver v (rêver pendant le sommeil V) : *allemand* träumen v - *anglais* dream v - *italien* sognare v - *espagnol* soñar v

rêver v_intr (imaginer, se représenter dans l'esprit) V : *allemand* träumen v_intr - *anglais* dream v_intr - *italien* sognare v_intr - *espagnol* soñar despierto v_intr

FeM : (dictionnaire français - anglais - malais)

v.intr.

faire des rêves : *anglais* dream - *malais* bermimpi

délirer, plaisanter : *malais* mengigau melamun

imaginer : *malais* membayangkan mengkhayalkan

souhaiter : *malais* berangan-angan

Je rêve de m'acheter une maison au bord de la mer - *malais* saya berangan-angan hendak membeli sebuah rumah di tepi laut

v.tr.

anglais dream - *malais* memimpikan mengidam-idamkan

Figure II.4 – Article RÊVER des dictionnaires ELRA et FeM

II.3 Ressources lexicales spécialisées

Les principes de la LEC ont également été mis en application pour la création de bases de données terminologiques. Nous présentons, dans cette section, la façon particulière dont l'approche lexico-sémantique perçoit et décrit les vocabulaires spécialisés. L'avantage de l'utilisation des FL sera présenté par la suite à travers l'exemple d'une base de données terminologique : le DiCoInfo.

Une approche lexicologique de la terminologie

L'approche classique de la terminologie, théorisée par Wüster au début du siècle dernier, a donné lieu à des courants ontologistes, encyclopédiques et conceptuels des vocabulaires spécialisés. Le terme y est vu comme un concept désignant un référent. Toute question de signifié est écartée et la polysémie est présumée inexistante au profit de la bi-univocité, c'est-à-dire que pour un terme n'existe qu'une notion et qu'une notion ne peut être désignée que par un seul terme. En somme, le terme n'est pas considéré comme étant un signe linguistique comme les autres.

Or, vers la fin du siècle dernier, des lacunes dans la méthodologie de description des unités terminologiques, centrée presque exclusivement sur des représentations ontologiques, ont conduit certains terminologues à s'intéresser de près aux modèles de description linguistique. Ceux-ci posent un nouveau point de vue sur le terme, qui n'est plus vu comme une simple étiquette désignant un concept, mais comme un signe linguistique au sens « réglé » [Bouveret, 1997]. Des courants terminologiques davantage centrés sur les termes en contexte ont fait leur apparition comme la *terminologie textuelle* proposée par Bourigault et Slodzian [1999], qui prône une étude des termes spécialisés dans leur contextes d'emploi. Dans une autre perspective, les partisans d'une *socioterminologie* [Gaudin, 2003] s'opposent également au courant classique en mettant en évidence les variations sociologiques d'emploi d'un terme et les usages oraux. Enfin, on peut noter, l'approche *lexico-sémantique* du terme que l'on pourrait également qualifier de *lexicographique* et qui met en avant le phénomène de polysémie, de synonymie ainsi que les relations sémantico-lexicales entre les termes d'un domaine [L'Homme, 2004]. Plusieurs approches terminologiques se sont tournées vers des modèles lexicologiques pour l'analyse des vocabulaires spécialisés ainsi que vers les méthodes lexicographiques pour leur description. À tel point que le terme *terminographie* se trouve parfois supplanté par celui de *lexicographie spécialisée* et que la distinction d'avec la lexicographie dite générale pourrait être remise en question. Les approches lexicologiques et lexicographiques de la terminologie s'intéressent aux propriétés linguistiques des termes en usage et s'attachent à décrire leur combinatoire syntaxique et lexicale. Les ressources descriptives de ce type sont donc radicalement différentes des ressources plus traditionnelles³ et comportent de nombreuses similitudes avec la lexicologie et la lexicographie. C'est pourquoi le choix d'un formalisme de description des relations lexicales semble tout à fait adapté aux vocabulaires spécialisés. Nous présentons un dictionnaire spécialisé développé à l'OLST qui suit une orientation lexicographique suivant le modèle théorique de la LEC et propose une formalisation des relations lexico-sémantiques entre termes au moyen des fonctions lexicales : le DiCoInfo.

³Se reporter à [L'Homme, 2004] pour une analyse comparative des principes méthodologiques de ces deux approches.

Le DiCoInfo

Le DiCoInfo est une ressource terminologique portant sur le vocabulaire de l’informatique, développée par Marie-Claude L’homme à l’OLST depuis 2002 [L’Homme, 2007]⁴. Vis-à-vis des autres ressources terminologiques, le DiCoInfo tire son originalité de son approche linguistique du vocabulaire spécialisé, plus particulièrement une approche lexico-sémantique directement inspirée du DiCo. La méthodologie d’élaboration du DiCoInfo est orientée corpus. Les termes décrits et ceux qui leur sont reliés sont extraits à l’aide d’outils d’extraction [Lemay et al., 2005] et validés selon des critères lexico-sémantiques [L’Homme, 2005]. Les descriptions reposent en premier lieu sur l’observation du fonctionnement des termes dans un corpus spécialisé. Un corpus contenant plus d’un million de mots a été assemblé à cette fin. Les entrées de la base de données comportent quatre types d’information :

1. le nom du terme ainsi que sa partie du discours et son numéro d’acception, par exemple : **Piratage 2a, n. m.**
2. la structure actantielle énumère les actants sémantiques du terme et décrit ainsi que leur rôle par rapport au terme décrit. La structure actantielle de **piratage 2a** est la suivante :

: ~ de PATIENT{données 1 ; logiciel 1} par AGENT{utilisateur 1}.

Contrairement au DiCo, les actants ne sont pas identifiés par des numéros mais par des rôles sémantiques (ex : patient, agent, destination, support, lieu, ...), dont l’objectif est de typer très généralement le rôle des actants par rapport au terme décrit. Une réalisation linguistique typique de ces actants est présentée dans un tableau en dessous de la forme propositionnelle :

AGENT	PATIENT
internaute 1, utilisateur 1	application 1, données 1, fichier 1, jeu 1, licence 1, licence 1, logiciel 1, musique, système d’exploitation 1

⁴La base de données est consultable à l’adresse suivante : <http://olst.ling.umontreal.ca/dicoinfo/>. Elle comprend environ 1 000 termes décrits et 20 000 termes reliés.

3. Une rubrique d'exemples vient illustrer l'emploi du terme décrit dans des exemples en contexte tirés directement du corpus.

Contexte(s) :	Source(s) :
<i>C'est vrai aussi que la démocratisation du graveur de CD a relancé la mode du piratage « innocent » qui consiste à copier tout et n'importe quoi pour des proches, sans idée d'en tirer profit.</i>	ANTHO
<i>C'est la nouveauté - avec Office XP - pour combattre le piratage informatique</i>	INSTXP
<i>Qu'est-ce que le piratage de logiciels ? Le piratage de logiciels est une usurpation des droits d'auteurs.</i>	macromedia.com

4. Puis vient la rubrique « **termes reliés** » qui présente la combinatoire des termes. Il s'agit des relations paradigmatiques et syntagmatiques entre termes. Les relations sont décrites au moyen de formules explicites normalisées qui correspondent *grosso modo* aux vulgarisations de FL. Un encodage plus formel à l'aide des FL est réalisé en parallèle mais reste réservé aux utilisateurs initiés.

Explication de la relation sémantique :	Terme(s) relié(s) :
Synonyme	~ informatique
Terme plus général	copie 1a
Verbe	pirater 2
Un patient qui a fait l'objet d'un « mot clé »	pirate 3
Une mesure qui vise à éliminer ou à réduire le « mot clé »	antipiratage 1
Qui se produit lors d'une connexion	~ en ligne 1
L'agent fait « le mot clé »	se livrer au ~
L'agent fait le « mot clé »	faire du ~
Qqn ou qqch. essaie de faire en sorte que le « mot clé » ne se produise pas	lutter 1 contre le ~
→ NOM	lutte 1a contre le ~
Qqn ou qqch. fait en sorte qu'un « mot clé » ne se produise pas sur le patient	protéger 2 ... contre le ~
→ NOM	protection 2a contre le ~

Le DiCoInfo propose une solution originale pour rendre compte du rôle sémantique (ou thématique) des actants d'une lexie. En effet, les numéros d'actants habituellement donnés dans les FL ne permettent pas d'indiquer le rôle sémantique des actants dans un événement donné. Par exemple, dans les trois relations suivantes, les valeurs de la FL S_3 jouent trois rôles différents vis-à-vis de leurs mots-clés :

$S_3(usine) = compagnie, propriétaire$ [de ART ~];

$S_3(timbre(1)\#1 \text{ 'moyen de paiement'}) = colis, enveloppe, lettre$;

$S_3(accusation\#I.1) = objet, motif$ [de ART ~].

De même, il existe des cas où le deuxième ou troisième actant d'une lexie est également l'instrument servant à la réaliser, par exemple :

$S_2/S_{instr}(canard\#II) = \lceil instrument \ à \ vent \rceil, clarinette, saxophone, etc.$

L'emploi des rôles 'actant', 'patient', 'instrument', permet de préciser le rôle des actants, qui sont vus plus largement comme des participants, et d'éviter toute ambiguïté dans l'interprétation des FL.

Le DiCoInfo offre un modèle aisément reproductible pour la structuration de bases de données terminologiques. Notons que ce modèle n'est pas incompatible avec un dictionnaire terminologique plus traditionnel. Un exemple de cette possible mixité est donné par Le *Dictionnaire Analytique de la Distribution*⁵. Il s'agit de la version informatisée du dictionnaire papier de Dancette et Réthoré [2000], dictionnaire spécialisé de nature encyclopédique qui a intégré la description de certains liens entre termes en utilisant les FL. Plusieurs autres chercheurs ont entrepris la modélisation des relations terminologiques au moyen des FL. Ainsi Frawley [1998] propose des exemples d'entrée de type DEC pour des termes de géologie, zoologie et physique ; Bouveret [2006] a encodé une partie des relations terminologiques d'un dictionnaire de biologie moléculaire (Biolex) à l'aide des FL et Faber et Sánchez [2001] propose de combiner le modèle lexématique fonctionnel et le modèle Sens-Texte avec les FL pour établir une méthodologie de représentation du vocabulaire des domaines spécialisés. On note également l'existence de ressources terminologiques inspirées de la LEC comme le DAFA [Binon et al., 2000] qui toutefois ne retient pas l'encodage « brut » des FL pour l'identification des relations terminologiques. Il y a donc fort à penser que ces travaux terminologiques fondés sur une méthode de description lexicographique et utilisant les FL comme outil systématique pour la description des liens terminologiques constitueront un moteur dans le domaine.

⁵La ressource est accessible en ligne à l'adresse suivante : <http://olst.ling.umontreal.ca/dad/>.

II.4 Ressources lexicales à visée pédagogique et outils d'apprentissage du lexique

Nous avons présenté, dans le deuxième chapitre de cette thèse, différents modèles relationnels du lexique. Le principe commun à ces modèles est de mettre l'accent sur l'interconnexion des éléments lexicaux. Une base de données lexicale reposant sur un tel modèle constitue donc un outil relativement cadré pour parcourir le lexique. Grimes [1996 : 104] considère les FL comme un outil heuristique pour l'apprentissage des langues et montre comment elles peuvent servir d'outil méthodique pour parcourir le lexique en prenant n'importe quelle unité lexicale comme point de départ.

That is why we can expect to cover the information we want by picking any lexeme whatever as starting point, finding out about its lexical correlates and the ways they relate to the starting lexeme, and following those paths until there are no more left to follow. The number of path to investigate is enormous, to be sure; but we always know where we are. No strategy can reduce the size of the task because the lexicon is innately complex and huge; but this way we can be sure of reasonably adequate coverage. It does not matter where we start : lexical structure is a clear case where “tout se tient”.

Nous présentons, dans cette section, les ressources lexicales à visée pédagogique développées dans le cadre de la LEC, soit, le LAF pour le français et le DICE pour l'espagnol. Dans le domaine de l'Apprentissage Assisté par Ordinateur (AAO), plusieurs projets sont développés autour des FL et de la LEC. Nous présentons les applications CALLex ainsi que le DAFLES et son environnement d'apprentissage Alfalex.

Le LAF

Le *Lexique Actif du Français* (= LAF) [Mel'čuk et Polguère, 2007] a été développé conjointement au DiCo à l'OLST. Les fiches du DiCo sont transposées vers des

article du LAF, comme nous l'avons montré à la figure 3.3 (p. 80) avec la lexie ABOIEMENT. Les relations lexicales y sont représentées au moyen d'un encodage vulgarisé des FL (présenté à la section 3.3.1). Associées au LAF, un certain nombre de fiches pédagogiques sont créées afin de guider le parcours du lexique ou comme support d'activités pour l'enseignement du lexique⁶ [Polguère, 2007].

Le DICE

Pour l'espagnol, il existe également un projet fondé sur la LEC : le *Diccionario de colocaciones del español* (DICE). Il s'agit d'une base de données lexicale développée à l'université de la Corogne sous la direction de Margarita Alonso Ramos⁷. Pour le moment, la nomenclature du DICE se limite aux unités lexicales des noms de sentiment. La base de données suit rigoureusement les principes du DEC et du DiCo tant pour le dégroupement de la polysémie que pour la représentation des collocations. Les collocations sont organisées en fonction de leur patrons syntaxiques : par exemple, pour l'unité lexicale CELOS (= 'jalousie'), les collocations sont regroupées selon les types suivants : attributs des participants (ex : **A**₁ *celoso*), CELOS + adjectif (ex : **Magn** : *grandes*), verbe + CELOS (ex : **IncepOper**₁ : *coger*), CELOS + verbe (ex : **Fact**₁ : *carcomer*) et nom + de + CELOS (ex : **Germ** : *germen*). Les vulgarisations des FL sont préférées à l'encodage brut des FL que l'on peut toutefois faire apparaître en activant un lien.

Le dictionnaire s'accompagne d'exercices pour l'apprentissage du vocabulaire espagnol. Il s'agit premièrement d'exercices de production qui s'articulent autour de cinq questions :

1. trouvez le bon collocatif parmi des propositions sous forme de questions à choix multiple ;
2. trouvez le bon collocatif sans propositions ;
3. reliez plusieurs collocatifs verbaux à leurs bases ;
4. trouvez l'intrus parmi les collocatifs proposés ;

⁶Ces fiches sont consultables à l'adresse suivante : <http://olst.ling.umontreal.ca/laf/>.

⁷La base de données est disponible à l'adresse suivante : <http://www.dicesp.com>.

5. reliez les collocatifs adjectivaux aux bases appropriées.

Les exercices de compréhension s'organisent autour de quatre questions portant sur le sens des unités lexicales et notamment sur le sens de certains collocatifs. Les réponses aux questions sont comptabilisées sur 100 et permettent ainsi une évaluation de l'utilisateur. On remarque que les FL ne sont pas mises en avant mais restent sous-jacente à la génération des exercices.

CALLex

Le système CALLex (Computer-Aided Learning of Lexical Functions) a été développé à l'IITPTRAS (Institute for Information Transmission Problems of the Russian Academy of Sciences) depuis les années 90s [Diachenko, 2006]. Il repose sur deux concepts de la théorie Sens-Texte : les FL et la théorie de la décomposition lexicale développée par Apresjan [2000]. Le projet a été développé pour l'anglais, le russe, l'espagnol ainsi que pour l'allemand dans une moindre mesure. Le système s'adresse aux apprenants de langue étrangère ainsi qu'aux locuteurs natifs qui souhaitent améliorer leur vocabulaire. Fondamentalement, CALLex est constitué de deux dictionnaires combinatoires de l'anglais et du russe convertis en deux bases de données s'élevant à plus de 65 000 entrées chacune. Chaque entrée contient les informations suivantes : partie du discours, traduction standard dans l'autre langue de travail, caractéristiques syntaxiques, caractéristiques sémantiques, schéma de régime, fonctions lexicales ainsi que l'ensemble des règles spécifiques à cette unité lexicale.

Plusieurs jeux sont créés directement à partir des FL [Diachenko, 2006 : 540].

– Jeu 1. FL

Ce jeu utilise le dictionnaire de FL. L'utilisateur doit remplir des valeurs d'une FL concrète pour chaque argument. Par exemple, la valeur de CausFact_0 (\rightarrow causer que \sim fonctionne) pour les lexies CLOCK et IMAGINATION :

$\text{CausFact}_0(\text{clock}) = \text{wind} \langle \text{wind up} \rangle, \text{start}$;

$\text{CausFact}_0(\text{imagination}) = \text{fire} (\text{somebody's imagination})$.

– **Jeu 2. Définition**

Le jeu consiste à proposer une définition analytique d'un lexème et à faire deviner à l'utilisateur de quel mot il s'agit. Par exemple, le système peut proposer la définition suivante : *A more favourable position of a person P1 in a certain situation as compared with the position of another participant P2*, à quoi l'utilisateur devrait répondre qu'il s'agit de la lexie ADVANTAGE.

– **Jeu 3. Traduction**

Le système montre à l'utilisateur un mot-vedette et sa définition analytique et l'utilisateur doit fournir une traduction de ce mot-vedette.

– **Jeu 4. Mot**

L'utilisateur joue avec l'ensemble des FL et des traductions présentes pour une entrée. Pendant le jeu, on lui montre la définition de la lexie ainsi qu'une définition et des exemples de chaque FL. Ce jeu permet à l'utilisateur d'améliorer son habileté à employer les unités lexicales appropriées.

– **Jeu 5. Mots avec indices**

Ce jeu est orienté sur le mode bilingue et fondé sur des paires de lexies : le système présente des questions dans la langue la mieux connue de l'utilisateur et celui-ci doit donner une traduction dans la langue qu'il apprend.

La version espagnole de CALLex : CALLEX-ESP [Boguslavsky et al., 2006] propose des exercices selon un angle légèrement différent. Les jeux prennent la forme d'exercices qui utilisent les gloses des fonctions lexicales. Par exemple : **AntiVer** va être vulgarisée par \rightarrow **no auténtico**.

– *cuadro* (\rightarrow **no auténtico**) = *cuadro* ... (réponse : *falsificado*);

– *acusación* (\rightarrow **no auténtica**) = *acusación* ... (réponse : *gratuita*).

Le DAFLES et Alfalex

Le *Dictionnaire d'Apprentissage du Français Langue Etrangère ou Seconde* (= DAFLES) est une base de données lexicale développée à l'Université de Louvain par Serge Verlinde, Jean Binon et Thierry Selva au sein du Groupe de Recherche

en Lexicographie Pédagogique (GRELEP)⁸. La méthodologie d'élaboration de la base de données repose en très grande partie sur les principes de la LEC. On n'observe toutefois pas de formalisation des relations lexicales. Les auteurs n'ont gardé des FL que le concept et rejeté le formalisme jugé trop aride et trop complexe. Le DAFLES est intégré dans un environnement multifonctionnel pour l'apprentissage du lexique appelé *Base Lexicale du Français* [Verlinde et al., 2006]. La BLF offre plusieurs interfaces de consultation de la base de données correspondant à plusieurs types d'accès aux unités lexicales.

L'interface « Recherche dans la base » offre un certain nombre de critères correspondant aux propriétés sémantiques, syntaxiques et combinatoires des unités lexicales. Ils permettent d'interroger la base sur :

- les catégories grammaticales ou parties du discours des unités lexicales ;
- les domaines de connaissance des unités lexicales ;
- les fonctions lexicales : la notion de fonction lexicale est très simplifiée et réduite à quinze relations non formalisées ;
- la fréquence des unités lexicales ;
- la compositionnalité du sens des éléments de l'unité lexicale ;
- les types de composition de mot : expression, locution, proverbe, etc. ;
- recherche directe dans l'inventaire des collocations ou compositions ;
- recherche par sens exprimé : Les différentes catégories sémantiques définies par les fonctions lexicales couplées aux parties du discours possible : exemple, un nom exprimant le début d'une action, etc. ;
- le genre et le type de terminaison des unités lexicales ;
- les modèles morphologiques des noms et des adjectifs ;
- la conjugaison des verbes ;
- les constructions syntaxiques possibles pour les verbes et les noms ;
- la place de l'adjectif : antéposé ou postposé ;

⁸La base de données peut être consultée à l'adresse suivante : www.kuleuven.be/dafles/

- l’exploration des définitions : permet l’accès aux définitions contenant un ou plusieurs mots précisés.

Couplé au DAFLES, le projet Alfalex⁹ est une ressource fournissant un certain nombre d’exercices pour l’apprentissage du vocabulaire. Alfalex se sert du DAFLES et d’un corpus pour générer des exercices portant sur les propriétés suivantes des unités lexicales : la formation des mots (morphologie, conjugaison, dérivation), le genre des noms, la combinatoire des mots, les champs sémantiques ainsi que la traduction du français vers le néerlandais et inversement. L’environnement d’apprentissage permet de situer le niveau de français des utilisateurs : il calcule, dans un premier temps, le score pour chaque exercice, garde en mémoire les erreurs commises et propose à nouveau des exercices sur ces erreurs. Il offre un corrigé pour chaque exercice à l’aide des informations lexicales du DAFLES et soumet des exercices adaptés au niveau évalué des apprenants.

II.5 Applications des FL en Traitement Automatique des Langues

Après la revue des ressources créées sur les principes de la LEC, nécessaires pour construire un inventaire formalisé et exploitable des relations lexicales, nous présentons, dans cette section, trois applications en traitement automatique des langues mettant en œuvre les FL, soit en synthèse de paraphrases, traduction automatique et génération automatique de résumés. Notons que nous mettons plus particulièrement l’accent sur les travaux dans la description desquels le rôle des FL est explicite.

⁹Le projet Alfalex (= Actieve leeromgeving Frans voor anderstaligen - lexicon) peut être consulté à l’adresse suivante <http://www.kuleuven.be/alfalex/>.

Synthèse de paraphrases et génération de texte

Il existe un certain nombre de systèmes informatiques de génération et synthèse de phrases fondés sur la théorie Sens-Texte. On peut citer *GOSSIP* [Iordanskaja et al., 1988], *LFS* [Iordanskaja et al., 1992], *FoG* [Goldberg et al., 1994], le modèle de reformulation de Nasr [1996], *RealPro* [Lavoie et Rambow, 1997], *AlethGen* [Coch, 1996], *MATE* [Bohnet et al., 2000] ou encore *Sentence Garden* [Lareau, 2002]. Nous en présentons très brièvement quelques-uns ci-dessous.

Nasr [1996] a proposé un modèle de reformulation automatique de phrases fondé sur la théorie Sens-Texte pour le domaine des langues contrôlées du milieu documentaire industriel et plus particulièrement des textes de maintenance aéronautique. Les langues contrôlées imposent aux textes techniques des contraintes diverses dont des contraintes syntaxiques et lexicales. L'objectif d'A. Nasr est d'utiliser les capacités paraphrastiques de la théorie pour réaliser la reformulation des structures interdites par des structures autorisées. Le modèle proposé comporte trois phases principales : premièrement, une phase d'analyse à l'issue de laquelle une représentation syntaxique profonde de la phrase à formuler est construite, deuxièmement, une phase de transformation consistant à modifier la phrase d'origine par application de règles de paraphrasage, et finalement, une dernière phase de régénération qui consiste à synthétiser une nouvelle phrase en s'assurant de ne pas réintroduire de structures interdites.

Le système *Realpro* [Lavoie et Rambow, 1997] est un synthétiseur pour des systèmes de génération de texte fondé sur la syntaxe de dépendance de la théorie Sens-Texte. Il prend en entrée des structures syntaxiques profondes et suit étape par étape les différents modules de la théorie Sens-Texte¹⁰ soit les modules syntaxiques (profond et de surface) et morphologiques (profond et de surface). Par la suite, un composant graphique insère la ponctuation. Le système se réfère à une grammaire ainsi qu'à un lexique pour intervenir aux différentes étapes. Il est principalement développé

¹⁰Se reporter aux paragraphes 3.2.1 (p. 69) ou 5.3.2 (p. 193) pour une définition des différents modules de la Théorie Sens-Texte.

pour l'anglais.

MATE [Bohnet et al., 2000] est un générateur automatique de phrases développé au département d'informatique de Stuttgart par Bernd Bohnet, Andreas Langiahr et Leo Wanner. Il repose également sur le modèle linguistique de la Théorie Sens-Texte. L'environnement de développement prend en entrée une structure conceptuelle dérivée des données présentes dans une base de données lexicale de type DEC. Le processus de génération consiste en une série de mises en correspondance des différentes structures liées aux niveaux de la TST, déjà mentionnés plus haut. Une fois que la structure morphologique de surface est atteinte, le résultat prend la forme d'une chaîne linéaire de mots-forme. Le module central de l'environnement de développement est un compilateur mettant en correspondance deux à deux les structures des différents niveaux de représentation. Un débogueur ainsi qu'un inspecteur sont réalisés pour venir en aide à un utilisateur lors de l'analyse du traitement des structures. L'utilisateur peut ainsi interagir avec le compilateur.

Lareau [2002] a réalisé un synthétiseur automatique de texte, *Sentence Garden*, permettant de générer, pour une phrase donnée, l'ensemble de ses paraphrases possibles. Il se distingue des autres modèles par le fait qu'il cherche à générer, à partir d'une représentation sémantique, l'ensemble des réalisations possibles d'un message et non une seule. Cet outil vise également à vérifier la validité des modèles linguistiques Sens-Texte pour la génération de phrases.

Traduction automatique : le projet ETAP-3

ETAP-3 est un processeur multifonctionnel et multilingue développé par le laboratoire de linguistique computationnelle de l'Académie Russe des Sciences pour diverses applications en traitement automatique des langues naturelles¹¹. Les prin-

¹¹Ce projet et les applications reliées sont consultables à l'adresse suivante : http://proling.iitp.ru/etap/index_e.html.

cipales réalisations d'ETAP-3 sont la traduction automatique, la communication multilingue par le biais d'une interlangue et le paraphrasage. Notons que l'environnement d'apprentissage du lexique CALLex présenté à la section II.4 est intégré dans ETAP-3. Le principal système développé au sein du projet est un système de traduction automatique bidirectionnel anglais-russe et russe-anglais. Quelques essais ont également été proposés sur le français, l'allemand, l'espagnol et le coréen. ETAP-3 est constitué de plusieurs ensembles de règles et de dictionnaires. Toutes les règles sont subdivisées en trois types principaux :

1. les règles générales, qui s'appliquent à toutes les phrases en cours de traitement ;
2. les règles spécifiques aux différentes classes, qui concernent des groupes d'unités lexicales et qui sont appelées par leur nom dans les entrées de dictionnaires des dites unités lexicales ;
3. des règles spécifiques aux unités lexicales, caractéristiques des unités lexicales individuelles et qui sont stockées directement dans leur entrée de dictionnaire.

Dans le système de paraphrasage, les règles sont appliquées dans l'ordre suivant : (a) analyse morphologique ; (b) analyse syntaxique ; (c) interprétation par FL de la structure syntaxique ; (d) normalisation de la structure interprétée ; (e) règles de canonisation ; (f) règles de paraphrasage ; (g) règles de réinterprétation et de correction ; (h) introduction du matériau lexical comme les prépositions spécifiques requises par les FL concernées ; (i) synthèse syntaxique ; (j) synthèse morphologique.

Apresjan et al. [2003] présentent deux avantages des FL dans les étapes du processus de traduction automatique :

1. Outil de désambiguïsation pour l'analyse syntaxique

Les FL peuvent servir d'outil pour lever les ambiguïtés syntaxiques et lexicales grâce à l'analyse des éléments par les FL. Par exemple, le syntagme *support of the parliament*, peut être interprété de deux façons : dans un premier cas, *parliament* est agentif et peut être paraphrasé par *support given by*

the parliament ; dans la deuxième interprétation, *parliament* est objet et sera paraphrasé par *support given to the parliament*. Les FL de verbes supports sont très utiles pour lever les ambiguïtés car elles imposent des fortes restrictions sur le comportement syntaxique de leurs mot-clé dans les textes. Ainsi, dans la phrase suivante : *The president spoke in support of the parliament*. Le verbe *to speak* est une forme de Oper_1 du nom *support*. Oper_1 donne des verbes dont le sujet grammatical correspond au premier actant du mot-clé. Étant donné que le premier actant de Oper_1 est déjà donné dans la phrase en tant que sujet grammatical de *to speak*, il n'est pas besoin de l'exprimer une fois de plus. Par conséquent, la phrase ne peut être interprétée que comme décrivant le support qui a été donné au parlement. À l'inverse, *the president enjoyed the support of the parliament*, sera interprétée dans l'autre sens compte-tenu du fait que *enjoy* est un Oper_2 dont le sujet est le deuxième actant du mot-clé.

2. Pointeurs vers les équivalents idiomatiques d'une langue cible pour la traduction automatique

Les propriétés d'universalité et d'idiomaticité interlangue font des FL un outil idéal pour la traduction de groupes d'expressions dans un système de traduction automatique. Par exemple, les prépositions locatives utilisées pour former des syntagmes prépositionnels dénotant des lieux, directions, repères temporels, périodes, intervalles, etc., sont très diverses et versatiles dans une langue donnée et même d'une langue à une autre. L'utilisation des FL permet d'éviter l'écriture fastidieuse d'un grand nombre de règles. Par exemple, les valeurs de la FL *Loc* en russe varient entre les prépositions *v* et *na* :

$\text{Loc}(\textit{institute}) = \textit{at}$ (the institute) ; $\text{Loc}(\textit{institut}) = \textit{v}$ (institut) ;

$\text{Loc}(\textit{work}) = \textit{at}$ (work) ; $\text{Loc}(\textit{rabota}) = \textit{na}$ (rabote), etc.

Pour assurer la production de ces équivalents en traduction automatique, il suffit simplement d'identifier les arguments et la valeur de la FL pendant l'analyse automatique et de substituer la valeur correcte dans la langue cible

pendant la génération.

Les implémentations réalisées dans les applications en TAL permettent de mettre en valeur certaines lacunes au niveau des FL. L'équipe du projet ETAP-3 a introduit une vingtaine de nouvelles FL avec un potentiel paraphrastique plus riche. Certaines de ces nouvelles fonctions sont des aménagements de FL existantes dont on a augmenté la granularité. Voyons en un exemple avec la FL Anti. Deux types sémantiques d'antonymes ont été dégagés qui diffèrent l'un de l'autre par la position de la négation dans leur représentation sémantique [Apresjan et al., 2003 : 10].

The first type of antonyms is based on inner negation. Consider the well-known pair *begin* and *stop*. *To stop doing P* can be explicated as 'to begin not to do P'. This type of antonymy is called ANTI1. The second type of antonyms is based on outer negation. Consider the pair *observe* (the regulations) and *violate* (the regulations); *to violate* means 'not to observe'. This type of antonymy is called ANTI2. [...]

À partir de ces deux types de négation, on peut formuler deux règles de paraphrasage :

1. $X \Leftrightarrow \text{Neg} + \text{ANTI2}(X)$:
He violated the regulations - He did not observe the regulations ;
2. $X+Y \Leftrightarrow \text{ANTI1}(X) + \text{ANTI2}(X)$:
He began to observe the regulations - He stopped violating the regulations.

Résumé automatique

Dans la problématique de la génération de résumés automatiques fluides et non seulement d'extraits de texte, Bélanger [2003] et Bélanger et Kittredge [2005] ont réfléchi à un modèle linguistique pour la reformulation et la condensation dans lequel les FL joueraient un rôle important. Pour ce faire, plusieurs articles de journaux en français et en anglais ont été analysés. L'étude montre qu'environ 70% des substitutions à partir de textes complets vers les textes résumés pourraient être

encodées par un ensemble de FL standard ainsi que quelques non standard. En plus des nombreux types de synonymes et antonymes, les FL les plus fréquentes sont d'abord les FL de dérivations (V_0 , S_0 , A_0 , Adv_0) puis les hyperonymes (**Gener**). Au contraire, les occurrences de paraphrases impliquant des constructions à verbes support se sont révélées relativement peu fréquentes. Par ailleurs, certaines opérations productives de reformulation repérées ne sont pas décrites en tant que règles de paraphrase standard dans la TST.

Il arrive que le matériau, pour les phrases du texte résumé, vienne de l'agrégation de deux phrases différentes du texte complet. Dans ce cas, les suppressions, substitutions synonymiques ainsi que les opérations de paraphrases semblent devoir être appliquées. Ainsi, dans l'exemple suivant, il doit y avoir une substitution synonymique impliquant une FL dans la première proposition et une opération de passivation appliquée aux deux propositions.

1. Texte initial : *Les feuilles ont un effet inhibiteur de l'efflux et peuvent, de ce fait ...*
L'appex, au contraire, a un effet stimulateur de l'efflux.
2. Résumé : *...l'efflux est freiné par les feuilles et stimulé, au contraire, par l'appex.*

Les FL en jeu dans ces opérations de transformation sont les suivantes :

- $Oper_1(\text{effet})$: *avoir*,
- $S_0(\text{stimuler})$: *stimulation*,
- $Able_1(\text{stimuler})$: *stimulateur*,
- $S_0(\text{inhiber})$: *inhibition*,
- $Able_1(\text{inhiber})$: *inhibiteur*,
- $Syn(\text{inhiber})$: *freiner*.

Notons que ce projet n'est resté qu'à une étape expérimentale d'étude préalable sur la faisabilité de l'utilisation des FL pour remplir cette tâche. Il n'a pas encore donné lieu à une implémentation visant la réalisation d'un programme de résumé

automatique.

Annexe III

Codification des classes du système de FL

Le système de FL est implémenté dans une table de données SQL contenant des codes. Ces codes renvoient aux différentes classes établies et présentées au chapitre 5. Nous présentons, dans cette annexe, la codification des classes et sous-classes des quatre perspectives du système, soit la hiérarchie sémantique (figure III.1), la liste des valences syntaxiques (figure III.2), la liste des couples de parties du discours (figure III.3) ainsi que la liste des cibles communicationnelles (figure III.4). La table peut être consultée à partir du site de l'OLST à l'adresse suivante : <http://www.olst.ling.umontreal.ca>.

- ▼ **Sem 1** : ÉQUIVALENCE
 - ▷ **Sem 1.1** : Lexies similaires
 - ▷ **Sem 1.2** : Conversifs
 - ▷ **Sem 1.3** : Reformulation syntaxique
- ▼ **Sem 2** : OPPOSITION
 - ▷ **Sem 2.1** : Opposition
 - ▷ **Sem 2.2** : Conversifs
 - ▷ **Sem 2.3** : Équivalent masculin/féminin
- ▼ **Sem 3** : ÉLÉMENT/ENSEMBLE
 - ▷ **Sem 3.1** : Hiérarchiques
 - ▷ **Sem 3.2** : Partie/tout
 - ▷ **Sem 3.3** : Petit de ~
 - ▷ **Sem 3.4** : Unité/ensemble régulier
 - ▷ **Sem 3.5** : Chef/équipe
 - ▷ **Sem 3.6** : Mesure
- ▼ **Sem 4** : QUALIFICATIFS
 - ▷ **Sem 4.1** : Qualificatifs de jugement
 - ▷ **Sem 4.1.1** : jugement positif
 - ▷ **Sem 4.1.2** : jugement négatif
 - ▷ **Sem 4.3** : Qualificatif d'intensité
 - ▷ **Sem 4.3.1** : Forte intensité
 - ▷ **Sem 4.3.2** : Faible intensité
 - ▷ **Sem 4.4** : Autres qualificatifs portant sur l'apparence physique
- ▼ **Sem 5** : PARTICIPANTS
 - ▷ **Sem 5.1** : Actants du mot-clé
 - ▷ **Sem 5.2** : Participants extérieurs
- ▼ **Sem 6** : PHASES/ASPECT
 - ▷ **Sem 6.1** : Préparation
 - ▷ **Sem 6.2** : Début
 - ▷ **Sem 6.3** : Milieu
 - ▷ **Sem 6.4** : Fin
 - ▷ **Sem 6.5** : Résultat
 - ▷ **Sem 6.6** : Répétition
 - ▷ **Sem 6.7** : Expression de la durée
- ▼ **Sem 7** : MANIÈRE
 - ▷ **Sem 7.1** : Instrument/moyen/manière
 - ▷ **Sem 7.2** : Manière dont l'entité ~ est utilisée/exprimée
 - ▷ **Sem 7.3** : Manière de réaliser l'action/événement ~
- ▶ **Sem 8** : CAUSE
- ▼ **Sem 9** : LOCALISATION
 - ▷ **Sem 9.1** : Lieu typique
 - ▷ **Sem 9.2** : Position spatiale/temporelle
- ▼ **Sem 10** : ACTION/ÉVÉNEMENT
 - ▷ **Sem 10.1** : Création/causation d'existence
 - ▷ **Sem 10.2** : Manifestation
 - ▷ **Sem 10.3** : Utilisation/fonctionnement typique
 - ▷ **Sem 10.4** : Augmentation/amélioration
 - ▷ **Sem 10.5** : Diminution/dégradation
 - ▷ **Sem 10.6** : Disparition/Fin d'existence
 - ▷ **Sem 10.7** : Non fonctionnement/Refus
 - ▷ **Sem 10.8** : Tentative

Figure III.1 – Codes associés aux classes sémantiques

▼ Syntp 1 : 1 actant

▷ Syntp 1.1 : [#]

▼ Syntp 2 : 2 actants

▷ Syntp 2.1 : [# , 1]

▷ Syntp 2.2 : [# , 2]

▷ Syntp 2.3 : [# , 3]

▷ Syntp 2.4 : [# , 4]

▷ Syntp 2.5 : [# , @]

▷ Syntp 2.6 : [1 , #]

▷ Syntp 2.7 : [2 , #]

▷ Syntp 2.8 : [3 , #]

▷ Syntp 2.9 : [4 , #]

▷ Syntp 2.10 : [@ , #]

▼ Syntp 3 : 3 actants

▷ Syntp 3.1 : [# , 1 , 2]

▷ Syntp 3.2 : [# , 2 , 1]

▷ Syntp 3.3 : [# , 2 , 3]

▷ Syntp 3.4 : [# , 2 , 4]

▷ Syntp 3.5 : [# , 3 , 2]

▷ Syntp 3.6 : [1 , # , 1]

▷ Syntp 3.7 : [1 , # , 2]

▷ Syntp 3.8 : [1 , # , 3]

▷ Syntp 3.9 : [1 , # , 4]

▷ Syntp 3.10 : [1 , # , @]

▷ Syntp 3.11 : [1 , 1 , #]

▷ Syntp 3.12 : [1 , 2 , #]

▷ Syntp 3.13 : [1 , 3 , #]

▷ Syntp 3.14 : [2 , # , 1]

▷ Syntp 3.15 : [2 , # , 3]

▷ Syntp 3.16 : [2 , 1 , #]

▷ Syntp 3.17 : [2 , 2 , #]

▷ Syntp 3.18 : [3 , # , 1]

▷ Syntp 3.19 : [3 , # , 2]

▷ Syntp 3.20 : [3 , # , 4]

▷ Syntp 3.21 : [3 , 1 , #]

▷ Syntp 3.22 : [3 , 3 , #]

▷ Syntp 3.23 : [@ , # , 1]

▷ Syntp 3.24 : [@ , # , 2]

▷ Syntp 3.25 : [@ , # , 3]

▷ Syntp 3.26 : [@ , 1 , #]

▷ Syntp 3.27 : [@ , 2 , #]

▼ Syntp 4 : 4 actants

▷ Syntp 4.1 : [1 , # , 2 , 3]

▷ Syntp 4.2 : [1 , # , 3 , 2]

▷ Syntp 4.3 : [2 , # , 1 , 3]

▼ Syntp Mod : Modification simple

▷ Syntp Mod.1 : [# ^]

▼ Syntp Mod1 : Modification à un actant

▷ Syntp Mod1.1 : [# ^ , 1]

▷ Syntp Mod1.2 : [# ^ , 2]

▷ Syntp Mod1.3 : [# ^ , 3]

▷ Syntp Mod1.4 : [1 ^ , #]

▷ Syntp Mod1.5 : [2 ^ , #]

▷ Syntp Mod1.6 : [3 ^ , #]

▷ Syntp Mod1.7 : [@ ^ , #]

Figure III.2 – Codes associés aux classes de valences syntaxiques

- ▼ **Pdd 1 : Valeur : N**
 - ▷ **Pdd 1.1** : $N \rightarrow N$
 - ▷ **Pdd 1.2** : $A \rightarrow N$
 - ▷ **Pdd 1.3** : $V \rightarrow N$
 - ▷ **Pdd 1.4** : $Adv \rightarrow N$
 - ▷ **Pdd 1.5** : $Claus \rightarrow N$
- ▼ **Pdd 2 : Valeur : V**
 - ▷ **Pdd 2.1** : $N \rightarrow V$
 - ▷ **Pdd 2.2** : $A \rightarrow V$
 - ▷ **Pdd 2.3** : $V \rightarrow V$
 - ▷ **Pdd 2.4** : $Adv \rightarrow V$
 - ▷ **Pdd 2.5** : $Claus \rightarrow V$
- ▼ **Pdd 3 : Valeur : A**
 - ▷ **Pdd 3.1** : $N \rightarrow A$
 - ▷ **Pdd 3.2** : $A \rightarrow A$
- ▷ **Pdd 3.3** : $V \rightarrow A$
- ▷ **Pdd 3.4** : $Adv \rightarrow A$
- ▷ **Pdd 3.5** : $Claus \rightarrow A$
- ▼ **Pdd 4 : Valeur : Adv**
 - ▷ **Pdd 4.1** : $N \rightarrow Adv$
 - ▷ **Pdd 4.2** : $A \rightarrow Adv$
 - ▷ **Pdd 4.3** : $V \rightarrow Adv$
 - ▷ **Pdd 4.4** : $Adv \rightarrow Adv$
 - ▷ **Pdd 4.5** : $Claus \rightarrow Adv$
- ▼ **Pdd 5 : Valeur : Claus**
 - ▷ **pdd 5.1** : $N \rightarrow Claus$
 - ▷ **pdd 5.2** : $A \rightarrow Claus$
 - ▷ **pdd 5.3** : $V \rightarrow Claus$
 - ▷ **pdd 5.4** : $Adv \rightarrow Claus$
 - ▷ **pdd 5.5** : $Claus \rightarrow Claus$

Figure III.3 – Codes associés aux classes de couples de parties du discours des liens lexicaux encodés par les FL

- ▶ **Cible #** : Mot-clé de la FL
- ▶ **Cible @** : Participant extérieur
- ▶ **Cible 1** : Premier actant du mot-clé
- ▶ **Cible 2** : Deuxième actant du mot-clé
- ▶ **Cible 3** : Troisième actant du mot-clé
- ▶ **Cible 4** : Quatrième actant du mot-clé
- ▶ **Cible 5** : Cinquième actant du mot-clé
- ▶ **Cible 6** : Sixième actant du mot-clé
- ▶ **Cible V** : Valeur de la FL

Figure III.4 – Codes associés aux cibles communicationnelles des FL

