

Université de Montréal

Termes et relations sémantiques en corpus spécialisés : rapport entre patrons
de relations sémantiques (PRS) et types sémantiques (TS)

par
Claudine Bodson

Département de linguistique et de traduction

Faculté des arts et des sciences

Thèse présentée à la Faculté des études supérieures
en vue de l'obtention du grade de Philosophiæ Doctor (Ph.D.)

en linguistique

option traduction

Août 2004

© Claudine Bodson, 2004

Université de Montréal
Faculté des études supérieures

Cette thèse intitulée :

Termes et relations sémantiques en corpus spécialisés : rapport entre patrons
de relations sémantiques (PRS) et types sémantiques (TS)

présentée par :

Claudine Bodson

a été évaluée par un jury composé des personnes suivantes :

Jeanne Dancette
présidente-rapporteuse

Marie-Claude L'Homme
directrice de recherche

Louise Dagenais
membre du jury

Anne Condamines
examinatrice externe

représentant du doyen de la FES

RÉSUMÉ

Les travaux entrepris dans le cadre de la présente thèse se situent dans le domaine de l'extraction semi-automatique d'informations sémantiques. Nous présentons un modèle dont l'objectif est de faciliter les recherches dans les corpus spécialisés. Nous cherchons à répondre aux besoins de la terminologie en repérant dans les textes des éléments utiles à la formulation de définitions.

Notre objectif principal consiste à élaborer un modèle d'association des patrons de relations sémantiques (PRS) aux types sémantiques (TS) auxquels appartiennent les termes. Les travaux antérieurs ont montré l'intérêt général des PRS pour la terminologie. L'association de patrons aux TS permettrait des requêtes encore plus précises dans les corpus spécialisés.

Nous postulons que les PRS peuvent être associés à des TS spécifiques indépendamment des domaines auxquels se rattachent les termes. Pour le démontrer, nous avons utilisé deux corpus; l'un porte sur la médecine et l'autre sur l'informatique. Ils sont composés de textes didactiques et de vulgarisation français : ces genres textuels sont le plus susceptibles de comporter des contextes définitoires. Nous avons étudié trente-trois termes associés à près de 710 contextes contenant un PRS. Nous nous sommes penchée sur les relations sémantiques d'hyponymie, de méronymie, de finalité et de causalité ainsi que sur les patrons métalinguistiques. Les termes choisis appartiennent à des types sémantiques différents et leur classement s'inspire de la hiérarchie des noms de WordNet.

Cette recherche montre qu'il est possible de mettre en rapport des relations sémantiques à des TS et ce, indépendamment des deux domaines étudiés. Cependant, elle montre également que la plupart des PRS sont spécifiques à un corpus.

Mots clés

Terminologie, sémantique, patron de relation sémantique (PRS), relation sémantique, type sémantique (TS), linguistique sur corpus, corpus spécialisés, extraction semi-automatique.

ABSTRACT

The research undertaken for this thesis is in relation to semi-automatic extraction of semantic information on terms. The objective of the model we are presenting is to facilitate searches in specialized corpora. Our aim is to meet needs specific to terminology by identifying within a text items that can be used to write definitions.

Our main objective is to develop a model for linking semantic relation patterns (SRPs) and semantic types (STs) to which the terms belong. Previous work has shown how SRPs can be of use in terminology, and linking the patterns to specific STs would result in more precise queries.

Our work is based on the premise that SRPs can be linked to specific STs regardless of the domain. For this research, we used two corpora, one in the field of medicine and the other in computer science. These corpora included both didactic and popularized texts in French, since these are the most likely to have the highest number of defining contexts. We studied 33 key terms in almost 710 contexts that included an SRP. We focused on the following semantic relations: hyperonymy, meronymy, function and causality, as well as metalinguistic patterns. Selected terms were linked to various STs and the classification is based on the unique beginners in the WordNet hierarchy.

Our research shows that it is possible to link semantic relations to semantic types, regardless of the domains. However, most SRPs are found to be specific to one of the corpora.

Keywords

Terminology, semantics, semantic relation patterns (SRPs), semantic patterns, semantic relations, semantic types (STs), corpus linguistics, specialised corpora, semi-automatic extraction.

TABLE DES MATIÈRES

RÉSUMÉ	iii
ABSTRACT	v
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX	xi
LISTE DES FIGURES	xv
LISTE DES SIGLES, DES ABRÉVIATIONS ET DES SYMBOLES	xvi
LISTE DES CONVENTIONS TYPOGRAPHIQUES.....	xvi
REMERCIEMENTS	xvii
1. INTRODUCTION	1
1.1 Intérêt des patrons de relations sémantiques pour la terminologie	3
1.2 Limites des travaux antérieurs	5
1.3 Objectifs de la recherche.....	6
1.4 Contenu de la thèse	10
2. ÉTAT DE LA QUESTION.....	12
2.1 Corpus spécialisés.....	13
2.1.1 Définition de corpus	14
2.1.2 Choix des textes et représentativité	17
2.1.3 Genres textuels et types de textes	22
2.1.4 Niveaux de spécialisation des corpus spécialisés	24
2.1.4.1 Registres de Pearson (1998).....	24
2.1.4.2 Registres de Rebeyrolle (2000).....	26
2.1.4.3 L'intérêt des registres didactiques et de vulgarisation pour étudier les énoncés définitoires.....	28
2.1.5 Taille du corpus	29

2.1.6	Disponibilité des corpus	31
2.1.7	Synthèse.....	31
2.2	Relations sémantiques.....	33
2.2.1	Relation hyperonymique.....	37
2.2.2	Relation méronymique	39
2.2.3	Relation causale	43
2.2.4	Relation de finalité.....	53
2.2.5	Synthèse.....	54
2.3	Information sémantique dans les corpus spécialisés.....	56
2.3.1	Noms donnés aux PRS selon les chercheurs	57
2.3.2	Travaux terminologiques.....	62
2.3.2.1	Ahmad et Fulford (1992).....	62
2.3.2.2	Rebeyrolle (2000), Auger (1998)	67
2.3.2.3	Pearson (1998).....	73
2.3.2.4	Davidson <i>et al.</i> (1998) et Meyer <i>et al.</i> (1999).....	75
2.3.2.5	Marshman <i>et al.</i> (2002)	80
2.3.2.6	Marshman (2002).....	83
2.3.2.7	Condamines <i>et al.</i> (1999)	84
2.3.2.8	Séguéla (1999).....	86
2.3.2.9	Synthèse	92
3.	MÉTHODOLOGIE.....	94
3.1	Choix des textes et mise en forme des corpus spécialisés	95
3.1.1	Choix des textes.....	95
3.1.2	Taille des corpus.....	96
3.1.2.1	Disponibilité des textes	96
3.1.3	Niveaux de spécialisation	97
3.2	Sélection des termes.....	99
3.2.1	Notion de terme	99
3.2.2	Choix des termes clés	100
3.2.3	Fréquence des contextes pertinents dans le corpus et diversité des auteurs	101
3.2.4	Liste des termes clés retenus.....	101
3.3	Extraction des contextes	106

3.3.1	Extension de sens du terme.....	107
3.3.2	Reprise anaphorique	107
3.3.3	Forme du PRS.....	108
3.3.4	Difficultés liées à certains patrons.....	109
3.4	Traitement des PRS.....	112
3.4.1	Exemples de neutralisation des patrons	115
3.4.2	Regroupement des patrons.....	118
3.4.2.1	Problèmes d’ambiguïté de certains patrons.....	118
3.4.2.2	Regroupement de certains patrons aux fins de l’interprétation des résultats	119
3.5	Regroupement des termes par type sémantique TS	121
3.5.1	Grands génériques de WordNet.....	121
3.5.2	Classement des termes.....	125
3.6	Description de la base de données	127
3.7	Conclusion	130
4.	RÉSULTATS.....	131
4.1	Remarques générales sur les contextes extraits	132
4.1.1	Résultats obtenus pour chacune des relations.....	133
4.1.2	Proportion des contextes retenus par rapport au nombre de contextes dépouillés pour chacun des termes clés.....	135
4.1.3	Répartition des relations sémantiques dans les corpus	139
4.1.4	Répartition des relations sémantiques dans les corpus pour lesquelles deux termes par corpus ont été étudiés.....	141
4.2	TS : [ENTITY OBJECT].....	144
4.2.1	Répartition des relations sémantiques par corpus.....	149
4.2.2	Répartition des PRS pour le TS [ENTITY OBJECT].....	151
4.2.3	Remarques générales sur le TS [ENTITY OBJECT].....	154
4.3	TS : [EVENT].....	154
4.3.1	Répartition des relations sémantiques par corpus.....	158
4.3.2	Répartition des PRS pour le TS [EVENT].....	161
4.3.3	Remarques générales sur le TS [EVENT]	163
4.4	TS : [ACTIVITY]	163

4.4.1	Répartition des relations sémantiques par corpus.....	170
4.4.2	Répartition des PRS pour le TS [ACTIVITY].....	171
4.4.3	Remarques générales sur le TS [ACTIVITY].....	173
4.5	TS : [PSYCHOLOGICAL FEATURE (FIELD)].....	174
4.5.1	Répartition des relations sémantiques par corpus.....	178
4.5.2	Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE (FIELD)].....	180
4.5.3	Remarques générales sur le TS [PSYCHOLOGICAL FEATURE (FIELD)].....	181
4.6	TS : [ENTITY ORGANISM]	181
4.6.1	Répartition des relations sémantiques par corpus.....	187
4.6.2	Répartition des PRS pour le TS [ENTITY ORGANISM]	189
4.6.3	Remarques générales sur le TS [ENTITY ORGANISM]	190
4.7	TS : [PHENOMENON]	191
4.7.1	Répartition des relations sémantiques par corpus.....	197
4.7.2	Répartition des PRS pour le TS [PHENOMENON].....	199
4.7.3	Remarques générales sur le TS [PHENOMENON]	201
4.8	TS : [STATE].....	202
4.8.1	Répartition des relations sémantiques par corpus.....	209
4.8.2	Répartition des PRS pour le TS [STATE]	211
4.8.3	Remarques générales sur le TS [STATE].....	213
4.9	TS : [ABSTRACTION]	213
4.9.1	Répartition des relations sémantiques par corpus.....	221
4.9.2	Répartition des PRS pour le TS [ABSTRACTION].....	223
4.9.3	Remarques générales sur le TS [ABSTRACTION]	225
4.10	TS : [PSYCHOLOGICAL FEATURE].....	226
4.10.1	Remarques générales sur le TS [PSYCHOLOGICAL FEATURE]	231
4.11	Synthèse des résultats	231
5.	CONCLUSION.....	239
	INDEX DES NOTIONS.....	246
	BIBLIOGRAPHIE.....	247

ANNEXE A : Description des corpus	xix
ANNEXE B : Acceptions des termes retenus pour le corpus de médecine et pour le corpus d'informatique.....	xxxii
ANNEXE C : Définitions des grands génériques de WordNet	xxxvi

LISTE DES TABLEAUX

Tableau 1.	Définitions de « corpus » selon différents auteurs.....	15
Tableau 2.	Liste des relations conceptuelles de Sager (1990 : 35).....	35
Tableau 3.	Adaptation par Barrière (2001-2002) de la classification de la relation causale de Talmy (1988, 2000), et simplifiée par Marshman (2002 : 16)	49
Tableau 4.	Tableau de synthèse contenant les définitions des différents noms donnés aux PRS selon les chercheurs	58
Tableau 5.	Relations sémantiques d’Ahmad et Fulford (1992).....	64
Tableau 6.	Liste de patrons (Ahmad et Fulford 1992)	66
Tableau 7.	Énoncés définitoires directs de Rebeyrolle (2000).....	69
Tableau 8.	Énoncés définitoires indirects de Rebeyrolle (2000 : 164).....	71
Tableau 9.	Résultats de Meyer <i>et al.</i> (2001 : page Internet).....	80
Tableau 10.	Tableau récapitulatif du contenu du corpus de l’étude	98
Tableau 11.	Termes retenus pour chacun des corpus	102
Tableau 12.	Patrons neutralisés de la relation hyperonymique	116
Tableau 13.	Patrons neutralisés de la relation méronymique	116
Tableau 14.	Patrons neutralisés de la relation causale.....	117
Tableau 15.	Patrons neutralisés de la relation de finalité	117
Tableau 16.	Neutralisations des patrons métalinguistiques.....	118
Tableau 17.	Exemples d’entrées dans la base de données.....	128
Tableau 18.	Nombre de contextes retenus pour chacun des termes clés/Nombre de contextes dépouillés	136
Tableau 19.	Répartition des relations sémantiques	140
Tableau 20.	Répartition des relations sémantiques pour lesquelles deux termes par corpus ont été étudiés	142
Tableau 21.	Répartition des PRS pour le TS [ENTITY OBJECT].....	145
Tableau 22.	Répartition des contextes par relation sémantique par ordre décroissant pour le TS [ENTITY OBJECT]	147
Tableau 23.	Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ENTITY OBJECT].....	149
Tableau 24.	Répartition des relations sémantiques par ordre décroissant pour	

le corpus d'informatique pour le TS [ENTITY OBJECT].....	149
Tableau 25. Répartition des relations sémantiques par terme clé pour le TS [ENTITY OBJECT]	150
Tableau 26. Répartition des patrons par corpus	152
Tableau 27. Répartition des relations sémantiques pour le TS [EVENT]	156
Tableau 28. Répartition des relations sémantiques par ordre décroissant pour le TS [EVENT]	156
Tableau 29. Répartition des relations sémantiques du corpus de médecine pour le TS [EVENT]	159
Tableau 30. Répartition des relations sémantiques du corpus d'informatique pour le TS [EVENT]	159
Tableau 31. Répartition des relations sémantiques par terme clé pour le TS [EVENT]	161
Tableau 32. Répartition des PRS par corpus	162
Tableau 33. Répartition des PRS pour le TS [ACTIVITY]	164
Tableau 34. Répartition des relations sémantiques par ordre décroissant pour le TS [ACTIVITY].....	165
Tableau 35. Répartition des relations sémantiques par terme clé pour le TS [ACTIVITY].....	169
Tableau 36. Répartition des relations sémantiques du corpus de médecine pour le TS [ACTIVITY]	170
Tableau 37. Répartition des relations sémantiques du corpus d'informatique pour le TS [ACTIVITY]	170
Tableau 38. Répartition des PRS par corpus	172
Tableau 39. Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE (FIELD)] 174	
Tableau 40. Répartition des relations sémantiques par ordre décroissant pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]	175
Tableau 41. Répartition des relations sémantiques par terme clé pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]	177
Tableau 42. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [PSYCHOLOGICAL FEATURE (FIELD)].....	179
Tableau 43. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [PSYCHOLOGICAL FEATURE (FIELD)].....	179
Tableau 44. Répartition des PRS par corpus	180

Tableau 45. Répartition des PRS pour le TS [ENTITY ORGANISM]	182
Tableau 46. Répartition des relations sémantiques par ordre décroissant pour le TS [ENTITY ORGANISM]	183
Tableau 47. Répartition des relations sémantiques par terme clé pour le TS [ENTITY ORGANISM]	186
Tableau 48. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ENTITY ORGANISM]	188
Tableau 49. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [ENTITY ORGANISM]	188
Tableau 50. Répartition des PRS par corpus	190
Tableau 51. Répartition des PRS pour le TS [PHENOMENON]	192
Tableau 52. Répartition des relations sémantiques par ordre décroissant pour le TS [PHENOMENON]	193
Tableau 53. Répartition des relations sémantiques par terme clé pour le TS [PHENOMENON]	196
Tableau 54. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [PHENOMENON]	198
Tableau 55. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [PHENOMENON].....	198
Tableau 56. Répartition des PRS par corpus	200
Tableau 57. Répartition des PRS pour le TS [STATE]	203
Tableau 58. Répartition des relations sémantiques par ordre décroissant pour le TS [STATE]	204
Tableau 59. Répartition des relations sémantiques par terme clé pour le TS [STATE]	208
Tableau 60. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [STATE]	210
Tableau 61. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [STATE]	210
Tableau 62. Répartition des PRS par corpus	212
Tableau 63. Répartition des PRS pour le TS [ABSTRACTION]	215
Tableau 64. Répartition des relations sémantiques par ordre décroissant pour le TS [ABSTRACTION]	216
Tableau 65. Répartition des relations sémantiques par terme clé pour le TS [ABSTRACTION]	220

Tableau 66. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ABSTRACTION]	222
Tableau 67. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [ABSTRACTION].....	222
Tableau 68. Répartition des PRS par corpus pour le TS [ABSTRACTION].....	224
Tableau 69. Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE].....	227
Tableau 70. Répartition des relations sémantiques par ordre décroissant pour le TS [PSYCHOLOGICAL FEATURE]	227
Tableau 71. Répartition des relations sémantiques par terme clé pour le TS [PSYCHOLOGICAL FEATURE]	230
Tableau 72. Relation la plus importante selon le TS	232
Tableau 73. Relations n'étant jamais associées à certains TS.....	235
Tableau 74. PRS le plus important pour une relation sémantique.....	237

LISTE DES FIGURES

Figure 1 : Axe des registres de Rebeyrolle (2000 : 194).....	28
Figure 2 : Relations hyperonymiques représentées dans une arborescence (Sager 1990 : 31).....	38
Figure 3 : Relations méronymiques représentées dans une arborescence (Sager 1990 : 32)	42
Figure 4 : Extrait des relations causales efficientes (Garcia 1997 : 11).	47
Figure 5 : Termes complexes repérés au moyen du terme clé <i>ordinateur</i>	104
Figure 6 : Termes complexes repérés au moyen du terme clé <i>angiographie</i>	105
Figure 7 : Grands génériques (Unique Beginners) de WordNet (Miller 1998 : 30).....	123

LISTE DES SIGLES, DES ABRÉVIATIONS ET DES SYMBOLES

BCT = Banque de connaissances terminologiques

GDT = Grand dictionnaire terminologique

PM = patron métalinguistique

PRS = patron de relation sémantique

TA = traduction automatique

TAL = traitement automatique des langues

TS = type sémantique

LISTE DES CONVENTIONS TYPOGRAPHIQUES

information : l'information qui se rapporte au terme et qui est introduite par un patron de relation sémantique est soulignée

patron de relation sémantique : le patron de relation sémantique est écrit en caractères gras

terme = le terme est écrit en italique

[TYPE SÉMANTIQUE] = le type sémantique est écrit entre crochets et en petites majuscules

REMERCIEMENTS

Un grand merci à ma directrice de recherche, Marie-Claude L'Homme, de m'avoir si bien encadrée tout au long de mes études de maîtrise et de doctorat. Je pense tout particulièrement à la préparation de l'examen de synthèse ainsi qu'à certaines questions cruciales qui se sont posées en cours de recherche et qui m'apparaissaient alors impossibles à résoudre.

Je tiens aussi à remercier Ingrid Meyer, professeure à l'École de traduction et d'interprétation de l'Université d'Ottawa, d'avoir accepté de superviser ma thèse, mais qui a dû se retirer en cours de route pour des raisons de santé. Étant spécialiste de la question de l'extraction de patrons riches en connaissances, M^{me} Meyer m'a fourni des commentaires fort judicieux vu sa grande expérience de recherche dans le domaine.

Je souhaite également exprimer ma reconnaissance à l'ensemble des membres de l'Équipe de Recherche en Syntaxe et Sémantique (ERSS) de l'Université Toulouse-le Mirail (Toulouse, France) qui m'a accueillie comme une des leurs pour un stage en 2000-2001. Mes remerciements vont tout particulièrement à Anne Condamines, chercheure et spécialiste de la question des patrons et des relations conceptuelles, qui a supervisé mes travaux de recherche lors de mon séjour en France. Ses questions fort pertinentes m'ont forcée à pousser ma réflexion plus loin. Je tiens à remercier aussi Josette Rebeyrolle avec qui j'ai pu discuter de sa thèse et de questions pertinentes sur les PRS qui se rapportaient à mon propre travail de recherche.

Je tiens aussi à remercier les organismes suivants de m'avoir accordé des bourses ou des subventions durant ma scolarité : l'Université de Montréal, l'Association canadienne de traductologie et le Conseil de recherche en sciences humaines (CRSH), le Fonds pour la formation des chercheurs et l'aide à la

recherche (FCAR), Technologies Lionbridge Canada, l'Office franco-québécois pour la jeunesse (OFQJ) et l'Office Québec-Amériques pour la jeunesse (OQAJ).

Pour terminer, j'exprime aussi ma reconnaissance aux responsables de l'Observatoire de linguistique Sens-Texte du Département de linguistique et de traduction de l'Université de Montréal pour m'avoir donné accès à ses corpus de textes ainsi qu'à ses outils informatiques.

Merci aussi aux parents et amis de leur soutien moral.

1. INTRODUCTION

Les travaux dont les résultats sont présentés dans les pages qui suivent portent sur les patrons de relations sémantiques (PRS), appelés *knowledge-rich patterns* par Meyer *et al.* (1999)¹, et le rôle qu'ils sont appelés à jouer dans la recherche terminologique. D'une manière générale, un PRS explicite un lien entre un terme et des éléments d'information sur son sens. Les exemples 1 et 2 ci-après – qui ont été tirés de nos corpus – illustrent ce que nous entendons par *PRS* et la notation que nous utiliserons.

1. À la différence des acquisitions séquentielles classiques, les *scanners* de nouvelle génération, à rotation continue et en mode spiralé, **permettent** d'acquérir tout un volume en un seul balayage pendant une seule manoeuvre de Valsalva ou une phonation, à condition que la coopération du patient soit suffisante (lary)².
2. Les *ordinateurs* de **type** IBM et Macintosh peuvent recevoir les mêmes disquettes, mais procèdent à un formatage différent (plaisen1).

Dans l'ensemble de la thèse (à moins d'indication contraire), le terme de la requête apparaît en italique et le PRS en caractères gras, tandis que l'information sémantique qui se rapporte au terme est soulignée. De plus, dans le texte, les patrons sont ramenés à une forme neutralisée.

Ainsi, le patron **X permet Y** (1), exprime une relation sémantique de finalité (qui indique ce à quoi sert un objet particulier) entre le terme *scanners* et l'information qui s'y rapporte, c'est-à-dire d'acquérir tout un volume en un seul balayage pendant une seule manoeuvre de Valsalva ou une phonation. Quant au

¹ D'autres appellations ont été proposées par différents auteurs; nous y reviendrons à la section 2.3.1.

² Les noms inscrits entre parenthèses renvoient aux textes du corpus (voir l'annexe A).

patron **type de X Y** (2), il montre une relation hyperonymique³ qui lie le terme générique *ordinateur* à ses spécifiques, c'est-à-dire IBM et Macintosh.

Dans cette recherche, les PRS sont étudiés dans des corpus spécialisés. Ils sont associés à des termes et sont recherchés dans deux corpus composés de textes de médecine et d'informatique. Ces textes sont principalement de nature didactique et de vulgarisation et sont disponibles ou ont été convertis en format électronique.

De nombreux chercheurs ont fait des recensements des patrons de relations sémantiques trouvés dans les textes spécialisés afin de faciliter l'extraction d'informations sur les termes. Nous avons voulu apporter des raffinements à ces premières recherches en étudiant la dépendance entre les PRS et le type sémantique (TS) des termes à décrire. Nous avons également voulu vérifier l'hypothèse selon laquelle il est possible de mettre en relation des PRS et des TS et de transcender ainsi les différences entre les domaines et les corpus.

Par exemple, pour les termes *scanner* et *maladie*, qui appartiennent tous deux à des TS différents (le premier renvoie à un appareil; le second, à un état), nous postulons que ces deux TS peuvent être associés à des PRS distincts dans les corpus. Pour le vérifier, nous utilisons un concordancier qui permet de repérer ces termes dans nos corpus. Ensuite, nous importons les contextes contenant ces termes, à condition qu'ils répondent à certains critères comme la présence d'un PRS de nature lexicale, ainsi que l'expression d'une information sémantique se rapportant au terme étudié. Nous étudions les relations sémantiques suivantes : la relation hyperonymique, méronymique, causale et de finalité. Nous nous sommes également penchée sur les patrons métalinguistiques. Nous dénombrons ensuite les types de PRS et les relations sémantiques exprimées pour chacun des termes étudiés.

³ Dans ce texte, *relation hyperonymique* englobe la relation hyperonymique et la relation hyponymique à moins d'indication contraire.

Pour notre étude, nous avons sélectionné trente-trois termes (que nous classons en fonction des types sémantiques auxquels ils appartiennent en nous inspirant de la hiérarchie des noms de WordNet) afin de voir avec quels PRS ils se combinent. Nous mettons au point un modèle caractérisant les différents TS et nous dégageons pour chaque type les PRS qui sont susceptibles de leur succéder. À leur tour, les PRS sont classés selon les relations sémantiques (hyponymie, méronymie, finalité, cause et patron métalinguistique) qu'ils expriment.

1.1 Intérêt des patrons de relations sémantiques pour la terminologie

Les patrons de relations sémantiques ont retenu l'attention des chercheurs appartenant à différents domaines et pour différentes applications. Notre optique est résolument terminologique et définitoire : 1) terminologique, car nos corpus sont spécialisés et les PRS qui nous intéressent sont associés à des termes; 2) définitoire, car les PRS apparaissent dans des contextes qui peuvent être utilisés dans la préparation de définitions. Les terminologues ou terminographes, qui cherchent à identifier et à décrire le sens des termes spécialisés, ont recours à des corpus pour effectuer leurs recherches, lesquels leur permettront de constituer des vocabulaires ou des lexiques.

La méthode traditionnelle utilisée pour rechercher des renseignements sur un terme consiste à interroger des textes au moyen d'un concordancier afin d'y repérer le terme en question. Toutefois, les informations obtenues sont de nature assez diversifiée si la requête porte uniquement sur le terme. Nous le démontrons au moyen des exemples 3 et 4 ci-après.

L'exemple 3 présente une requête formulée uniquement à partir du terme *programmation*. Cet exemple montre des contextes qui n'apportent aucune information utile à la préparation d'une définition. En effet, les quatre occurrences de *programmation* constituent de simples attestations du terme, mais ne nous disent pas en quoi consiste cette activité.

3. La qualité de la documentation, et surtout sa cohérence avec la *programmation*, est un élément primordial de l'efficacité du mainteneur. Un mainteneur récupère le travail de plusieurs programmeurs, il est donc souhaitable que le style de la documentation et de la *programmation* soit le plus homogène possible, ce que permet l'adoption de normes rigoureuses de documentation et de *programmation*. Un trop grand laisser-aller en phases de conception et de *programmation* se traduira inmanquablement par des coûts de maintenance prohibitifs qui peuvent ruiner complètement la rentabilité du projet (genilo).

L'exemple 4, en revanche, montre qu'une requête sur un terme jumelé à un PRS permet d'atteindre un contexte utile. La requête sur *programmation* et le PRS **X permet Y** localise un contexte dans lequel une relation de finalité est associée au terme.

4. La *programmation orientée objet* (POO) nous **a permis de concevoir un système** en tant que jeu d'objets collaborant les uns avec les autres (aspect).

En repérant les PRS de façon automatique ou semi-automatique, la recherche du terminographe est grandement facilitée. En principe, en combinant un terme et un PRS, le terminographe peut retrouver dans un corpus uniquement les informations dont il a besoin. Par exemple, il peut être nécessaire, pour rédiger la définition d'un terme, de connaître la finalité ou la composition d'un objet. Pour ce faire, il interrogera un corpus avec un concordancier (tel Naturel Pro) au moyen du terme en question accompagné de patrons⁴.

⁴ Les modes d'interrogation varient d'un logiciel à l'autre. Par exemple, il est possible avec Naturel Pro, pour tenir compte d'un certain nombre de variations flexionnelles, de faire suivre le terme, pour lequel on recherche une description, de ***ser*** (pour **X sert à Y, X servent à Y**, etc.) ou ***utilis*** (pour **X est utilisé pour Y**, etc.) afin de se renseigner sur sa finalité, ou encore ***compos*** (pour **X est composé de Y, X se compose de Y**, etc.) ou ***cont*** (pour **X contient Y**, etc.) afin d'obtenir des informations sur sa composition.

1.2 Limites des travaux antérieurs

Les travaux antérieurs – que nous décrivons en détail dans le chapitre 2 (section 2.3) – ont montré que les PRS utilisés pour extraire des informations précises de façon automatique ou semi-automatique étaient pertinents, mais qu’il restait encore des problèmes importants à résoudre (d’après Davidson *et al.* (1998) et Meyer *et al.* (1999)). L’un de ces problèmes réside dans le fait que les auteurs dressent des listes de patrons regroupés en fonction de la relation qu’ils expriment sans les associer à des types sémantiques.

Nous croyons que pour améliorer l’extraction d’informations se trouvant dans les corpus il serait avantageux de pouvoir se référer à un modèle faisant état des associations possibles entre les PRS et les TS auxquels appartiennent les termes. Le TS renvoie à une notion dont parlent plusieurs auteurs, dont Apresjan (1992) et Levin (1993), c’est-à-dire un regroupement d’éléments lexicaux ayant un même comportement linguistique. Apresjan, quant à lui, nomme cette notion *lexicographic type* et la définit ainsi :

« By a lexicographic type I mean a more or less closed group of lexemes which have a number of such linguistically relevant properties in common and require therefore a unified description in the dictionary and a unified set of rules for their grammatical treatment. The greater the number of such properties and of the rules referring to them the more interesting the lexicographic types » (Apresjan 1992 : 9).

L’information dont on a besoin pour rédiger la définition d’un terme serait ainsi plus rapidement repérée. Par exemple, pour trouver ce à quoi sert un *scanner*, il serait avantageux de pouvoir identifier les patrons susceptibles d’accompagner ce terme. Pour ce faire, il suffirait de se référer au TS auquel appartient ce terme, grâce au modèle que nous nous proposons d’élaborer.

1.3 Objectifs de la recherche

L'objectif général de notre recherche est de mettre au point un modèle caractérisant les différents TS et, pour chaque type, les PRS susceptibles de leur être associés. Le modèle devra, en premier lieu, servir à mieux cibler les recherches dans les corpus. En outre, nous souhaitons vérifier s'il est possible de mettre en rapport un type sémantique et une ou des relations sémantiques, d'une part, et un ou des patrons de relations sémantiques, d'autre part.

Nous espérons ainsi démontrer que les patrons de relations sémantiques et les relations sémantiques associés à des types sémantiques permettent d'outrepasser les différences entre les domaines de spécialité et les corpus.

Pour illustrer notre propos, nous présentons en 5 à 8 des exemples qui montrent des similitudes observables dans deux domaines de spécialité différents.

Informatique

5. L'*imprimante* (line printer) **permet** d'obtenir des résultats imprimés (zmspc1).
6. Les *ordinateurs* **permettent** à l'utilisateur de traiter et de communiquer de l'information (adevim).

Médecine

7. Radiologiquement, le *scanner*, en particulier en reconstruction volumique tridimensionnelle, **permet** d'illustrer de façon très précise les anomalies osseuses (snc).
8. Le *myéloscanner* **permet** une meilleure visualisation de la compression du sac dural et de la moelle et des racines par la collection épidurale (rach).

Les exemples 5, 6, 7 et 8 montrent des contextes puisés dans les corpus d'informatique et de médecine où le patron **X permet Y** exprime la finalité et

relie un objet à l'utilisation qu'on peut en faire. Dans ces exemples, les termes *imprimante, ordinateurs, scanner* et *myéloscanner* (qui renvoient tous à du matériel) peuvent tous être regroupés sous le même TS, tant dans le domaine de la médecine que de l'informatique. Dans les deux domaines, le PRS **X permet Y** exprime la même relation, soit la finalité. Examinons encore d'autres exemples (9, 10 et 11) où des termes représentant des concepts appartenant au même TS (à savoir des phénomènes) peuvent appeler des patrons semblables.

Informatique

9. L'*évolution* exponentielle des performances, alimentée par l'évolution de la technologie des semi-conducteurs, **a entraîné** une évolution concomitante des chemins de données (trajet).
10. Comme nous l'avons vu, l'*évolution* des ordinateurs **a entraîné** le besoin de disposer d'un système d'exploitation réellement multitâche capable de gérer des ordinateurs qui devenaient interconnectés et fonctionnaient en temps partagé pour fournir des services à des utilisateurs de plus en plus nombreux (internet).

Médecine

11. Le *développement latéral* de la masse **entraîne** une déformation progressive de la pièce sacro-coccygienne avec aspect de sacrum « en cimenterre » correspondant à une déformation de l'os, plutôt qu'à son agénésie (snc).

Comme le montrent les exemples 9, 10 et 11, le PRS **X entraîne Y** est employé pour exprimer la cause. Que ce soit en médecine ou en informatique, le même patron est utilisé pour des termes associés au même TS.

Dans un autre ordre d'idées, des types sémantiques distincts appellent des PRS différents. Les exemples 12, 13, 14 et 15 présentent des contextes dans lesquels figurent des termes (soit *maladies* et *maladie de Krabbe, scanner sans injection* et *scanner à balayage spiralé*) appartenant à des TS différents et auxquels sont assortis des patrons différents (**X lié à Y** et **X permet Y**) et ce, dans un même domaine, à savoir la médecine. En outre, les relations sémantiques

exprimées par ces patrons ne sont pas les mêmes. En effet, ici le patron **X lié à Y** exprime une relation causale, alors que le patron **X permet Y** exprime une relation de finalité.

Médecine

12. [...] ces *maladies* [AFFECTIONS DYSMYÉLINISANTES OU LEUCODYSTROPHIE] **sont liées** à un trouble métabolique d'origine enzymatique (image7).
13. La *maladie de Krabbe* **est liée** à un déficit en [[bêta]]-galactocérébrosidase (image7).
14. Radiologiquement, le *scanner*, en particulier en reconstruction volumique tridimensionnelle, **permet** d'illustrer de façon très précise les anomalies osseuses (snc).
15. Le *scanner à balayage spiralé* **permet** actuellement une exploration très satisfaisante du rachis traumatique dans un temps réduit (traum2).

On constate que la nature des TS a une incidence sur les PRS employés. Ainsi, un terme appartenant au même TS que *scanner* (à savoir du matériel) est susceptible d'être décrit par une relation de finalité, tandis qu'un terme appartenant au même TS que *maladie* (à savoir un état) est plus susceptible d'être associé à une relation causale. Il ne serait pas logique de s'attendre à trouver une relation de finalité (qu'exprimerait un patron du type **X permet Y**) pour *maladie*.

Cependant, on peut également s'attendre à ce que le rapport entre relations sémantiques et types sémantiques (et, par conséquent, entre patrons de relations sémantiques et types sémantiques) ne soit pas systématique, notamment pour les relations les plus répandues comme l'hyponymie. Les exemples 16 et 17 montrent que la relation hyperonymique est valable pour différents TS. *Traitement* et *imprimante* appartiennent à des TS différents (soit une activité et du matériel).

16. Le *traitement* de la cardiopathie sousjacente (par **exemple** le traitement anti-angineux) [...] (lechat).
17. On retrouve deux grands **types d'imprimantes** dans cette classe : par caractère et par points (plaisen1).

Un modèle présentant les différents TS accompagnés des PRS susceptibles de leur être associés selon un ordre d'importance permettrait donc de cibler une relation sémantique spécifique. Si l'on sait qu'en médecine et en informatique on retrouve souvent, pour un terme appartenant à un TS donné, un PRS exprimant la finalité, on pourra améliorer la recherche dans les corpus. Les termes des exemples 18 et 19 (*langage* et *électrocardiogramme*) sont associés à un même TS (c'est-à-dire une abstraction).

18. Ce *langage* [htm ou html] **permet** principalement de spécifier l'affichage des données, i.e. mettre en forme un document hypertexte via des commandes de formatage similaires à celles utilisées en Latex (tag en anglais) (rollan).
19. Il [*électrocardiogramme*] **permet** de faire le diagnostic soit d'extrasystoles, soit de tachycardie soutenue (troub1).

Le modèle que nous souhaitons mettre au point permettrait donc de dégager des généralisations sur le regroupement des termes par TS. Nous cherchons à rendre la recherche dans les corpus encore plus méthodique grâce à l'association des PRS à des TS. Il suffirait de s'y référer pour localiser le TS auquel appartient un terme et d'utiliser les PRS proposés (par ordre d'importance) dans les requêtes.

Par ailleurs, notre recherche permettra de dégager les relations sémantiques que font ressortir les différents TS, ce qui aidera le terminologue à définir les termes en fonction du TS auquel appartient le terme à traiter. En d'autres mots, notre modèle permettra non seulement de faciliter les recherches terminologiques, mais aussi de formuler les définitions en fonction des caractéristiques intrinsèques du TS dont relève un terme.

Enfin, notre recherche permettra de mettre au jour des patrons de relations sémantiques moins connus. On y trouvera les patrons les plus répandus pour une relation sémantique donnée et déjà documentés, mais notre étude systématique de deux corpus révélera d'autres patrons de relations sémantiques.

1.4 Contenu de la thèse

La présente thèse est divisée en trois grandes parties. Le chapitre 2 présente l'état de la question. Il est lui-même subdivisé en trois sections. La première section (2.1) aborde différents éléments liés aux corpus spécialisés : caractéristiques principales des corpus spécialisés par rapport aux corpus dits *généraux*, représentativité, niveaux de spécialisation et taille. La deuxième section (2.2) présente les notions relatives aux relations sémantiques. Nous y traitons des caractéristiques des relations sémantiques que nous avons étudiées dans notre travail, à savoir l'hyponymie, la méronymie, la causalité et la finalité. Nous décrivons également la notion de « patron métalinguistique ». Enfin, la troisième section (2.3) porte sur l'extraction de l'information sémantique dans les textes et décrit les principaux travaux terminologiques sur les PRS.

Le chapitre 3 décrit la méthodologie mise au point pour mener à bien cette étude. Ce chapitre est lui-même subdivisé en huit sections. En premier lieu, nous abordons la question du choix des textes et de la mise en forme des corpus spécialisés (section 3.1). Nous traitons, entre autres, de la taille de nos corpus, de la disponibilité des textes et des niveaux de spécialisation. Nous présentons ensuite la méthode de sélection des termes clés (3.2). Nous expliquons comment s'est fait le choix des termes clés de l'étude. Nous décrivons également les critères de sélection des termes et des contextes dans lesquels ces termes apparaissent : fréquence minimum des termes; contextes pertinents retenus dans notre analyse. Ensuite, nous abordons la question de l'extraction des contextes contenant l'expression d'une relation sémantique (3.3) et les différents problèmes rencontrés à cette étape. Les sections suivantes portent sur le traitement des PRS (3.4) et sur le regroupement des termes par TS (3.5). Nous expliquons comment

s'est fait le regroupement des termes et nous expliquons le rôle joué par les grands génériques de WordNet (3.6). Enfin, dans la dernière section, qui porte sur la description de la base de données (3.7), nous présentons l'analyse des contextes.

Le chapitre 4 présente l'analyse des résultats. Ce chapitre est subdivisé en onze parties. Nous présentons d'abord les généralités sur les résultats obtenus par l'étude des contextes extraits de nos corpus (section 4.1). Puis, dans les sections 4.2 à 4.10, nous analysons les résultats obtenus pour chacun des neuf TS étudiés. La section 4.11 présente la synthèse des résultats.

Enfin, dans la conclusion (chapitre 5), nous résumons l'essentiel de notre contribution et dégageons des pistes de recherche qui restent à explorer.

2. ÉTAT DE LA QUESTION

« Tous ces travaux reposent sur l'idée que le sens se construit en contexte mais aussi par le contexte. C'est donner un rôle central au corpus. On a souligné ce point dans le travail de G. Grefenstette. Celui de M. Sussna converge à cet égard. [...] ... mais c'est la confrontation des contextes entre eux qui fait émerger le sens. Les connaissances projetées sur le corpus ne servent alors que de révélateurs » (Habert *et al.* 1997 : 115).

Ce chapitre traite principalement des travaux antérieurs qui ont porté sur le repérage automatique ou semi-automatique d'informations sémantiques dans les corpus. Nous consacrons également deux autres sections à la présentation des concepts principaux auxquels nous ferons appel.

D'abord, nous consacrons une section aux corpus tels que les définit la terminologie (2.1). Le corpus spécialisé est à la base même de notre étude et il est nécessaire de préciser les différents éléments qui le composent. En outre, l'avènement des outils informatiques a considérablement bouleversé les méthodes de travail dans le domaine.

Nous présentons et définissons ensuite les relations sémantiques sur lesquelles nous nous pencherons, soit celles d'hyponymie, de méronymie, de finalité et de causalité (2.2). Nous avons retenu ces relations, car ce sont celles qui sont les plus documentées dans les travaux de terminologie et qui ont retenu l'attention des chercheurs qui se sont intéressés aux patrons de relations sémantiques. Nous avons étudié, en plus de ces quatre relations principales, les contextes contenant un patron métalinguistique.

Enfin, la dernière section (2.3) est réservée aux recherches de patrons dans les corpus. La section 2.3.1 porte sur les noms différents que les chercheurs ont donnés aux PRS; puis, nous présentons en détail les travaux les plus significatifs par rapport au nôtre (2.3.2).

2.1 Corpus spécialisés

Selon Ahmad et Rogers (2001), la terminologie a mis plus de temps à accepter l'utilisation de corpus que la lexicographie (où, toujours selon les auteurs, les corpus occupent une place centrale depuis longtemps) en raison de son orientation plus prescriptive et de sa démarche onomasiologique. Aujourd'hui, cependant, le corpus revêt une importance capitale dans presque tous les travaux de terminologie.

Bien que la terminologie ait, en théorie, une orientation onomasiologique, en pratique, le terminologue débute souvent sa recherche à partir d'un terme et de ses réalisations en contexte. Depuis les tout débuts de cette discipline, le terminologue a été appelé à faire du dépouillement de corpus et à repérer des termes en contexte (contrairement à ce qu'ont avancé Ahmad et Rogers (2001)). Ce travail lui permet, entre autres, de répertorier les néologismes et de s'assurer que la terminologie à recenser dans les lexiques et les vocabulaires soit représentative de celle employée par les spécialistes du domaine. Avec l'avènement de l'ordinateur, la technique a changé, mais la méthode est demeurée fondamentalement la même.

L'importance accordée désormais au concept de « corpus » est due au fait que l'informatique est venue bouleverser les méthodes de travail. On n'a qu'à penser à la taille du corpus qui n'a presque plus de limite, ou encore à la facilité avec laquelle on peut obtenir des textes en format électronique.

Dans cette section, nous présentons les éléments qui distinguent les corpus « général » (celui qui intéressera le lexicographe) et « spécialisé » (celui sur lequel travaillera le terminographe). Nous passons en revue différents points de vue d'auteurs, selon qu'ils sont issus d'une tradition lexicographique ou terminologique. Nous examinons d'abord la définition d'un corpus (2.1.1). Puis, nous traitons du choix des textes et de leur représentativité (2.1.2), des genres textuels et des types de textes (2.1.3) des niveaux de spécialisation des corpus

spécialisés (2.1.4), de la taille des corpus (2.1.5) et de leur disponibilité (2.1.6). Enfin, la synthèse (2.1.7) présentera les choix que nous avons faits pour notre recherche.

2.1.1 Définition de corpus

Cette section présente les définitions de « corpus » élaborées par différents spécialistes de la question, tant lexicographes que terminologues. Nous commenterons chacune des définitions, mais d'abord, le Tableau 1 présente le tout sous forme de résumé.

Tableau 1. Définitions de « corpus » selon différents auteurs

Auteurs	Spécialistes	Définitions
McEnery et Wilson (1996 : 177)	Lexicographes	« a finite collection of machine readable text, sampled to be maximally representative of a language or variety ».
Francis (1982 : 7)	Lexicographe	« a collection of texts assumed to be representative of a given language, dialect, or other subset of language, to be used for linguistic analysis ».
Atkins, Clear et Ostler (1992 : 1)	Lexicographes	« a subset of an ETL ⁵ built according to explicit design criteria for a specific purpose, e.g. the Corpus Révolutionnaire (Bibliothèque Beaubourg, Paris), the Cobuild Corpus, the Longman/Lancaster corpus, the Oxford Pilot Corpus ».
Sinclair (1991 : 171)	Lexicographe	« a collection of naturally-occurring language texts... chosen to characterize a state or variety of a language ».
Sinclair (1994 : 20)	Lexicographe	« ... a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of language ».
Pearson (1998 : 43)	Terminologue	« ... a corpus is an artefact; it is selected, chosen or assembled according to explicit criteria. It is stored in electronic form. It consists of pieces of naturally occurring language. In this context, we understand naturally occurring to mean that the pieces of language have not been tampered with or edited. [...] A corpus, may be used as a "sample of the language" (Sinclair 1994) or because it is "representative of a given language" (Francis 1982). A corpus may be a collection of transcribed spoken and/or written pieces of language, contrary to what the use of the word text might suggest ».
Bowker et Pearson (2002 : 9)	Terminologues	« A corpus can be described as a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria ».

⁵ **Electronic text library** : « a collection of electronic texts in standardized format with certain conventions relating to content, etc., but without rigorous selectional constraints ».

La définition de McEnery et Wilson (1996) incorpore les notions fondamentales de « collection », d'« échantillon » et de « représentativité ». L'intérêt de cette définition repose en effet sur ces trois éléments, qui sont essentiels à tout travail sur corpus.

Dans la définition de Francis (1982), on retrouve la notion de « représentativité », l'une des trois notions fondamentales décrites par McEnery et Wilson (1996). Les textes devraient en effet être représentatifs de la langue et du phénomène à décrire ou à étudier, et l'on souhaite toujours s'approcher de cet objectif.

Atkins, Clear et Ostler (1992) amènent l'idée selon laquelle le corpus devrait, dans l'absolu, avoir été créé en fonction d'objectifs définis au préalable. Toutefois, comme on le verra plus loin, les difficultés pratiques liées à la constitution d'un corpus nous empêchent parfois de respecter fidèlement ce critère.

Dans les définitions des terminologues Pearson (1998) et Bowker et Pearson (2002), on retrouve bien entendu les notions fondamentales de « collection », d'« échantillon » et de « représentativité », mais aussi les notions de « critères explicites » et de « format électronique ». Pour notre propre définition, nous retenons tous les éléments amenés par Bowker et Pearson, car ils sont en adéquation avec nos objectifs. C'est d'ailleurs sur ces définitions que s'alignera notre propre définition de « corpus » qui est présenté formellement à la section 2.1.6.

D'une manière générale, le corpus constitue un ensemble de données linguistiques représentatives observables dans leur environnement « naturel » et doit constituer le reflet d'usages réels. Il s'oppose aux ouvrages de référence linguistiques tels les dictionnaires, qui sont le résultat d'analyses faites par une communauté de spécialistes (L'Homme 2004). De plus, la sélection des textes doit reposer sur des critères explicites en fonction des objectifs de la recherche.

Tous les ensembles de textes ne peuvent donc pas être considérés pour la constitution d'un corpus.

Par ailleurs, en terminologie, la valeur d'une recherche dépend plus directement de la qualité du corpus de l'étude. Bien que cela tende à changer, par tradition, le terminologue s'appuie essentiellement sur les observations qu'il a faites dans les corpus alors que le lexicographe fait souvent intervenir son intuition et ses connaissances sur la langue :

« Le corpus est le préalable à partir duquel toute la recherche terminographie [*sic*] s'organise. Pendant longtemps, on a cru important de limiter l'intervention du terminographe à sa plus simple expression. Tout ce qu'il pouvait dire sur les termes devait forcément apparaître dans un texte rédigé par un spécialiste ou dans un autre ouvrage de référence. On reconnaît aujourd'hui qu'une plus grande latitude est non seulement nécessaire, mais souhaitable (L'Homme 2004 : 125-126) ».

2.1.2 *Choix des textes et représentativité*

Par corpus *représentatif*, on entend un corpus qui constitue le reflet d'un usage dans une langue et plus précisément des faits de langue que l'on souhaite étudier. Dans cette section, nous verrons que le choix des textes et la représentativité des corpus varient selon que l'étude porte sur la langue générale ou sur la langue de spécialité. Enfin, les auteurs précisent que, selon le type d'étude, il peut ne pas être approprié de sélectionner des échantillons de textes plutôt que des textes entiers, puisque certains faits de langue peuvent se retrouver dans certaines parties d'un texte seulement, ce qui peut avoir des répercussions sur les résultats.

En ce qui a trait aux corpus généraux, Biber (1993 : 243) croit que la représentativité peut être assurée par un échantillon de textes incluant toute la gamme de variation d'un vaste ensemble. Pour Leech (1991 : 28, note en bas de page), un corpus est dit *représentatif* dans la mesure où les résultats peuvent être généralisés à un corpus de plus grande taille. Pour McEnery (1996 : 22), un

corpus est dit *représentatif* s'il reflète le plus fidèlement possible les faits de langue ainsi que la fréquence.

Les travaux de Muller (1979, 1992a) avaient pour objet de déterminer le vocabulaire spécifique à des ouvrages (pièces de théâtre, romans, etc.) en appliquant des méthodes de statistique lexicale à des corpus de nature littéraire. Selon Muller (1992a : 77), le calcul statistique sert à quantifier la notion subjective de vocabulaire caractéristique d'un texte. Pour pouvoir faire des observations justes sur la fréquence, elles doivent être comparées à une fréquence théorique calculée sur un ensemble textuel plus grand. Ce type de comparaison permet d'identifier le vocabulaire qui se démarque de façon significative à l'intérieur d'un corpus. Pour l'auteur, le vocabulaire est considéré comme caractéristique dans les cas où la fréquence observée dans un échantillon ne se démarque que de 5 % de la fréquence théorique attendue.

Selon Muller (1992b : 13), si, par exemple, l'on veut connaître le nombre de mots moyen par vers dans une pièce de théâtre, le seul moyen parfait de s'assurer que la fréquence est juste est d'effectuer des calculs sur l'ensemble de l'œuvre. Cependant, pour éviter d'avoir à faire un travail aussi fastidieux, il est possible de faire des calculs sur un échantillon. Le calcul effectué sur cet échantillon constitue une estimation. Deux moyens permettent d'y parvenir : constituer l'échantillon en prenant dans chaque partie caractéristique de la pièce un nombre de vers proportionnel à son étendue, ou encore plus simplement, tirer au sort 100 vers sur lesquels on fera le comptage :

« Si l'on veut opérer sur un échantillon très petit, la première méthode est préférable, à condition évidemment de connaître les caractères de l'ensemble; le prélèvement aléatoire est plus sûr, à condition d'admettre un échantillon assez grand » (Muller 1992b : 13).

L'auteur mentionne qu'il est également possible de combiner les deux méthodes. Enfin, il résume en ces mots la notion d'échantillon et de représentativité :

« ... une partie d'un texte est un échantillon de ce texte [...] si elle est faite de fragments assez nombreux prélevés sur toute l'étendue du texte [...] ou choisis de telle façon que l'ensemble du texte y soit représenté, on peut le considérer comme représentatif du texte » (Muller 1992b : 14).

En plus de devoir tenir compte du critère de la représentativité, le chercheur doit prendre en compte d'autres éléments (Flowerdew 1993 : 232). Étant donné qu'un corpus est un objet construit pour un objectif donné, le chercheur doit, entre autres, choisir ses textes en fonction de son domaine d'étude, afin de s'assurer que les éléments de l'étude seront bien présents dans le corpus et ce, en quantité suffisante. Comme nous le verrons dans la section 2.1.5, c'est l'objectif de l'étude (et aussi les moyens dont le chercheur dispose) qui influence la taille des corpus. De plus, il doit prendre en compte la variété des auteurs des textes retenus pour éviter une sur-représentativité des formulations privilégiées par l'un ou l'autre des auteurs.

Certains auteurs sont d'avis qu'il est souhaitable que les textes soient présents en entier dans les corpus (Sinclair (1991), Pearson (1998 : 60) et Meyer et Mackintosh (1996b)). D'autres auteurs, comme Habert *et al.* (1997), croient plutôt qu'il vaut mieux ne pas choisir des textes entiers afin d'éviter que des parties spécifiques, telle l'introduction, ne soient sur-représentées :

« ... voire à ne pas retenir les empanns de texte continus, de manière à éviter de sur-représenter des « lieux » du texte particulier (l'introduction, par exemple) » (Habert *et al.* 1997 : 146).

À ce sujet, Sinclair a remplacé, en 1994, le terme *text* par *pieces of language* pour définir les composantes d'un corpus. À son avis, le terme *text* peut porter à confusion puisqu'un texte n'est pas toujours complet, c'est-à-dire allant de l'introduction à la conclusion. Selon Roe (1977), le corpus devrait contenir plusieurs grands échantillons :

« For an investigation into the text of a restricted and relatively clearly defined subject area, a corpus of around a million words, but consisting of a small number of large samples would be appropriate » (Roe 1977 : 21).

Ainsi, Roe préconise la diversité des auteurs et des styles dans un même corpus, mais des extraits plus petits.

La question des échantillons semble se poser différemment pour d'autres auteurs étudiant les langues de spécialité également. Dans un corpus spécialisé, les termes, les patrons et les contextes intéressants peuvent apparaître dans n'importe quelle section du texte. Un concept important peut être défini seulement dans les premières parties du texte; un terme écrit au long peut n'apparaître qu'une seule fois au début du texte (ailleurs il est repris sous forme d'abréviation). Bowker et Pearson (2002 : 49) suggèrent donc d'utiliser des textes entiers dans les études portant sur la langue de spécialité (LSP). Enfin, tout corpus devrait être constitué de textes provenant de différentes sources, rédigés par des auteurs différents afin d'avoir une idée d'ensemble des mots ou des termes qui sont employés dans la langue générale ou la langue de spécialité, selon le cas.

Muller (1992b : 11-12) traite aussi de délimitation c'est-à-dire des limites du texte à l'étude. Il est très souvent nécessaire, selon le type d'étude, de fixer les limites de l'ensemble étudié en réfléchissant sur la façon dont seront traités ces éléments. En effet, pour qu'il soit possible de comparer deux études, les critères qui ont été appliqués doivent être identiques :

« Quand on choisit un texte littéraire comme champ d'expérience, on devra préciser sur quelle édition on travaille, et fixer les limites du texte. Il peut paraître superflu, si l'on dit que dans *Le Cid*⁶ il y a 2 943 substantifs, de préciser que l'on a considéré le texte dit par les acteurs, à l'exclusion de la préface, des indications scéniques, des noms des personnages qui précèdent chaque réplique et chaque scène, etc.; il est pourtant bien des cas où de telles indications sont nécessaires pour fixer les limites de l'ensemble étudié. On sera amené d'autre part à définir de façon rigoureuse des termes comme « mot », « substantif », etc.; car si l'on dit que dans *Le Cid* 4,01 % des mots sont des adjectifs, alors que dans *Phèdre* cette proportion monte à 5,99 %, cela n'a d'intérêt que si dans les deux textes on a appliqué des critères strictement identiques, pour le découpage du texte en mots aussi

⁶ « Sauf indication contraire, ces données sont empruntées aux *Concordances* et aux *Index* de B. Quemada (Laboratoire d'Analyse Lexicologique de Besançon) (v. Bibliographie) ».

bien que pour l'identification de ceux qui ont le caractère « adjectif » (Muller 1992b : 11-12).

En ce qui concerne la délimitation des corpus spécialisés, elle s'appuie principalement sur les notions de domaine et de niveaux de spécialisation. Meyer et Mackintosh (1996) traitent de l'importance et de la difficulté de la délimitation du domaine dans les corpus. En effet, la délimitation peut être complexe, car les domaines tendent à être de plus en plus multidisciplinaires, et que les textes peuvent être plus ou moins spécialisés, selon le cas. Une fois que ces frontières ont été délimitées, les auteurs précisent qu'il faut choisir des textes qui couvrent bien l'ensemble du domaine, c'est-à-dire qu'il faut bien balancer les sous-domaines et les domaines apparentés.

Biber (1993 : 247-248) – pour qui les corpus sont censés fournir aux chercheurs des échantillons de langue représentatifs, c'est-à-dire contenant toute la gamme de la variation linguistique d'une langue – utilise divers traits linguistiques pour caractériser les textes. Les regroupements de traits forment ce que l'auteur appelle des *dimensions*. L'auteur étudie aussi la répartition des textes sur les différentes dimensions dans le but de regrouper ceux qui partagent le plus grand nombre de propriétés communes pour constituer des types de textes. En fin de compte, ces types de textes sont définis sur la base de leur fonctionnement linguistique (lexical, grammatical, syntaxique) et non plus sur la base de caractéristiques externes (fonctionnelles, situationnelles). Biber (1988) en arrive à la conclusion suivante :

« Selon la dimension considérée, il y a plus de similarité de fonctionnements linguistiques entre X et Y qu'entre Y et Z alors que Y et Z paraissent intuitivement proches; par exemple, du point de vue de la dimension narrative, les dialogues spontanés sont plus proches des biographies que des conversations téléphoniques (Biber 1988 : 136 », dans Condamines 2003 : 148).

2.1.3 *Genres textuels et types de textes*

Cette section traite des points de vue d'auteurs sur les genres textuels et les types de textes. Selon Biber (1988 : 170), les genres textuels (ou registres) caractérisent les textes sur la base de propriétés extra-linguistiques concernant l'objectif et le sujet de l'auteur ou de l'interlocuteur; elles sont attribuées sur la base de leur emploi plutôt que de leur forme. Condamines (2003) abonde dans le même sens et résume ainsi la notion de genre textuel :

« ... ensemble d'éléments extra-linguistiques qui s'élaborent conjointement avec des régularités linguistiques pour constituer une sorte de norme qui vient se superposer à la norme linguistique (Bakhtine, 1984) (Condamines 2003 : 147) ».

Quant aux types de textes, ce sont des textes définis sur la base des propriétés linguistiques spécifiques qui les caractérisent. Par exemple, l'objectif de l'auteur d'un article de revue de niveau universitaire sur l'histoire asiatique est didactique. Cependant, sa forme linguistique pourrait être narrative et plus près de certains types de fiction plutôt que d'articles didactiques en sciences ou en génie. Ou encore, un texte de science fiction représente un genre de fiction (concerne l'objectif de l'auteur), mais pourrait représenter un type de résumé et de texte technique (du point de vue de sa forme linguistique), semblable à certains types de communications de nature didactique et différent de la plupart des autres textes de fiction.

Certains genres ont plusieurs sous-classes qui sont assez différentes les unes des autres. Par exemple, les textes didactiques comprennent des articles en ingénierie, des analyses politiques et historiques et des discussions littéraires. La forme linguistique des textes dans d'autres genres peut être relativement variée. En résumé, dans une analyse de types de textes, les textes de genres différents sont regroupés lorsqu'ils sont semblables du point de vue linguistique; les textes qui composent un seul genre peuvent comporter plusieurs types de textes différents.

Comme le mentionne Bronckart (1996 :76), les genres textuels sont vagues. Ils sont difficiles à classer à cause de la diversité des critères utilisés pour définir un genre. Cependant, Condamines (2003 : 38) fait remarquer que certains genres sont marqués linguistiquement et que cette caractéristique doit être prise en compte : « le roman, la presse, la petite annonce, le manuel, le sermon, le cours... » (Condamines 2003 : 38). Ces dénominations peuvent également être subdivisées en sous-genres : « roman historique, roman d’aventure, roman épistolaire, presse quotidienne, presse hebdomadaire, presse féminine, sportive... manuel de logiciel, d’appareil ménager... » (Condamines 2003 : 38). Bien que ces sous-genres soient difficilement cernables, ce phénomène de dénomination montre une certaine régularité et il est important de s’en inspirer au départ quitte à corriger le tir.

Cependant, en langue de spécialité, la classification des genres se pose différemment de celle proposée entre autres par Biber (1988). Selon Meyer et Mackintosh (1996), les textes de langue de spécialité peuvent varier énormément en genres textuels (que nous appelons « niveaux de spécialisation » (consulter la section 2.1.4). Le genre textuel, possède, comme nous l’expliquent les auteures, ses propres conventions linguistiques et pragmatiques, telles que l’historique médical d’un patient, le brevet d’invention dans un texte de loi, etc. Cependant, un élément plus important encore, rapportent les auteures, concerne la variation de genres textuels à l’intérieur d’une même langue de spécialité, et plus précisément la variation de technicité des textes. Par exemple, les experts d’un domaine utilisent des textes hautement techniques, alors que les techniciens et les journalistes ont besoin de textes d’un degré de technicité moindre. S’inspirant de Bowker (1995), Meyer et Mackintosh (1996) ont subdivisé les genres textuels de la façon suivante : didactiques (ex. manuels), avancés (ex. : articles scientifiques) et de vulgarisation. (ex. : articles de journaux). Selon Ahmad (1995 : 61), la diversité des genres textuels dans un corpus permet un meilleur équilibre. Bowker (1995 : 13) abonde dans le même sens et précise que l’inclusion de textes appartenant à différents niveaux de technicité permet de brosser un tableau complet de la diversité terminologique de la langue de spécialité étudiée.

2.1.4 *Niveaux de spécialisation des corpus spécialisés*

La section précédente traite des généralités sur les genres textuels, alors que la présente section détaille les niveaux de spécialisation⁷ des textes dans les corpus spécialisés. Nous verrons deux points de vue sur la question, soit celui de Pearson (1998) et celui de Rebeyrolle (2000). Ces auteures ont travaillé sur les PRS et ont également fait des rapprochements entre les niveaux de spécialisation et les énoncés définitoires⁸. Comme nous le verrons, certains types de textes sont plus susceptibles que d'autres de contenir des énoncés de ce type.

2.1.4.1 Registres de Pearson (1998)

Pearson (1998 : 36-38) distingue des niveaux de spécialisation en fonction des participants à la communication spécialisée : 1) communication expert – experts; 2) communication expert – initiés; 3) communication semi-expert – non-initiés; 4) communication enseignant – apprenants. Nous examinons ci-dessous chacun de ces types de communication.

- 1) La communication expert – experts utilise souvent un jargon hautement spécialisé. On présume que l'auteur et le lecteur partagent tous deux une langue de spécialité commune et que, lorsque certains mots ou expressions sont employés, chacun en comprend le sens. Cette langue diffère de la langue générale, car des sens spécifiques ont été attribués aux termes utilisés et définis antérieurement à l'acte de communication par une autorité externe (un organisme de normalisation, un dictionnaire spécialisé ou, encore, ce qui est généralement perçu comme tel dans le domaine). Les auteurs s'entendent pour comprendre et utiliser la terminologie telle

⁷ Pour notre recherche, nous avons choisi le terme *niveau de spécialisation*, aussi employé par L'Homme (2004), plutôt que *genre textuel*, car nous trouvons le terme plus approprié étant donné que nous travaillons sur des corpus spécialisés.

⁸ Un énoncé définitoire pour Rebeyrolle (2000) est un énoncé dans lequel on trouve des patrons métalinguistiques (voir la section 2.3.2.2).

qu'elle a été définie à l'origine. On présume que l'auteur et le lecteur (ou l'auditeur) ont le même ou presque le même niveau d'expertise. Dans un corpus, ce contexte de communication particulier devrait offrir le plus de termes.

- 2) Dans la communication expert – initiés, le niveau d'expertise des initiés est inférieur à celui des experts bien que les premiers aient certaines connaissances du domaine (ex. : étudiants du 3^e cycle universitaire dans une discipline particulière, ou encore, des destinataires ayant une formation dans le même domaine mais à un niveau d'études différent comme les ingénieurs et les techniciens). Dans ce type de communication, les experts emploient la terminologie qu'ils utiliseraient pour communiquer avec leurs pairs, mais ils sont portés à expliquer certains termes qu'ils croient inconnus du destinataire ou mal compris par lui. Ce qui caractérise la communication expert – initiés par rapport à la communication expert – experts est la différence du niveau d'expertise entre l'auteur et le lecteur (ou l'auditeur). Par conséquent, le nombre de termes devrait être inférieur étant donné que la communication comprendra des explications qui pourraient, si nécessaire, faire usage de vocabulaire plus général. Puisque, dans ce contexte, l'objectif de la communication est d'assister le lecteur dans son apprentissage du domaine, les explications sont en général assez détaillées et spécifiques.
- 3) Communication semi-expert – non-initiés. Les non-initiés ne travaillent pas nécessairement dans le domaine faisant l'objet de la communication. Les textes écrits pour ce public-cible devraient contenir moins de termes que les textes associés aux deux premiers types de communication. Puisqu'il ne suppose aucune connaissance spécifique du sujet, l'auteur peut même choisir d'utiliser un mot de la langue générale plutôt que le terme spécialisé approprié pour faciliter la compréhension du lecteur. Lorsque l'auteur emploie des termes, il les explique ou indique qu'il croit que le concept est déjà connu du lecteur. Contrairement aux deux types de

communication qui précèdent, pour la communication semi-expert – non-initiés, il n'est pas nécessaire pour l'auteur et le lecteur d'atteindre le même niveau de compréhension des termes employés tant que l'essentiel du message est compris. Ainsi, ce type de communication ne conduit généralement pas à l'emploi rigoureux des termes proprement dit.

- 4) Communication enseignant – apprenants. Le terme *apprenant* est employé pour désigner une personne qui n'a aucune connaissance préalable dans un domaine particulier, mais qui doit se familiariser avec celui-ci pour des raisons professionnelles ou scolaires. Les ouvrages utilisés pour l'apprentissage sont des ouvrages d'introduction ou des manuels. L'auteur utilisera la terminologie appropriée, mais présupera d'un niveau d'expertise moins élevé qu'en communication expert – initiés. Il fournira beaucoup plus d'explications et de définitions formulées soit en langue générale, soit en langue technique simplifiée. Bien qu'il soit important que les concepts soient maîtrisés, les apprenants ne sont pas tenus d'atteindre le même niveau de compréhension que les spécialistes en herbe de la deuxième catégorie. Néanmoins, lorsqu'on emploie un terme, on lui donne une signification particulière dans un domaine de spécialité donné et les apprenants sont tenus de comprendre le terme de cette façon.

2.1.4.2 Registres de Rebeyrolle (2000)

Le découpage de Rebeyrolle (2000) diffère sensiblement de celui proposé par Pearson (1998). En effet, le vocabulaire de Pearson (1998) met l'accent sur l'émetteur et le destinataire, alors que celui de Rebeyrolle (2000), met plutôt l'accent sur le type de texte, le niveau de spécificité et la fonction. Rebeyrolle (2000) distingue quatre registres spécialisés présentés ci-après.

- 1) Le registre *didactique*, illustré par un manuel de géomorphologie qui s'adresse à des étudiants de premier cycle qui se forment à une discipline. Sa fonction est descriptive et expositive, c'est-à-dire que le texte vise à la

fois la description des formes et l'explication de leur mode de formation, puisque l'objectif essentiel est d'exposer un savoir à maîtriser.

- 2) Le registre *scientifique*, illustré par des articles scientifiques dans le domaine de l'ingénierie des connaissances, s'adresse à des spécialistes (chercheurs). Sa fonction est également descriptive et expositive.
- 3) Le registre *technique*, illustré par des guides techniques en génie logiciel provenant d'une entreprise, s'adresse à des professionnels. Sa fonction est descriptive et injonctive, c'est-à-dire que les textes présentent des méthodes de développement de projet et indiquent la manière de les réaliser.
- 4) Quant au registre *vulgarisation*, illustré par des articles extraits d'une encyclopédie portant sur la géomorphologie, la géographie, la géologie et la climatologie, il s'adresse à un vaste public. Sa fonction est descriptive et expositive.

La communication expert – experts et la communication expert – initiés de Pearson (1998) se situent entre le registre *scientifique* et le registre *technique* de Rebeyrolle (2000), selon la spécificité des textes. Ces registres sont peu susceptibles de contenir des énoncés définitoires puisque, généralement, la terminologie est bien connue du destinataire. Quant à la communication semi-expert – non-initiés de Pearson, elle correspond partiellement au registre *vulgarisation* dont parle Rebeyrolle (2000), puisque cette dernière a utilisé des textes provenant d'articles encyclopédiques plutôt que d'articles de revue ou de sources spécialisées.

2.1.4.3 L'intérêt des registres didactiques et de vulgarisation pour étudier les énoncés définitoires

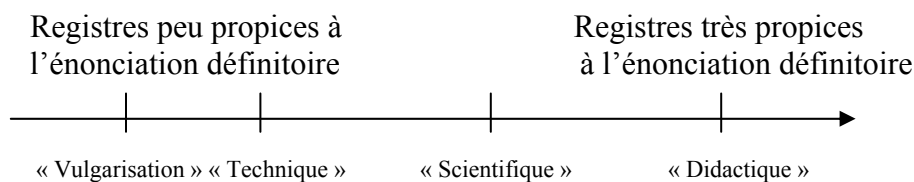
Rebeyrolle (2000) a montré que près de la moitié des occurrences des énoncés définitoires de l'ensemble du corpus qu'elle a étudié se trouvent dans le registre *didactique* (43,5 % des énoncés définitoires de l'ensemble du corpus sont réunis dans ce registre). Quant à l'autre moitié, elle se répartit de façon à peu près équilibrée entre les trois autres registres (environ 20 % dans chacun).

Rebeyrolle a hiérarchisé les quatre registres étudiés sur un axe dont les deux pôles permettent d'opposer des registres peu propices à l'énonciation définitoire (*technique* et surtout *vulgarisation*) à des registres qui leur sont très propices (*didactique*) :

« Le pôle gauche de l'axe est occupé par les registres "Techniques" et "Vulgarisation" qui représentent ainsi des registres, sinon inhospitaliers, du moins peu favorables à la définition. Au centre de cet axe, le registre "Scientifique" représente un registre où tous les types d'énoncés définitoires sont présents, mais avec une fréquence moindre que dans le registre "Didactique". Que cette position, dans laquelle on attendait effectivement le registre "Techniques", soit également occupée par le registre "Vulgarisation" ne manque pas d'étonner » (Rebeyrolle 2000 : 194).

La figure 1 présente l'axe des registres selon qu'ils sont peu ou très propices à l'énonciation définitoire selon Rebeyrolle.

Figure 1 : Axe des registres de Rebeyrolle (2000 : 194)



Les textes du registre *vulgarisation* s'adressent à des non-initiés et devraient donc, en principe, être favorables aux énoncés définitoires. Or, les résultats de Rebeyrolle (2000 : 194-195) montrent que les textes de ce registre sont ceux qui contiennent le moins grand nombre de définitions parmi les quatre registres de l'étude. Ces résultats sont plutôt surprenants.

Toutefois, ce résultat ne correspond ni à sa propre intuition ni aux observations de Loffler-Laurian (1983, 1994). Selon cette dernière, les définitions sont fort bien représentées dans les « discours de vulgarisation scientifique », textes provenant de revues destinées au grand public. Les deux constats ne sont pas irréconciliables puisque dans le corpus vulgarisation de Rebeyrolle (2000), les textes utilisés provenaient d'articles encyclopédiques plutôt que d'articles de revue ou de sources spécialisées.

2.1.5 *Taille du corpus*

Bien que tous s'entendent pour dire que la question de la taille d'un corpus est fort importante, il n'existe pas de consensus sur la taille idéale qu'il devrait avoir. Aujourd'hui, les centres de recherche disposent de données textuelles de millions et même de milliards de mots. Par exemple, les corpus de langue générale de Brown et LOB (Lancaster Oslo Bergen)⁹, comptant un million de mots chacun, sont aujourd'hui définis comme étant de petits corpus.

Pour certains auteurs, dont Pearson (1998) et Sinclair (1991), un corpus doit être le plus volumineux possible. Pearson (1998 : 56-57) ajoute que la taille

⁹ Le corpus Brown a été constitué en 1979 par W. N. Francis et H. Kucera, à l'Université Brown (Providence, RI). Il comprend 500 extraits de 2000 occurrences chacun provenant de textes écrits en anglais américain standard, tous publiés en 1961 et relevant de 15 genres, comme le reportage, les écrits scientifiques et techniques, etc. Quant au corpus Lancaster-Oslo-Bergen (LOB), il a été constitué par S. Johansson and G. N. Leech aux universités d'Oslo et de Lancaster entre 1970 et 1978. Il est composé de textes écrits en anglais britannique standard, sélectionnés selon les mêmes critères que le corpus Brown.

d'un corpus devrait être déterminée en fonction des objectifs de l'étude, de même qu'en fonction du caractère des textes ainsi que du domaine auxquels ils appartiennent. Cependant, il semble que la taille « idéale » ne soit pas définie de la même manière pour les corpus généraux et spécialisés. Nous examinons le point de vue de différents spécialistes de cette question.

Plusieurs auteurs, dont Engwall (1994) et James *et al.* (1994), sont d'avis qu'un corpus spécialisé ne doit pas obligatoirement être aussi grand que les corpus généraux.

Ahmad et Rogers (2001 : 736) vont même jusqu'à affirmer que, dans une discipline très pointue, un corpus de taille raisonnable peut contenir environ 100 000 mots. Selon ces auteurs, plusieurs raisons justifient le recours à des corpus moins volumineux en terminologie.

D'abord, on observe moins de variations lexicales et grammaticales en langue de spécialité qu'en langue générale. Les éléments linguistiques recherchés se présentent fréquemment, même dans un corpus de petite taille. De plus, le lexique des textes spécialisés est généralement plus dense que celui trouvé dans les textes de langue générale. Aussi, seuls les termes intéressent les terminologues, et non tous les mots du texte.

L'Homme (2004) ajoute que, puisque le texte spécialisé porte sur un sujet ciblé, il fait appel à un nombre limité de termes. On peut donc retrouver plusieurs occurrences d'un même terme sans avoir nécessairement à parcourir un corpus gigantesque. De plus, comme le corpus est défini dans les limites d'un domaine spécialisé, les termes ou les acceptions de certains termes à retenir en fonction des objectifs de l'étude devraient apparaître assez fréquemment et ce, même dans un corpus de petite taille.

Malgré sa taille réduite (si on le compare aux corpus de nature « générale »), le corpus de spécialité doit être assez volumineux pour qu'un nombre suffisant d'éléments permettent d'en tirer des généralisations.

2.1.6 *Disponibilité des corpus*

La disponibilité des textes peut obliger à faire des entorses à ce qui est jugé acceptable sur le plan de la taille idéale ou de la représentativité. En effet, la question des droits d'auteur, d'une part, des coûts et du temps nécessaires à la constitution d'un corpus – surtout lorsqu'il faut numériser les textes –, d'autre part, peuvent contraindre nos choix. Comme le note Pearson, ces contraintes, conjuguées à la disponibilité du matériel et aux considérations relatives aux droits d'auteurs, semblent expliquer la raison pour laquelle un corpus plus petit soit de meilleure qualité.

« ... the rationale for using a smaller corpus tends to be related less to a decision to keep it small than to the availability of material and copyright considerations. Compilers of such corpora hope, rather than know for certain, that the phenomena which they are investigating will appear with sufficient frequency in their smaller corpora to give them adequate results » (Pearson 1998 : 57).

La taille des corpus est aussi souvent limitée par la disponibilité des textes, comme en témoigne Engwal (1994) :

« ... but here lack of resources sometimes restricts the desired design » (Engwall 1994 : 51).

Il peut être encore plus difficile de mettre la main sur des textes spécialisés rédigés en français, comme le souligne Rebeyrolle :

« À ce sujet, nous voudrions souligner la difficulté qu'il y a à trouver des corpus électroniques en français dès qu'on s'intéresse aux textes spécialisés et qu'on s'écarte des textes journalistiques... » (Rebeyrolle 2000 : 100).

2.1.7 *Synthèse*

Dans cette section sur les corpus, nous avons vu que l'avènement des outils informatiques avait modifié la manière de conduire des recherches terminologiques. Ils permettent, entre autres, de constituer des corpus de très

grande taille, voire de taille pratiquement illimitée. Nous avons également vu que les corpus généraux et spécialisés ne sont pas constitués de manière identique. Par exemple, les spécialistes s'entendent pour dire que le corpus spécialisé est généralement de plus petite taille que le corpus général. De plus, nous partageons la même orientation que Biber (1993), mais nous nous différencions de sa méthode. L'objectif de ses travaux est de fournir une typologie des textes reposant fondamentalement sur leur fonctionnement linguistique. En effet, en ce qui a trait aux niveaux de spécialisation, l'important pour nous était simplement de sélectionner des textes riches en informations définitoires; nous n'avions pas pour objectif de caractériser le fonctionnement global d'un texte. Enfin, nous nous différencions également de la méthode de Biber (1993), qui utilise diverses méthodes statistiques (analyses multidimensionnelles) ainsi que différentes dimensions. Enfin, nous avons vu que les notions de base qui définissent le corpus moderne, peu importe l'application pour laquelle il est construit, sont celles de « collection », d'« échantillon » et de « représentativité », ainsi que les notions de « critères explicites » et de « format électronique ».

Notre conception de *corpus* s'inspire fortement de la définition de Pearson (1998), mais nous conservons uniquement les éléments nécessaires et suffisants pour décrire les corpus utilisés dans le cadre de notre étude. Chacun de nos corpus constitue donc une : collection de textes écrits – disponibles ou transformés en format électronique – portant sur un domaine de spécialité et sélectionnés en fonction d'objectifs de recherche, pouvant servir d'échantillon, car ce dernier devrait être représentatif des usages linguistiques dans un domaine spécialisé.

Chacun de nos deux corpus contient un peu plus d'un demi million de mots. Avec des corpus de cette taille, nous rejoignons les terminologues, qui constituent des corpus de moindre envergure que ceux utilisés par les lexicographes, pour qui les corpus d'un million de mots sont aujourd'hui considérés comme de petits corpus.

Nos corpus sont constitués de textes entiers provenant de sources variées. Notre corpus de médecine contient 117 documents et notre corpus d'informatique en contient 72. Ces textes sont rédigés par un grand nombre d'auteurs différents. Nos corpus sont constitués en partie de textes didactiques, puisque ce niveau de spécialisation est susceptible d'être riche en information sémantique, comme l'a démontré Rebeyrolle (2000). Nous y avons ajouté des textes de vulgarisation qui proviennent de sources spécialisées et non d'articles encyclopédiques. Étant donné que la terminologie est bien connue du destinataire, les textes des registres *technique* et *scientifique* de Rebeyrolle (2000) sont peu propices à l'énonciation définitoire. Pour cette raison, le registre *scientifique* a été écarté de notre étude, alors que seulement quelques textes du registre *technique* ont été retenus.

2.2 Relations sémantiques

Cette section porte sur les quatre relations sémantiques retenues pour notre recherche, soit la relation hyperonymique (2.2.1), la relation méronymique (2.2.2), la relation causale (2.2.3) et la relation de finalité (2.2.4). Nous avons retenu ces quatre relations parce qu'elles ont fait l'objet d'un certain nombre d'études terminologiques et qu'elles sont abondamment définies par différents auteurs. Nous croyons également que ces relations sont susceptibles de renfermer beaucoup de particularités propres aux différents types sémantiques (TS) choisis. En effet, des études en extraction d'informations conceptuelles semi-automatique portant sur ces quatre relations ont montré des résultats intéressants (Davidson *et al.* (1998), Marshman (2002), Marshman *et al.* (2002)). De plus, selon Condamines (2003), les relations d'hyperonymie, de méronymie et de causalité sont celles qui sont les plus fréquemment rencontrées et aussi celles qui sont le plus souvent l'objet de travaux de recherche sur les marqueurs :

« Les relations les plus fréquemment rencontrées semblent ainsi être les relations d'hyponymie, de méronymie et de causalité. C'est justement pour ces trois relations que les travaux de recherche de marqueurs sont les plus fréquents, aussi bien en terminologie/linguistique qu'en TAL, cf. par exemple : (Borillo, 1996), (Cabré et al., 1997), (Garcia, 1998), (Jackiewicz, 1996), (Jouis, 1993), (Morin, 1999) » Condamines (2003 : 122).

Outre les quatre relations sémantiques énumérées plus haut, nous nous intéresserons à des patrons un peu particuliers qui n'expriment pas toujours une relation sémantique à proprement parler. Il s'agit des patrons métalinguistiques.

Il importe de souligner que les relations sémantiques ou conceptuelles observées dans les corpus spécialisés ne s'arrêtent pas là. Les relations entre termes sont extrêmement variées comme l'expose Sager (1990) qui a dressé une liste de relations entre termes, accompagnées d'exemples. Cette liste est reproduite dans le Tableau 2.

Tableau 2. Liste des relations conceptuelles de Sager (1990 : 35)

RELATIONSHIPS		EXAMPLES
Cause	Effect	Explosion – fall-out
Material	Product	Steel – girder
Material	Property	Glass – brittle
Material	State	Iron – corrosion
Process	Product	Weaving – cloth
Process	Instrument	Incision – scalpel
Process	Method	Storage – freeze-dry
Process	Patient	Dying – textile
Phenomenon	Measurement	Light – Watt
Object	Counteragent	Poison – antidote
Object	Container	Tool – tool-box
Object	Material	Bridge – iron
Object	Quality	Petrol – high octane
Object	Operation	Drill bit – drilling
Object	Characteristic	Fuel – smokeless
Object	Form	Book – paperback
Activity	Place	Coalmining – coalmine

Nous travaillons sur des patrons de relations sémantiques (PRS) que nous croyons présents dans plusieurs corpus (nous reviendrons sur cette question). Pour cette raison, nous avons écarté les relations qui pourraient être trop spécifiques. Nous avons choisi de nous concentrer sur les quatre relations principales (hyperonymique, méronymique, causale et de finalité) qui sont plus susceptibles de faire ressortir certains liens fondamentaux entre nos corpus. Les sections qui suivent donnent une caractérisation de chacune des relations retenues. Seuls les travaux en terminologie ont été retenus.

On emploie dans la documentation les termes *relations conceptuelles* (par exemple, Sager (1990) et Condamines et Rebeyrolle (1998)) et *relations sémantiques* (par exemple, Ahmad et Fulford. (1992)). L'Homme (2004) explique la différence entre les relations conceptuelles et les relations sémantiques (appelées *relations lexico-sémantiques* par l'auteure) :

« Certaines *relations lexico-sémantiques* sont très proches de relations envisagées dans les représentations conceptuelles. C'est le cas, notamment, de celles qui interviennent dans les taxinomies et les méronymies. Ces relations concernent, le plus souvent, des termes qui renvoient à des entités. Il est en effet extrêmement difficile dans ces cas de faire un réel départ entre ce qui relève du conceptuel et ce qui est lié au lexical. D'autres séries de relations, toutefois, sont résolument lexicales, puisqu'elles engagent, par exemple, des termes appartenant à des parties du discours différentes » (L'Homme 2004 : 90-91).

Ainsi, l'emploi du terme *relation sémantique* plutôt que *relation conceptuelle* dépend, entre autres, du point de vue. En effet, la relation sémantique exprime un lien entre deux sens, alors que la relation conceptuelle met plutôt l'accent sur la structure de la connaissance. Nous retenons l'expression *relation sémantique* plutôt que *relation conceptuelle*, puisque les relations seront dégagées à partir de textes et entre éléments lexicaux.

2.2.1 *Relation hyperonymique*

La relation hyperonymique a beaucoup été étudiée en terminologie. Lyons (1968) la nomme *relation d'inclusion (inclusion)*, alors que Sager (1990) la nomme *relation générique (generic relationship)*.

Marshman *et al.* (2002) disent que cette relation est celle qui a le plus été étudiée en extraction d'information conceptuelle semi-automatique, car elle est à la base de la définition aristotélicienne sur laquelle s'appuient les définitions terminologiques (Flowerdew 1992a : 166; Sager 1990 : 33). Comme le mentionne Hearst, elle pourrait convenir tout particulièrement à l'extraction semi-automatique des connaissances, car en anglais, les patrons qui expriment la relation hyperonymique n'expriment généralement que cette relation; il y a donc peu de risque de bruit (Hearst 1992 : 542).

La relation hyperonymique établit un ordre hiérarchique entre termes d'une même classe sémantique. Bien que cette relation se subdivise en deux relations, il n'est pas rare de voir les auteurs employer une seule appellation générique, c'est-à-dire *hyperonymique*. En effet, la relation hyperonymique doit être considérée sous deux angles : soit celui du terme générique vers le terme spécifique (on parle dans ce cas d'*hyponymie*, donc d'une *relation hyponymique*), et celui du terme spécifique vers le terme générique (que l'on appelle *hyperonymie*, donc d'une *relation hyperonymique*). Dans les exemples 20 et 21, tirés de Marshman *et al.* (2002), l'hyperonyme est signalé par un soulignement double et l'hyponyme par un seul soulignement; le terme recherché est en italique.

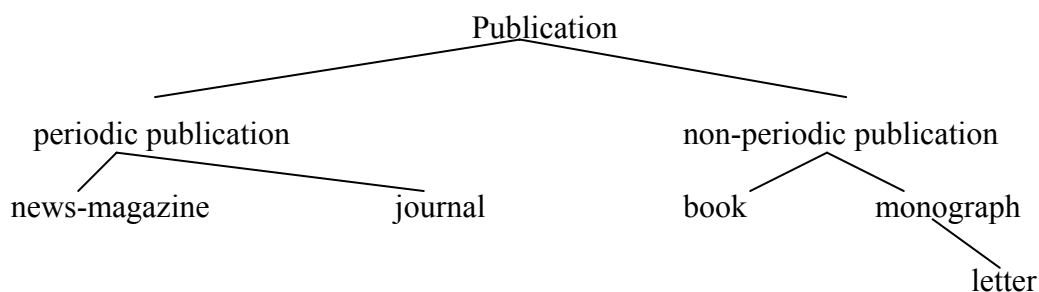
20. Hyperonymie : La gyrase est une topoisomérase de type II. (pour le terme recherché *gyrase*, on trouve l'hyperonyme *topoisomérase*).
21. Hyponymie : [Il y a] trois types de cartes à microcircuit : la carte à simple mémoire, la carte à logique câblée et la carte à microcalculateur (pour le terme *carte à microcircuit*, on trouve les hyponymes *carte à simple mémoire*, *carte à logique câblée* et *carte à microcalculateur*).

Comme le souligne Morgan (2000), la relation hyperonymique établit un ordre hiérarchique, l'hyponyme héritant des caractéristiques de l'hyperonyme :

« Hyperonymy establishes an hierarchical order in that it relates the hypernym, or generic term, to a hyponym, or more specific one. The hyponyms inherit the characteristics of the hypernym, but have at least one differentiating characteristic » (Morgan 2000 : 9).

Les liens hiérarchiques entre hyperonymes et hyponymes sont souvent représentés graphiquement, comme l'illustre la figure 2 :

Figure 2 : Relations hyperonymiques représentées dans une arborescence (Sager 1990 : 31)



Sager utilise les paraphrases anglaises suivantes pour valider les relations hyperonymiques et hyponymiques :

« X is a type of A.
 Or X, Y and Z are types of A.
 Or A has the specific concepts X, Y and Z.
 Or A has the subtype X » (Sager 1990 : 30).

Les paraphrases de Sager permettent de valider les relations d'hyperonymie et d'hyponymie partagées par les termes *letter* (hyponyme) et *monograph* (hyperonyme), *books* and *monographs* (hyponymes) et *non-periodic publications* (hyperonymes), *non-periodic publications* (hyperonyme) et *books*

and *monographs* (hyponymes), et *monographs* (hyperonyme) et *letters* (hyponyme).

A letter is a type of monograph

Or Books and monographs are types of non-periodic publications

Or Non-periodic publications have the specific concepts books and monographs

Or Monographs have the subtype letters (d'après Sager 1990 : 30).

En français, une des paraphrases permettant de valider la relation hyperonymique est : *Tous les X sont des Y*, où X est l'hyponyme et Y est l'hyperonyme. De même, une paraphrase permettant de valider la relation hyponymique prend la forme suivante : *X comprend les spécifiques Y et Z*, où X est l'hyperonyme et Y et Z sont les hyponymes.

2.2.2 *Relation méronymique*

La relation méronymique est aussi connue sous le nom de *relation partitive* ou de *relation partie – tout* (whole – part, Sager 1990). Loffler-Laurian décrit un type de relation qu'elle appelle *analyse*, et lui prête une extension qui semble s'appliquer à la relation méronymique.

« L'analyse à proprement parler est la décomposition d'un objet en éléments tels que si l'on ajoute tous les éléments les uns aux autres, on obtient l'objet en question. Pour étendre cette définition, on peut considérer comme analyse la somme des propriétés de l'objet à définir » (Loffler-Laurian 1983 : 17).

Cruse souligne l'importance de cette relation parfois mise de côté :

« There is no doubt about the central importance of fully integrated and cohesive physical objects, with well-differentiated parts, in the concept of 'part' and 'whole' » (Cruse 1986 : 157).

Selon Sager (1990 : 32), ce type de relation exprime « the connection between concepts consisting of more than one part and their constituent parts ».

L'Homme (2004) mentionne que les relations méronymiques sont fondées sur les notions vagues de « proximité » ou d'« association dans l'espace », et qu'elles engagent un terme dénotant un tout et un ou plusieurs termes dénotant des parties.

Cette relation, tout comme la relation hyperonymique (et la relation inverse, à savoir l'hyponymie) peut aussi être vue sous deux angles, soit du tout vers la partie (méronymie), soit de la partie vers le tout (holonymie). Les exemples 22 et 23, tirés de nos corpus, illustrent ce propos. Les termes désignant le tout sont signalés par un seul soulignement, alors que les termes désignant la partie sont signalés par l'italique.

Méronymie

22. En effet, qu'il s'agisse de micro-ordinateurs, mini-ordinateurs ou gros ordinateurs, tous disposent d'une unité centrale de traitement, d'une *mémoire principale*, d'*unités de mémoire secondaire* et d'*équipements d'entrée-sortie* (pierre1).

Holonymie

23. Tous les *utilisateurs* qui font partie de ce groupe recevront les droits du groupe (linux3p1).

La relation méronymique est une relation importante, certes, mais complexe, car elle se fragmente en différents sous-types. Certains auteurs, dont Iris *et al.* (1988), sont d'avis que cette relation est si complexe qu'elle doit être vue comme un ensemble de relations : « the part – whole relation should be treated as a collection of relations, not as a single relation » (Iris *et al.* 1988 : 262).

Cette variété de relations méronymiques explique qu'il existe une multitude de structures linguistiques pour exprimer la méronymie.

Winston *et al.* (1987 : 421) ont élaboré une taxonomie de relations méronymiques qui comprend six catégories :

- « component/integral object, e.g.: *handle/cup* »

- « member/collection, e.g.: *tree/forest* »
- « portion/mass, e.g.: *grain/salt* »
- « stuff/object, e.g.: *steel/bike* »
- « feature/activity, e.g.: *dating/adolescence* »
- « place/area, e.g.: *oasis/desert* ».

Nous reproduisons et expliquons ci-dessous le découpage des relations méronymiques proposé par Winston *et al.* (1987) et repris par Otman (1996).

partie fonctionnelle – tout

Dans cette relation, le tout est composé de parties détachables qui diffèrent les unes des autres. La partie peut difficilement être retranchée sans que cela nuise au fonctionnement du tout. Elle joue donc un rôle fonctionnel par rapport au tout.

élément – ensemble

Dans cette relation, le tout est composé de parties détachées qui sont toutes semblables les unes aux autres.

portion – masse

Dans cette relation, la partie retient toutes les propriétés du tout. La principale particularité de cette relation est donc de permettre de remplacer le méronyme par l'holonyme étant donné son caractère homéomère. La notion « homéomère » est définie par Van Campenhoudt de la manière suivante :

« Les parties dites homéomères sont matériellement identiques entre elles et par rapport au tout : un morceau de tarte est encore de la tarte, un grain de sel est encore du sel, etc. » (Van Campenhoudt 2004 : page Internet).

constituant – objet

Cette relation se caractérise par le fait que le constituant est la matière dont est fait le tout et ne peut être détachée ou séparée de lui.

phase – activité

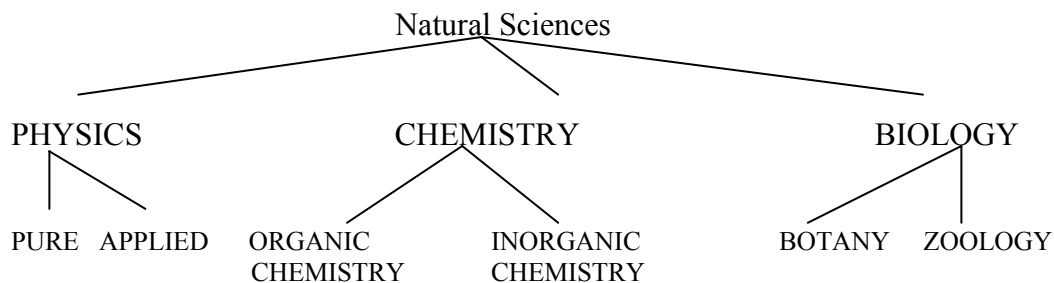
Cette relation concerne les activités et non les entités. Celles-ci peuvent être découpées en parties distinctes appelées phases, qui se succèdent les unes aux autres.

lieu – zone

Cette relation concerne la localisation de la partie dans le tout.

Les relations méronymiques peuvent, tout comme la relation hyperonymique, être représentées dans une arborescence. L'exemple reproduit à la figure 3 est tiré de Sager.

Figure 3 : Relations méronymiques représentées dans une arborescence (Sager 1990 : 32)



Sager utilise les paraphrases suivantes pour valider la relation méronymique :

« X is a constituent part of Y.
Or X, Y and Z are constituent parts of A.
Or A consists of X.
Or A consists of X, Y and Z » (Sager 1990 : 32).

Les paraphrases de Sager permettent de valider les relations de méronymie et d'holonymie partagées par les termes *botany* (méronyme) et *biology* (holonyme), *botany* and *zoology* (méronymes) et *biology* (holonyme), *chemistry* (holonyme) et *organic chemistry* and *inorganic chemistry* (méronymes).

Botany is a constituent part of biology
Or Botany and zoology are constituent parts of biology
Or Chemistry consists of organic chemistry and inorganic chemistry
 (d'après Sager 1990 : 32).

En français, une paraphrase qui permet de valider la relation méronymique est : *X a comme parties Y Z*, où X est l'holonyme et Y et Z sont les méronymes. Par ailleurs, une paraphrase qui permet de valider la relation holonymique est la suivante : *X est un constituant de Y*, où X est le méronyme et Y est l'holonyme.

Selon Marshman *et al.* (2002), la relation partie – tout peut être exprimée par des patrons du type *consists of*, *part of* et *includes* en anglais, et *consiste en*, *partie de* et *comporte* en français.

Selon Loffler-Laurian (1994), les mots types exprimant une relation méronymique sont *comprendre*, *contenir*, et les expressions les plus fréquentes sont *être composé de*, *être constitué de*, *se composer de*.

2.2.3 *Relation causale*

La relation causale lie une cause à son effet. Certains travaux portant sur la recherche de cette relation dans les corpus spécialisés se sont appuyés sur un

modèle proposé par Talmy (1985, 1988, 2000), qui s'inscrit dans le domaine des sciences cognitives. Le modèle de Talmy, qui s'appuie lui-même sur la théorie de la *force dynamique*, consiste à envisager l'interaction de forces opposées. Les forces sont appelées *agoniste (ago)* et *antagoniste (ant)*.

L'auteur a étudié leur interaction, tant du point de vue de la force dynamique (dans le cas où la puissance relative des forces change) que statique (dans le cas où la puissance relative des forces demeure la même). Le modèle de Talmy propose huit possibilités d'interaction entre forces opposées qui peuvent être utilisées pour le classement de la majorité des occurrences de la relation causale. Cependant, étant donné que nous ne poussons pas à ce point l'étude de cette relation pour notre recherche, nous nous contentons de résumer l'exploitation qui a été faite du modèle dans les travaux de terminologie. La représentation de la relation causale décrite par Talmy est intuitive et peut être appliquée à des données linguistiques. Quant aux autres représentations de la relation causale décrites ci-dessous, l'objectif des auteurs était de développer et de transposer le cadre théorique de la relation en mots en vue de l'extraction des connaissances.

Nuopponen (1994) a travaillé à l'élaboration d'une classification de la relation causale. L'auteure était intéressée par ses applications en terminologie. Dans plusieurs langues de spécialité, dont la médecine, le droit, la physique et la biologie, la relation causale est fortement représentée. Dans ces domaines, il est essentiel de rechercher les causes et les effets ainsi que les relations (forcément causales) qui les lient.

« In many subject fields causality is an important factor, and finding the causes and effects, and the relations between these, is essential. Examples might be medicine, law, physics, biology, etc. In medicine, for instance, the questions asked might include the following:

- What caused this disease?
- What are the complications of this disease?
- What effects does this medicine have?
- What side effects does this treatment have?
- How can we prevent this disease? »

... (Nuopponen 1994 : 36)

L'auteure divise la **cause** en quatre composants : a) agent causal (*causal agent*), b) processus causal (*producing cause*), c) cause explicative (*explanatory cause*) et d) effet neutralisant (*counteracting cause*).

- L'agent causal renvoie à des substances, à des matériaux ou à d'autres éléments qui causent un effet.
- Le processus causal peut être un événement, une action ou tout autre processus.
- La cause explicative est un fait ou un état.
- Un effet neutralisant est un agent, un événement, un état ou un fait qui neutralise le processus causal et prévient l'effet.

Enfin, l'auteure subdivise l'effet en trois principaux composants : a) état, telle une maladie ou un dommage; b) produit, telle la corrosion qui produit de la rouille; c) événement, telle la vaccination qui conduit à l'immunisation. Quant aux symptômes d'une maladie, ils font partie intégrante de l'effet et peuvent appartenir aux trois catégories qui précèdent. Un quatrième composant s'ajoute aux trois principaux, soit celui des d) complications, c'est-à-dire les effets causés par le premier effet. Par exemple, la rougeole peut être suivie de problèmes respiratoire ou neurologique. Enfin, certaines causes dites « neutralisantes », comme un médicament ou une opération, peuvent être liées aux effets afin de les éliminer.

Garcia (1996, 1997) a étudié la relation causale en profondeur. Elle a élaboré une hiérarchie de la relation causale, en s'inspirant de Talmy (1988) et mis au point COATIS, un outil d'aide à l'acquisition de connaissances causales par exploration de textes de langue française.

Garcia a étudié les relations causales exprimées principalement par les verbes et présentes dans les textes techniques. Deux familles de relations ayant

chacune leurs propriétés se détachent, soit la relation causale formelle (illustrée en 24) et la relation causale efficiente (illustrée en 25).

24. « Le niveau de la puissance produite par l'usine varie en fonction de l'hydraulicité » (Garcia 1996 : 98).
25. « Cette opération facilite l'implantation des futurs ouvrages » (Garcia 1996 : 98).

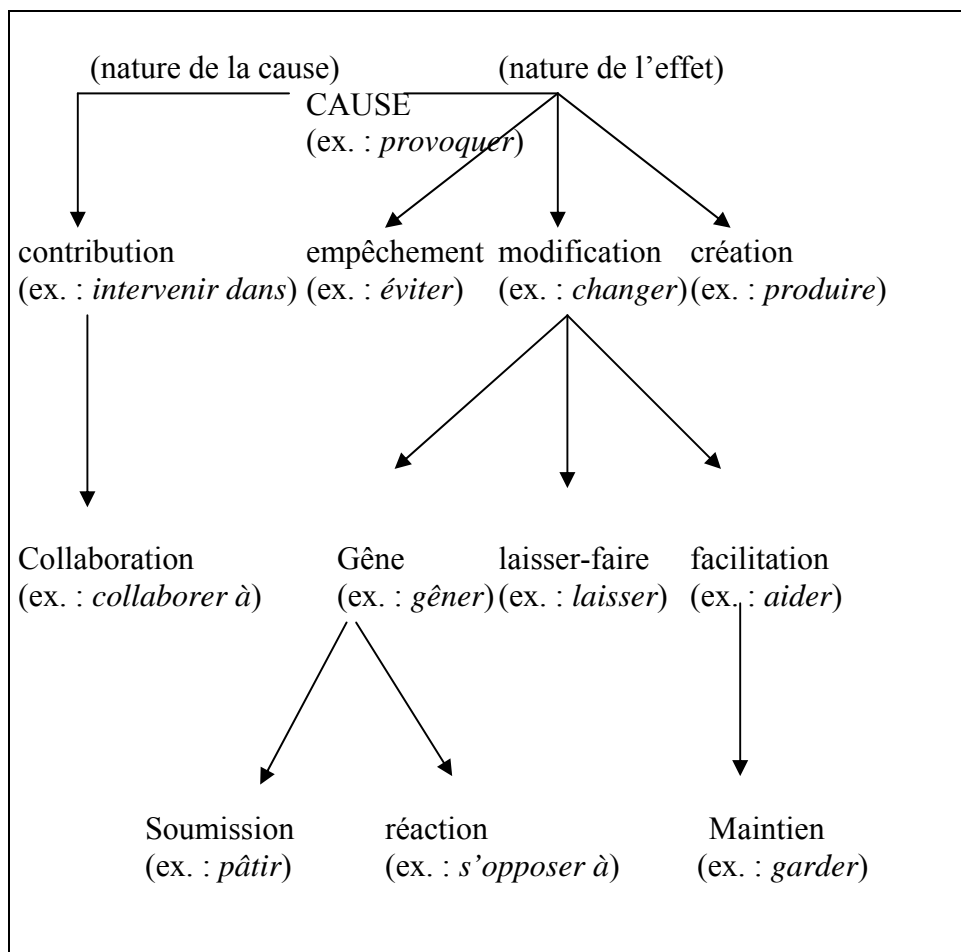
La cause efficiente produit un effet qui lui est distinct : cette relation causale est orientée dans le temps. Les indicateurs linguistiques qui traduisent les relations efficientes peuvent exprimer simplement, un rapport causal (26), préciser la nature de l'effet produit (27) ou préciser la nature de l'action causale (28).

26. « En France, l'élevage est la source agricole la plus importante de rejet de méthane » (Garcia 1996 : 98).
27. « La mise en parallèle de différents types d'ouvrages entrave le fonctionnement du réseau » (Garcia 1996 : 98).
28. « La longueur maximale du départ le plus long ramené à moins de 90 km contribue à améliorer la qualité de service » (Garcia 1996 : 98).

Puisque les textes techniques que Garcia (1996) a étudiés se caractérisent par l'expression de la causalité efficiente, l'auteure n'a pas privilégié la relation causale formelle et ne donne pas plus de détails sur cette relation.

La figure 4 présente un échantillon des relations causales efficientes que l'auteure a distinguées.

Figure 4 : Extrait des relations causales efficaces (Garcia 1997 : 11).



La figure 4 présente les relations causales efficaces qui découlent de la primitive *Cause*. L'arbre montre, à partir du premier fil droit, les causalités qui qualifient l'effet produit, et, à partir, du premier fil gauche, certaines causalités précisant la nature de l'action causale. À chacune des relations, le modèle associe une classe de verbes qui expriment, dans les textes, la causalité.

Quant au travail sur COATIS, il a amené l'auteure à étudier en priorité les indicateurs de causalité qui sont des verbes. L'auteure a aussi considéré, mais dans une moindre mesure, les relations causales qui s'expriment dans les textes par d'autres indicateurs (noms, conjonctions).

Barrière (2001, 2002) a adapté l'analyse de Talmy (1988, 2000) et a mis au point une hiérarchie de la relation causale, en la divisant en deux grandes catégories, chacune étant subdivisée encore plus finement : soit les dépendances existantes (création, destruction, maintien et prévention) et influente (modification et préservation).

Comme le montre le Tableau 3, l'effet de l'interaction de la force a été étendu afin de représenter non seulement un objet au repos ou en mouvement, mais aussi l'existence ou la non-existence d'une entité et l'existence ou la non-existence d'un événement (dépendance existante) ainsi que les changements dans le caractère d'un événement ou d'une entité (dépendance influente). Le Tableau 3 présente également des verbes susceptibles d'exprimer ce type de relation. Dans la colonne de gauche, on trouve le type de dépendance; dans la colonne de droite, sa sous-catégorie. La troisième colonne indique l'effet de l'interaction entre les forces opposées. Enfin, la quatrième et dernière colonne recense quelques verbes anglais pouvant exprimer une sous-catégorie de la relation causale.

Tableau 3. Adaptation par Barrière (2001-2002) de la classification de la relation causale de Talmy (1988, 2000), et simplifiée par Marshman (2002 : 16)

Dépendance existante	Création	$\sim E \rightarrow E$	Create Generate Produce
	Destruction	$E \rightarrow \sim E$	Kill Eliminate Destroy
	Maintien	$E \rightarrow E$	Allow Keep Maintain
	Prévention	$\sim E \rightarrow \sim E$	Prevent Discourage Control
Dépendance influente	Modification	$E_1 < > E_2$	Influence Change Modify
	Augmentation	$E_1 < E_2$	Increase Improve Promote Enhance
	Diminution	$E_1 > E_2$	Reduce Decrease Shorten Slow down Deter Discourage
	Conservation	$E_1 = E_2$	Maintain Keep Retain

Légende : E peut exister (E) ou non ($\sim E$) et, s'il existe, il aura des caractéristiques ou agents modificateurs ($f_{1...N}$).

Nous décrivons ci-dessous chacune des sous-catégories de la relation causale (en nous inspirant des explications données dans Marshman 2002).

Dépendance existante :

Cette catégorie désigne des relations causales qui déterminent l'existence ou la non-existence d'une entité ou d'un événement. Cette dépendance est divisée en quatre sous-types : création, destruction, maintenance et prévention.

Création :

Ce type de relation causale survient lorsque l'interaction entre les forces opposées fait naître une entité qui n'existait pas auparavant ou cause un événement qui n'avait pas lieu auparavant. Le contexte 29 en présente un exemple.

29. Le virus varicelle-zona (VZV)¹⁰ est **responsable de** la *maladie* [zona ophthalmique] réalisant une ganglioradiculite postérieure par atteinte du ganglion de Gasser (oph-mg).

Destruction :

Cette relation constitue l'opposé de la création. Elle survient lorsque l'interaction entre les forces opposées est la cause de la fin de l'existence d'une entité ou d'un événement. Le contexte 30 illustre cette relation causale.

30. En effet une *panne* de ventilateur peut directement **causer** la mort de votre processeur (optrep).

¹⁰ Comme nous le mentionnions au chapitre 1, dans l'ensemble de la thèse (à moins d'indication contraire), le terme de la requête apparaît en italique et le PRS en caractères gras, tandis que l'information sémantique qui se rapporte au terme est soulignée.

Maintien :

Cette relation exprime une situation dans laquelle une entité ou un événement existait déjà avant l'interaction des forces opposées et continue d'exister ou se poursuit après. Le contexte 31 présente un exemple de la relation causale de maintien.

31. Au contraire, l'Internet est un standard simple, qui n'existe que **grâce au développement collectif**, effectué dans la transparence et la recherche de l'efficacité plutôt que par l'emballage commercial (société).

Prévention :

En prévention, une entité ou un événement n'existait pas avant l'interaction entre les forces opposées et continue à ne pas exister ou ne pas se produire. Puisque nos corpus ne contiennent aucun exemple de cette relation causale, nous présentons en 32 un exemple tiré de Marshman (2002).

32. ... ces *composés* agissent lors de la phase précoce de multiplication virale en **empêchant** l'acidification de l'intérieur de la particule virale... (Marshman 2002 : 81).

Dépendance influente :

Cette catégorie désigne des relations où l'interaction entre forces opposées détermine certaines caractéristiques d'une entité ou d'un événement. Cette caractéristique peut être la modification, l'augmentation, la diminution ou la conservation.

Modification :

Dans ce type de relation causale, l'interaction des forces a un effet sur une caractéristique d'une entité ou d'un événement. Cependant, le type de modification peut ne pas être spécifié. Le contexte 33 en présente un exemple.

33. La quinidine **modifie** l'*électrocardiogramme*, allongeant les espaces PR, QRS et surtout QT, l'espace ST restant normal (zschmit).

Augmentation :

Dans ce type de relation causale, la caractéristique de l'entité ou de l'événement est intensifiée ou augmentée par l'interaction entre les forces opposées. Le contexte 34 en présente un exemple.

34. **Avec** le *vieillissement* de la population, les problèmes neurologiques du sujet âgé sont de plus en plus fréquents et leur analyse justifie souvent la réalisation d'une imagerie (image1).

Diminution :

Dans cette relation causale, qui constitue l'image miroir de l'augmentation, la caractéristique de l'entité ou de l'événement est diminuée par l'interaction entre les forces opposées, comme le montre l'exemple 35 :

35. Le plus grand **facteur de** ralentissement est le *développement* de fonctions inutiles, c'est ce que B. Boehm appelle le gold platulg (la dorure), très fréquent dans les grands projets (genilo).

Conservation :

Cette relation causale est analogue à la relation *maintien*. En effet, la caractéristique de l'entité ou de l'événement existe préalablement à l'interaction des forces et continue d'exister de la même façon après. Cependant, cette relation inclut aussi ces cas où une caractéristique n'est pas présente ni avant ni après l'interaction entre les forces. Le contexte 36 illustre la conservation.

36. Le développement technique de l'Internet bénéficie d'une dynamique originale. Cette dynamique **repose sur** [...] le *développement* et le déploiement accélérés des standards qui définissent l'Internet et ses applications (missio).

Comme le mentionne Barrière, c'est le contexte dans son ensemble, et non seulement le patron, qui permettra le classement adéquat de la relation causale :

« It is clear that some indicators are highly linked to a particular dependency category and we can assume with some confidence that an instance indicates that type of causal relation. For example *improve* and *increase* express the [increase] subrelation, while *affect* and *depend* communicate [modification] and *produce* [creation]. Some indicators, on the other hand, are not very informative. We see that instances of *achieve* fall into the general [influence dependency] category, and *help* and *result* express a variety of [causal] relations, as do many conjunctive patterns. This means that information from elsewhere in the sentence will be needed for categorization. Even for patterns with dual dominants – *control*, *inhibit*, *reduce* and *allow* – automatic extraction will need other clues from the sentence to assign a category » Barrière (2002 : 102-103).

En français, les paraphrases qui permettent de valider cette relation sont : *X cause Y*, où X est l'agent causal et Y est l'effet, comme dans *Le diabète cause la rétinopathie diabétique*, ou *X dû à Y*, où X est l'effet et Y est l'agent causal, comme dans *Cette maladie de la substance blanche est due à une artériosclérose des artères distales*.

2.2.4 Relation de finalité

Nous prêtons, à la relation de finalité, l'extension suivante : relation qui exprime la fonction d'une entité. Cette relation est aussi appelée *relation fonctionnelle (functionality)* par Miller qui la décrit ainsi :

« A functional feature of a nominal concept is intended to be a description of something that instances of the concept normally do, or that is normally done with or to them » Miller (1990 : 257).

Comme le mentionne Miller, ce type de relation se manifeste surtout dans le cas d'objets fabriqués :

« Particularly among the human artifacts there are things that have been created for a purpose... [and they can be] define both by structure and use » Miller (1990 : 259).

Un des cinq types de définitions utilisées par les différents textes scientifiques que distingue Loffler-Laurian (1994) porte sur la finalité, qu'elle appelle *fonction*. Ce type de définition est le propre de la vulgarisation. Elle la

caractérise ainsi : « Ce type de définition s'appuie sur la finalité de l'objet, sur son usage et ses possibilités, éventuellement sur ses effets » (Loffler-Laurian 1983 : 18). Selon l'auteure, les mots types exprimant cette relation sont les suivants : *capable de, permettant de, ... permettre de, employer pour, utiliser pour, conduire à, etc.*

En français, les paraphrases qui permettent de valider cette relation sont : *X sert à Y*, où X est l'entité et Y est la fonction qu'accomplit ou que peut accomplir cette entité, comme dans *Une imprimante sert à imprimer des documents* ou *l'entité X est utilisée pour accomplir la tâche Y*, où X est l'entité et Y est la fonction qu'accomplit ou que peut accomplir cette entité, comme dans *On utilise un ordinateur pour résoudre des problèmes symboliques*.

Enfin, Marshman *et al.* (2002) montrent que la relation de finalité peut être exprimée en anglais par les patrons suivants : *is used for, allows and function*, et, en français, *utilisé pour, permet et fonction*.

2.2.5 Synthèse

En résumé, nous avons discuté et défini les quatre principales relations sémantiques retenues, soit les relations hyperonymique, méronymique, causale et de finalité. En nous basant sur les définitions de différents spécialistes, nous avons pu tracer un portrait des caractéristiques de chacune.

Dans cette section, nous expliquons comment nous interprétons les relations sémantiques¹¹ dans notre étude et donnons les paraphrases utilisées pour les valider.

- Relation hyperonymique : Relation qui lie un spécifique à son générique ou un générique à un spécifique. Bien que, comme nous l'avons souligné dans la section 2.2.1, cette relation puisse être subdivisée en relations plus

¹¹ Afin d'alléger le texte, nous parlerons de *relations sémantiques*, mais elles comprennent souvent les patrons métalinguistiques.

spécifiques, nous conserverons l'appellation générique, *relation hyperonymique*. Pour valider cette relation, nous utiliserons les paraphrases *Tous les X sont des Y*, où X est l'hyponyme et Y est l'hyperonyme, et *X comprend les spécifiques Y et Z*, où X est l'hyperonyme et Y et Z sont les hyponymes.

- Relation méronymique : Relation qui exprime le lien entre un tout et ses parties. Bien que, comme nous l'avons souligné dans la section 2.2.2, cette relation puisse être subdivisée en relations plus spécifiques, nous conserverons l'appellation générique, *relation méronymique*. Nous utiliserons les paraphrases *X a comme parties Y Z*, où X est l'holonyme et Y et Z sont les méronymes et *X est un constituant de Y*, où X est le méronyme et Y est l'holonyme afin de valider cette relation. Nous tiendrons compte des six subdivisions de Winston *et al.* (1987), soit : partie fonctionnelle – tout, élément – ensemble, portion – masse, constituant – objet, phase – activité et lieu – zone.
- Relation causale : La relation causale lie une cause à son effet. Les paraphrases *X cause Y*, où X est l'agent causal et Y est l'effet, ou *X dû à Y*, où X est l'effet et Y est l'agent causal, sont employées pour valider cette relation. Dans l'analyse que nous ferons des contextes contenant une expression d'une relation causale, nous tiendrons compte des subdivisions de Barrière (2001, 2002), d'après l'analyse de Talmy (1988, 2000) : modification (diminution et augmentation), création, destruction, maintien, prévention et conservation.
- Relation de finalité : Relation qui exprime l'utilité d'une entité. Les paraphrases *X sert à Y*, où X est l'entité et Y est la fonction qu'accomplit ou que peut accomplir cette entité, ou *l'entité X est utilisée pour accomplir la tâche Y*, où X est l'entité et Y est la fonction qu'accomplit ou que peut accomplir cette entité, sont employées pour valider cette relation.

Outre ces quatre relations principales, nous avons étudié les contextes qui contiennent un patron métalinguistique. Nous les plaçons à part, car il ne s'agit pas toujours de contextes exprimant une relation sémantique à proprement parler. De plus, nous supposons que les patrons métalinguistiques sont valables pour tous les types sémantiques, et il est peu probable qu'ils se démarquent pour un TS donné. Nous avons quand même choisi de les étudier en raison de leur très grande utilité dans le travail terminologique.

Rappelons simplement que « le métalangage est défini comme un discours qui a pour thème le langage » (Auger 1998 : 44). Rey-Debove (1978), qui a étudié les propriétés de la fonction métalinguistique, a recensé des verbes et expressions métalinguistiques comme ceux qui suivent : *désigner, s'appeler, signifier, se dire, autement dit, c'est-à-dire, ou*.

Les patrons métalinguistiques qui nous intéressent sont des indices repérables dans les textes qui prennent la forme d'un verbe ou d'une expression plus longue. Les exemples 37 et 38 sont tirés respectivement de nos corpus de médecine et d'informatique.

37. Ce *traitement*, **appelé** photocoagulation, consiste à focaliser la lumière du laser sur les vaisseaux qui laissent passer le fluide dans la macula. (ret_diab).
38. On appelle compilateur un logiciel système dont la fonction essentielle est de traduire en langage machine ou en code objet (binaire) un programme source écrit en *langage évolué*, **c'est-à-dire** proche du langage humain, tout en préservant le sens et la structure logique de ce programme (pierre1).

2.3 Information sémantique dans les corpus spécialisés

Cette section est consacrée aux travaux antérieurs qui ont porté sur le repérage automatique ou semi-automatique d'informations sémantiques dans les corpus. Nous présentons d'abord la multitude d'appellations qui ont été données à ce que nous appelons *patron de relation sémantique* (PRS). Nous présentons un

tableau de synthèse dans lequel apparaissent de brèves définitions et les différentes appellations de *PRS* proposées par les auteurs (2.3.1). Ensuite, nous décrivons en détail les travaux les plus près du nôtre, c'est-à-dire, les travaux terminologiques qui ont porté sur l'extraction d'informations sémantiques à partir de corpus (2.3.4). Il en existe d'autres, dont les travaux de Barnbrook et Sinclair (1995) portant sur les articles de dictionnaire (textes généraux) que nous citons à la section 2.3.1 sans les détailler toutefois.

2.3.1 *Noms donnés aux PRS selon les chercheurs*

Plusieurs noms ont été donnés à la notion de « PRS ». Davidson *et al.* dressent une liste d'appellations anglaises utilisées par différents chercheurs :

« These linguistic patterns¹² have been designated by various terms in English, depending on the researcher : Lyons (1977) uses *formulae*; Cruse (1986) refers to *diagnostic frames* or *test frames*; Winston *et al.* (1987) simply use *frames*; Flowerdew (1992a) refers to *linguistic structures* that make certain definitional information salient (and breaks these down into *boosters* and *downgraders*); Pearson (1996) refers to *hinges*; Ahmad et Fulford (1992) call them *knowledge probes* » (Davidson *et al.* 1998 : 51).

Cette citation met en évidence les différentes appellations utilisées, mais par la même occasion, elle montre aussi que nombre de chercheurs s'intéressent à la question. Nous avons reproduit dans le Tableau 4 les appellations utilisées, l'auteur ou les auteurs qui les emploient, et l'extension qu'ils leur prêtent. Comme on le verra, ces termes ne sont pas interchangeables; les chercheurs choisissent, de façon générale, une appellation particulière en fonction des objectifs de leur étude.

¹² Meyer *et al.* (1999) ont changé d'appellation; ils utilisent maintenant *knowledge-rich patterns* (patrons riches en connaissances) ou *knowledge patterns* (patrons de connaissances).

Tableau 4. Tableau de synthèse contenant les définitions des différents noms donnés aux PRS selon les chercheurs

Auteurs	Appellations	Distinctions des PRS
Ahmad <i>et al.</i> (1992)	<u>Knowledge probes</u>	Ce sont des items lexicaux substituables servant à extraire des relations sémantiques dans les textes. Ex. : <i>X is Y, X is a kind of Y, X causes Y</i> , etc. Leur recherche dans un texte permet de trouver des relations sémantiques et permettra de découvrir les relations entre les termes dans un domaine en particulier.
Auger (1998)	<u>Énoncés définitoires (énoncés définitoires métalinguistiques et énoncés définitoires linguistiques)</u>	Discours de forme phrastique qui détermine, en tout ou en partie, les caractères qui appartiennent à un concept, les éléments qui caractérisent une chose ou encore un aspect propre à un signe linguistique (S ^{ant} ou S ^é). Le vocabulaire particulier de l'énoncé définitoire explicite cette détermination. Ex. : <i>signifier, désigner, s'écrire</i> . Les énoncés définitoires métalinguistiques comportent un sujet ou un terme défini [N _o] qui est autonome et des mots métalinguistiques. Ex. : <i>Un X est un Y qui, c'est-à-dire, permettre, servir à</i> . Un énoncé définitoire linguistique est référentiel; il ne comporte pas de mot métalinguistique ni de sujet autonome. Ex. : <i>désigner, vouloir dire, dénommer, s'appeler</i> .
Barnbrook et Sinclair (1995)	<u>Hinge</u>	Un certain nombre de verbes ou d'expressions verbales comme ces patrons ont une relation d'égalité '='. Le terme <i>hinge</i> est utilisé par Barnbrook et Sinclair (1995 : 21) pour décrire une partie de la structure d'une définition du Dictionnaire Cobuild. Par exemple, cette « charnière » est constituée d'un mot ou d'une expression comme <i>as X is/are Y, X were Y, X means Y, X consists of Y</i> , etc. lorsque utilisée pour introduire une définition.
Condamines et Rebeyrolle (1998)	<u>Conceptual relationship patterns, linguistic patterns, linguistic markers, semantic and lexico-syntactic patterns</u>	Ces patrons montrent les relations conceptuelles existantes. Ex. : <i>X is composed of Y, X is the result of Y, X is responsible for Y</i> , etc.
Cruse (1986)	<u>Diagnostic frames</u>	Paraphrases qui servent à valider une relation particulière. Par exemple, une relation d'hyponymie peut être exprimée par des expressions telles que <i>X is a type of Y, X is a species of Y</i> , etc.

Davidson <i>et al.</i> (1998), Meyer <i>et al.</i> (1999), Marshman (2002) et Marshman <i>et al.</i> (2002)	<u>Knowledge-rich contexts, contextes riches en connaissances</u> (traduction des auteurs)	Contextes qui illustrent des relations conceptuelles dans un domaine de spécialité. Ces contextes sont appelés « <i>knowledge-rich</i> ». Ex. : <i>Compost is a practical method of recycling organic matter and reducing solid waste</i> , où <i>is a</i> est un patron riche en connaissances et l'information qui se rapporte à <i>compost</i> est le contexte riche en connaissances. La relation exprimée dans cet exemple en est une d'hyperonymie (Synonymes : voir la définition de <i>metalanguage patterns</i> ou de <i>defining expositives</i> de Pearson 1998).
Davidson <i>et al.</i> (1998) et Meyer <i>et al.</i> (1999), Marshman (2002) et Marshman <i>et al.</i> (2002)	<u>Knowledge-rich patterns, knowledge patterns, patrons riches en connaissances, patrons de connaissances</u> (traduction des auteurs)	Patrons associés aux relations conceptuelles qui peuvent être classés sous trois catégories : 1) Patrons lexicaux (ex. : <i>X is a Y, X is known as Y, etc.</i>); 2) Patrons grammaticaux (ex. : <i>TERM + VERB : X transforms Y</i> , ou <i>ADJECTIVE + NOUN : ex. : a writable disk</i>); 3) Patrons paralinguistiques : (ex. : les parenthèses).
Flowerdew (1992a)	<u>Boosters et downtoners</u>	Les <i>boosters</i> sont des items linguistiques qui indiquent clairement la force illocutionnaire de l'acte de parole (Flowerdew 1992a : 172). Ex. : <i>we call X Y, X is called Y, X is known as Y, etc.</i> Quant aux <i>downtoners</i> , ils peuvent diminuer la « notoriété » d'une définition. Ex. : <i>X is one way of defining Y, etc.</i>
Harris (1982)	<u>Patterns of words</u>	Les <i>knowledge probes</i> d'Ahmad et Fulford (1992) sont des éléments de l'environnement de la langue générale. C'est dans cet environnement que les éléments de la LSP d'un texte prennent place. Selon Harris, les <i>knowledge probes</i> peuvent aussi être considérés comme des patrons (<i>patterns of words</i>) récurrents en LSP. Ces <i>patterns of words</i> se distinguent des <i>knowledge-rich patterns</i> , entre autres, parce qu'ils excluent les patrons paralinguistiques .
Jacobi (1994)	<u>Marqueurs métalinguistiques (forme d'auto-reformulation intradiscursive)</u>	Série d'opérations dans le discours (auto-reformulation intradiscursive) destinées en principe à aider les lecteurs à accéder au sens; mécanismes de reformulation dans les documents de vulgarisation scientifique. Ex. : <i>X en d'autres termes Y, X c'est-à-dire Y, X est ainsi baptisé en Y, X que l'on appelle Y, etc.</i>
Loffler-Laurian (1994)	<u>Définitions</u>	« Les définitions de la vulgarisation scientifique ont un sémantisme tel que le définissant donne : la description de l'objet

		par ses caractéristiques ou ses parties constitutives; l'histoire de l'objet ou de sa désignation, son mode de formation ou de fabrication; l'avenir de l'objet, sa fonction ou sa finalité, ses effets ». Ex. : <i>X est capable de Y, X englobe Y, X est Y</i> , etc. (Loffler-Laurian 1994 : 111). En plus d'étudier les PRS , cette auteure a étudié également la liaison par ponctuation comme Meyer <i>et al.</i> (1999) (patrons paralinguistiques), entre autres, mais aussi l'illustration .
Lyons (1977)	<u><i>Formulae</i></u>	Le terme <i>formulae</i> est employé pour décrire les patrons lexicaux utilisés pour exprimer des relations conceptuelles. Les <i>formulae</i> témoignent de la relation entre au moins deux items lexicaux. Par exemple, une relation d'hyponymie peut généralement être exprimée par la formule <i>X is a kind of Y</i> .
Pearson (1998)	<u><i>Connectives, connective verb, connective phrase</i></u>	Puisque le terme <i>hinge</i> (voir explications ci-dessous sous Barnbrook et Sinclair (1995)) a deux sens, Pearson a choisi d'utiliser un terme plus général, soit celui de <i>connectives</i> (connectifs) pour décrire tous les verbes et les expressions qui lient un terme à l'information qui s'y rattache. Elle utilise le terme de <i>connective verbs</i> pour décrire les verbes et les expressions verbales qui sont utilisés pour lier un terme et un contexte définitoire, et le terme <i>connective phrases</i> est utilisé, quant à lui, pour décrire d'autres patrons qui lient un terme à l'information qui s'y rattache, mais où le « connective » n'est ni un verbe ni une expression verbale . Ex. (<i>connective verbs</i>): <i>X comprise(s) Y, X is/are Y, X is/are known as Y</i> , etc.; (<i>connective phrases</i>) : <i>X i.e.Y, X e.g. Y, X is called Y, X is known as Y</i> , etc.
Pearson (1998)	<u><i>Metalanguage patterns ou statements</i></u>	Ces patrons sont présents dans certains types de textes spécialisés et donnent des indications sur le sens des termes auxquels ils réfèrent. Ils sont extraits des textes grâce à un ensemble de <i>connectives</i> (connectifs) ou de termes comme moyen de recherche. Ils indiquent des relations synonymiques, hyponymiques, méronymiques, etc. (Synonymes : voir la définition de <i>defining expositives</i> ci-dessous ou celle des <i>knowledge rich-contexts</i> de Davidson <i>et al.</i> (1998) et Meyer <i>et al.</i> (1999)).

Pearson (1998)	<u><i>Defining expositives</i></u>	Ce type de définition apparaît dans l'entourage d' expressions telles que : <i>X is/are defined as Y, X denote(s) Y, X consist(s) of Y, X comprise(s) Y</i> , etc. pour définir un terme peu familier des lecteurs. Après une définition, l'on voit aussi : <i>X is/are known as Y, X is/are called Y</i> , etc. Les segments de textes dans lesquels ces patrons apparaissent sont susceptibles d'être des <i>defining expositives</i> (Synonymes : voir la définition de <i>metalinguage patterns ou statements</i> ci-dessus ou celle des <i>knowledge rich-contexts</i> de Davidson <i>et al.</i> (1998) et Meyer <i>et al.</i> (1999)).
Pearson (1998)	<u><i>Linguistic signals</i></u>	Les <i>linguistic signals</i> (signaux linguistiques) sont des patrons comme ceux-ci : <i>X i.e.Y, X denotes Y, X is called Y, X is known as Y</i> , etc. Si l'un de ces patrons est accompagné d'une référence générique, cela confirme la présence d'un terme.
Sager <i>et al.</i> (1980)	<u><i>Connective verbs</i></u>	Sager <i>et al.</i> (1980 : 186) utilisent le terme <i>connective verbs</i> (verbes connecteurs) pour décrire les verbes qui lient un terme à sa description. Ex. : " <i>to be, to have or to give</i> ".
Séguéla (1999)	<u><i>Marqueurs lexicaux-syntaxiques</i></u>	Deux types : les marqueurs génériques et les marqueurs spécifiques . Un marqueur générique est un patron lexico-syntaxique qui désigne une relation de façon plus ou moins stable à travers différents corpus et domaines techniques. Un marqueur spécifique est, quant à lui, un patron lexico-syntaxique qui désigne une relation avec une certaine précision relativement à un corpus.

2.3.2 *Travaux terminologiques*

Nous présentons les travaux de terminologie traitant des PRS qui sont analogues au nôtre. Nous commençons par les premiers travaux publiés sur le sujet, soit ceux d'Ahmad et Fulford (1992) (2.3.3.1). Puis, nous avons regroupé les travaux qui suivent en fonction de leurs similarités : Rebeyrolle (2000) et Auger (1998) (2.3.3.2), qui ont travaillé sur les énoncés définitoires; Pearson (1998) (2.3.3.3), Davidson *et al.* (1998) et Meyer *et al.* (1999) (2.3.3.4), Marshman *et al.* (2002) (2.3.3.5), Marshman (2002) (2.3.3.6), dont les objectifs consistaient principalement à dresser des inventaires de patrons de relations sémantiques en vue d'aider le travail terminologique; Condamines *et al.* (1999) (2.3.3.7) et Séguéla (1999) (2.3.3.8), dont les travaux ont permis de mettre en lumière les caractéristiques distinctives des corpus. Enfin, nous terminons par les travaux les plus récents, soit ceux de Condamines (2003) (2.3.3.9), qui sont une réflexion sur la prise en compte de la nature des corpus dans l'étude des patrons de relations sémantiques.

Souvent, les auteurs que nous citons utilisent une terminologie qui diffère de celle que nous avons utilisée jusqu'ici. En effet, on trouve par exemple *marqueur de relation* (ex. : Condamines (2003)), alors que nous utilisons *patrons de relations sémantiques* (PRS); on trouve également *relation conceptuelle* (ex. : Davidson *et al.* (1998)), alors que nous utilisons *relation sémantique*. Dans ce qui suit, nous avons respecté la terminologie des auteurs dans la mesure du possible.

2.3.2.1 Ahmad et Fulford (1992)

Les travaux d'Ahmad et Fulford (1992) – les premiers du genre – ont porté sur les relations sémantiques (qu'ils appellent *sense relations* ou *lexical-semantic relations*) qui peuvent être extraites des textes de façon semi-automatique dans le but de construire des banques de termes à usages

multiples. Les auteurs étudient les termes d'un domaine liés les uns aux autres par différentes relations sémantiques qui sont présentées dans le Tableau 5.

D'abord, les auteurs dressent un tableau récapitulatif des relations sémantiques dont font état différents chercheurs : Lyons (1977), Palmer (1981), Jackson (1988) et Cruse (1986). Ils ont étudié de façon empirique les lexèmes interchangeables (*substitutable lexical items*) que l'on trouve dans la documentation, qu'ils ont appelé *knowledge probes*. Les chercheurs décrivent une technique d'extraction de relations sémantiques à partir de textes au moyen de *probes* pour chaque relation. La recherche de *probes* pour dégager des relations sémantiques à partir d'un texte met au jour les relations entre les termes, ce qui permet de construire la hiérarchie structurée d'un domaine. Le Tableau 5 présente les relations sémantiques étudiées par les auteurs.

Tableau 5. Relations sémantiques d'Ahmad et Fulford (1992)

Relation	Définition	Paraphrase
Synonymique	Relation qui exprime une égalité entre deux expressions.	<i>X is Y</i>
Hyponymique	Pour les auteurs, ce terme englobe tant les relations hyperonymiques que hyponymiques. Cette relation établit un ordre hiérarchique entre concepts de même catégorie.	<i>X is a kind of Y,</i> <i>X includes Y</i>
Partitive	Ce terme englobe l'ensemble des relations méronymiques à l'exception des relations compositionnelles.	<i>X is part of Y</i>
Causale	Relation de cause à effet	<i>X causes Y</i>
Compositionnelle ¹³	Relation exprimant le rapport entre un objet et ses éléments constitutifs ou ceux dont il peut être constitué. Cette relation équivaut à une relation méronymique de type constituant – objet.	<i>X is made of Y</i>

¹³ Traduction de *compositional relation*, terminologie des auteurs.

Les auteurs ont utilisé le programme d'analyse Kon Text (Mate) pour l'identification semi-automatique de termes au moyen des équivalents des patrons pour chacune des relations. La première partie de la recherche a consisté à identifier des « probes » en les recherchant manuellement dans une variété de textes provenant du *Surrey English automotive engineering corpus*. Ensuite, les auteurs ont consulté des dictionnaires de synonymes pour le repérage de synonymes des lexèmes utilisés dans la formule de chaque relation. Par exemple, dans la relation d'hyponymie (qui correspond en fait à une relation d'hyperonymie) exprimée par *X is a kind of Y*, ils ont identifié les synonymes de *kind*¹⁴. Par la suite, les *knowledge probes* de chaque relation ont été testés dans des textes du *English TWB automotive engineering corpus*.

Des raffinements ont été apportés aux *probes* dans la phase 2. Les auteurs ont utilisé des *wildcards patterns* (sortes de patrons à caractères de substitution) plus sophistiqués pour limiter ou étendre la recherche selon les cas. Dans la phase 3 du processus d'identification des *knowledge probes*, ils ont analysé brièvement un corpus encyclopédique, soit le thésaurus de Bernard (1990), où chacune des « probes » existantes y a été repérée. Quant aux synonymes et autres mots ou expressions, ils ont aussi été identifiés comme *probes* potentielles de relations sémantiques.

Dans la partie expérimentale, les auteurs apportent des exemples des relations sémantiques typiques. Ils y présentent sous forme de tableau un résumé des résultats des recherches dans l'échantillon de textes sur les *knowledge probes* pour chacune des relations. Ils y présentent aussi les données lexicales recueillies pour chacune des relations, soit causale, hyponymique, compositionnelle, partitive et synonymique (voir le Tableau 6). Les résultats de la recherche montrent que l'utilisation de *knowledge probes* semble concluante pour extraire des relations sémantiques à partir de textes.

¹⁴ Les recherches sur les PRS portent aussi sur leurs formes fléchies.

Tableau 6. Liste de patrons (Ahmad et Fulford 1992)¹⁵

Relations	Patrons
Causales	Affect*, alter*, causal, changing, induce*, cause*, bring* about, causation, gave rise to, inducing, effect*, causally, causing, give* rise to, influenc*, causative*, brought about, change %, giving rise to, initiate %
Hyponymiques	Action %, category, classes of, grade % of, make % of, are, class, collection %, group*, manner of, brand %, class of, event %, include %, member %, breed %, classed as, form %, is a, process, categories, classes, generic, kind % of, processes
Compositionnelles	Build, composition %, fabric, material, stuff, built, construct*, form*, material %, substance %, composed of, constructed from, made of, produce %, composition, create %, manufacture %, made out of
Partitives	Area %, category, constituent %, division %, encompass*, bit %, component %, contain*, domain %, factor %, branch*, compris*, department %, element %, feature %, categories, consist* of, district %, embrace*, includ*
Synonymiques	Also called, another way of saying, familiarly known as, identical to, known* as, also known as, better, frequently, identical with, mean, also referred to as, better known as, generally, in general, meaning, also termed, equivalent, i.e., is a, more commonly

¹⁵ Dans le 0, les symboles % et * sont utilisés pour indiquer les formes fléchies. L'astérisque (*) remplace n'importe quel nombre de caractères, alors que le symbole % remplace un seul caractère, comme dans l'exemple suivant : **g%v* rise to** correspond à : **gives rise to, giving rise to, given rise to et gave rise to.**

2.3.2.2 Rebeyrolle (2000), Auger (1998)

Rebeyrolle (2000) s'est intéressée aux formes de la définition dans les corpus spécialisés. La définition est représentée par une structure de base comportant : un terme à définir, le *definiendum*, noté *A*, une séquence descriptive du sens de *A*, le *definiens*, constituée d'un élément, *B*. En d'autres mots, la définition consiste en un rapport entre une idée (le *definiendum*) et d'autres idées (le *definiens*). Le tout est accompagné d'une périphrase permettant de faire le départ entre *A* et *B*, notée *X* (modifieur). Dans l'exemple 39, *carré* correspond à *A*; *figure* correspond à *B* et l'expression *à quatre côtés égaux et à angles droits* correspond à *X* :

39. Le carré est une figure à quatre côtés égaux et à angles droits (Rebeyrolle 2000 : 128).

Rebeyrolle (2000) a montré la variété des moyens linguistiques qui permettent d'établir l'équivalence entre *A* et *B-X*. L'auteure a étudié la manière d'établir l'équivalence comme lieu de variation entre les définitions, en se basant sur l'hypothèse de Riegel, hypothèse selon laquelle :

« L'acte propositionnel effectué par une définition consiste à référer aux deux termes de l'équation définitoire (le *definiendum* et le *definiens*) et à prédiquer le rapport qui les unit » (Riegel 1987 : 36).

Rebeyrolle (2000) précise les propriétés syntaxiques et sémantiques des structures définitoires. Elle distingue, sur la base de critères purement formels, les énoncés définitoires directs qui, pour reprendre Riegel (1987 : 37), « affichent explicitement le caractère métalinguistique de l'opération définitoire » et les énoncés définitoires indirects qui « ne présentent pas les deux marques typiques de l'opération définitoire : l'interprétation autonymique du *definiendum* et la verbalisation d'un rapport métalinguistique entre *definiendum* et *definiens* ».

D'abord, Rebeyrolle (2000) analyse les énoncés définitoires métalinguistiques aussi appelés *énoncés définitoires directs*. La typologie de Rey-Debove (1978) a servi de cadre à sa méthodologie. Cette typologie oppose deux types de verbes métalinguistiques en tenant compte de la nature de la relation qu'ils marquent :

- 1) une relation référentielle entre le signe et ce qu'il désigne : la relation s'établit entre le mot et la classe référentielle;
- 2) une relation sémantique entre un signe et son signifié : la relation reste intralinguistique.

Dans le premier cas (1), deux types de verbes sont possibles : *désigner* et *s'appeler*. Ces deux verbes partagent la même propriété d'asserter l'existence d'un lien référentiel stable entre le mot et l'objet qu'il désigne ou dénomme. Notons que le verbe *désigner* marque une relation d'explicitation ou de désignation (où le thème est le terme à définir), orientée du mot vers la classe référentielle. Le verbe *s'appeler* marque, au contraire, une relation de dénomination (où le thème est l'explication définitoire) orientée cette fois de la classe référentielle vers le mot. Dans le deuxième cas (2), c'est le verbe *signifier* qui marque la relation. Le Tableau 7 résume les observations faites dans ce paragraphe.

Tableau 7. Énoncés définitoires directs de Rebeyrolle (2000)

Orientation de la relation structure définitoire		Exemple
Mot → chose	[SNa V désigner SNx – X]	Librairie désigne un magasin où l'on vend des livres.
	[N ₀ V désigner SNa SNx – X]	On appelle « librairie » un magasin où l'on vend des livres.
Chose → mot	[SNx –X V s'appeler SNa]	Un magasin où l'on vend des livres s'appelle « librairie ».
Signe → signifié	[A V signifier B – X]	Le mot librairie signifie « magasin où l'on vend des livres ».

Légende : Na est un terme hyponyme; Nx est un terme hyperonyme; S correspond à *syntagme*; Sna correspond à : déterminant + terme hyponyme; SNx correspond à : déterminant + terme hyperonyme; N₀ est un sujet agentif (on, je, nous); X est un modifieur; il correspond à un ensemble de caractéristiques; A a pour signifié B; B est le signifié de A, et V correspond à *verbe*.

En plus des énoncés définitoires directs, Rebeyrolle (2000) a aussi étudié les énoncés définitoires indirects de forme [SNa est un Nx – X]. L'auteure définit cette structure lexico-sémantique en ces mots :

« La structure lexico-sémantique [Sna est un Nx – X] marque une relation d'inclusion entre les noms *Na* et *Nx*, où *Na* est l'hyponyme de *Nx*, l'hyperonyme. Cette phrase attributive est une structure couramment considérée (cf. p. 53) comme le schéma canonique des phrases dites définitoires qui s'appuient sur une relation d'inclusion entre le défini, *Na*, et un mot classifieur, *Nx*, accompagné d'un ensemble de caractéristiques, *X*, qui permettent de faire départ entre l'extension du mot défini et les extensions respectives des autres mots qui appartiennent à la même classe, ou co-hyponymes » (Rebeyrolle 2000 : 127).

L'auteure a relevé les propriétés syntaxiques et sémantiques des structures servant à introduire un énoncé définitoire indirect. Elle a délimité les variations intrinsèques de chacun des éléments qui composent la structure définitoire copulative de forme [SNa est un Nx – X]. Le Tableau 8 fait la synthèse de ces propriétés que l'auteure a rassemblées.

Tableau 8. Énoncés définitoires indirects de Rebeyrolle (2000 : 164)

Propriétés grammaticales des éléments				Propriétés Discursives
Na	Copule	Nx	X	
Article défini (ex. : le, la, les) Article indéfini (ex. : un, une, des) Tout Chaque	Présent (ex. : est, sont) Passé (ex. : était) Futur (ex. : sera)	Article indéfini (ex. : un, une, des) Article défini (ex. : le, la, les)	Sprep (ex. : ... de comblement) Sadj (ex. : ...sous- marines) Pprésent (ex. : ... fournissant une lecture normalisée...) Ppassé (ex. : ... creusées par un glacier de vallée) Relative (... qui contiennent une certaine proportion de calcaire) Prop. indép. (Elles contiennent une certaine proportion de calcaire.)	Reprise fidèle de Na (ex. : ... situés sous un <u>titre</u> . <u>Le</u> <u>titre est un</u> <u>bloc...</u>)

Le contexte 40, tiré de Rebeyrolle (2000 : 141), illustre un exemple :

40. Le flysch est un faciès de comblement.

Dans l'exemple 40, *Le* est un article défini, *flysch* est un Na (ou hyponyme), *Le flysch* est un SNa, *est* est une copule (au présent), *un* est un article indéfini, *faciès* est un Nx (hyperonyme), *de* est une préposition et *comblement* est un X.

Auger (1998) s'est, lui aussi, intéressé aux énoncés définitoires. Il a classé les marqueurs d'énoncés d'intérêt définitoire en huit catégories distinctes :

- 1) analyse (ex : *constituer de, partie de*);
- 2) caractérisation (*genre de, sorte de*);
- 3) causalité (*être à l'origine de, provenir de*);
- 4) dénomination (*appeler, on appeler* (sic));
- 5) désignation (*le mot, vouloir dire*);
- 6) équivalence (*c'est-à-dire, il s'agit*);
- 7) fonction (*destiné à, permettre de*);
- 8) systémique (*la locution, le substantif*).

À la base de sa recherche, Auger (1998) a choisi les formes définitoires les plus étudiées, soit *désigner, signifier, dénommer* et *s'appeler*. Il est d'avis qu'il faudrait nuancer l'affirmation de Loffler-Laurian (1983) et de Condamines (1993) selon laquelle ces formes définitoires sont beaucoup plus fréquentes dans les discours scientifiques et techniques. Ces discours ont en effet une vocation didactique et ils tendent à définir les termes qu'il emploient, mais il s'agit plutôt d'une distribution complémentaire des formes définitoires entre discours littéraire et discours scientifiques. Un exemple type est certainement la comparaison entre *signifier* et *vouloir dire*. Selon les résultats d'Auger (1998), *signifier* est spécifique aux discours scientifiques. Par contre, *vouloir dire* est davantage typique du discours littéraire et de la langue générale. En conclusion, cette

recherche a permis d'enrichir la liste du vocabulaire définitoire et de découvrir que certaines formes sont plus fréquentes dans la langue de spécialité, alors que d'autres sont plus fréquentes dans la langue générale.

2.3.2.3 Pearson (1998)

Les travaux de Pearson (1998) portent sur l'identification de patrons annonçant des éléments qui pourraient être utilisés dans les définitions terminologiques, que l'auteure appelle *metalanguage patterns* (patrons métalinguistiques). L'auteure utilise trois corpus spécialisés en format électronique qui totalisent près de 6 millions de mots. Les corpus se prêtent bien à l'extraction de ces patrons métalinguistiques. Ils sont composés de textes appartenant aux trois niveaux de spécialisation suivants : 1) communication expert – experts; 2) communication expert – initiés et 3) communication semi-expert – non-initiés (voir la section 2.1.3.1 pour la description de ces niveaux de spécialisation).

Pearson (1998) a procédé à l'identification et à l'extraction de termes contenus dans ses corpus. L'auteure voulait ainsi identifier un certain nombre de termes accompagnés de contextes définitoires. Les contextes recherchés par Pearson contenaient une forme de définition, un hyperonyme ou un synonyme. Il est nécessaire de savoir quelles formes prennent les termes afin de pouvoir les mettre en relation avec des patrons qui seront identifiés pour extraire par la suite des énoncés définitoires.

Concrètement, le protocole de repérage mis au point par Pearson se déroule en deux étapes distinctes : d'abord, il identifie des séquences susceptibles d'être des termes; ensuite, il localise des contextes contenant l'expression d'une relation sémantique.

D'abord, Pearson a fait une analyse manuelle de chacun des corpus. À la suite de cette analyse, l'auteure a été en mesure de spécifier certains critères qui permettent la reconnaissance des termes-candidats de façon semi-automatique.

Plus tard, elle a raffiné son analyse en ajoutant certaines restrictions pour réduire le bruit.

Par la suite, l'auteure a utilisé un programme qui permet d'extraire les termes-candidats correspondant à un patron de formation. Par exemple, un patron valable pour l'anglais est composé de la séquence d'étiquettes suivante : adj. + nom + nom. Le programme construit une représentation de chacune des séquences de patrons. Les patrons sont ensuite appliqués aux corpus étiquetés. Le programme donne le nombre de patrons, la séquence étiquetée et les mots ou expressions qui ont été repérés. Ces mots ou expressions sont appelés *termes-candidats*.

Les opérations suivantes portent plus spécifiquement sur la recherche de patrons. L'auteure a isolé les termes-candidats susceptibles d'être jumelés à l'une des formes de patrons métalinguistiques. Elle a recherché les *defining expositives*¹⁶, qui peuvent non seulement être étudiés à l'intérieur d'une seule phrase, mais également à l'intérieur de quelques phrases (*simple defining expositives vs. complex defining expositives*).

Au terme de sa recherche, l'auteure propose une étude des classes de termes et cherche à savoir si certaines ne seraient pas décrites en fonction de leur finalité ou de leur composition :

« Further investigation will be required to establish whether it is possible to predict what type of distinguishing characteristic will be specified, e.g. whether a particular class of term is more likely to be explained in term of its purpose or its properties, and how one might distinguish between different characteristics » (Pearson 1998 : 209).

Cette proposition de recherche se rapproche de nos propres objectifs. En effet, il n'y a qu'un pas à franchir entre les renseignements fournis sur différentes

¹⁶ Rappelons que les *defining expositives* sont des segments de textes dans lesquels certains patrons apparaissent. Les patrons sont de type : *X is/are defined as Y*, *X denote(s) Y*, *X consist(s) of Y*, *X comprise(s) Y*, etc. pour définir un terme peu familier des lecteurs. Après une définition, on voit aussi : *X is/are known as Y*, *X is/are called Y*, etc.

classes conceptuelles que Pearson propose et notre étude sur les patrons qui leur sont associés. Toutefois, l'auteure ne pousse pas l'investigation plus loin.

2.3.2.4 Davidson *et al.* (1998) et Meyer *et al.* (1999)

Davidson *et al.* (1998) et Meyer *et al.* (1999) s'intéressent particulièrement au réseau de concepts qui sous-tend les termes d'un domaine. Pour les auteurs, un contexte est dit *riche en connaissances* lorsqu'il présente une relation conceptuelle. Cette relation conceptuelle est exprimée par des *patrons de connaissances* qui lui sont propres.

Les auteurs s'intéressent aux patrons de connaissances dans la mesure où ils leur permettent de trouver, de façon automatique ou semi-automatique, des renseignements sur le sens des termes. Leur objectif principal consiste donc à dresser des listes de patrons et à mesurer leur efficacité.

Les auteurs distinguent trois types de patrons qui peuvent chacun indiquer différentes relations :

- a) Les *patrons lexicaux* sont représentés par des mots ou des groupes de mots (et leurs variantes morphologiques). Ils peuvent exprimer un grand nombre de relations. Des exemples de patrons lexicaux qui expriment une relation hyperonymique comprennent **is a, classified as, and other**; des exemples de patrons lexicaux qui expriment une relation méronymique comprennent **its, is a part of, contains**; enfin, des exemples de patrons lexicaux qui expriment une relation de finalité comprennent **needed for, in order to, functions of**. Tout comme Marshman *et al.* (2002), notre recherche repose sur les patrons lexicaux qui semblent génériques, c'est-à-dire qui sont susceptibles d'être présents dans des textes de spécialité portant sur divers domaines :

« This article focuses on *lexical* knowledge patterns commonly used in French, and more specifically on lexical patterns which appear to be

generic, that is, potentially relevant to many domains » (Marshman *et al.* 2002 : 3).

- b) Les *patrons grammaticaux* sont représentés par des mots sémantiquement pleins. Ils sont représentés par des combinaisons de parties du discours. Ils se répartissent en effet en deux groupes, soit les patrons **TERMES + VERBES** et les patrons **ADJECTIF + NOM** (pour l'anglais, alors que pour le français la séquence correspondante est **NOM + ADJECTIF**). Les patrons **TERMES + VERBES** sont des patrons tels que *X assure Y*, *X effectue Y*, *X transforme Y*, etc. (Meyer, non publié). Lorsqu'un nombre très élevé de verbes peuvent prendre place après X, il est plus utile de travailler avec des patrons grammaticaux. Il suffit alors de préciser, lors de la requête, que X doit être suivi d'un verbe. Une liste d'exceptions doit comprendre, entre autres, les verbes d'état (ex. : être), afin d'éviter le bruit. Quant aux patrons **ADJECTIFS + NOMS**, l'auteure donne l'exemple suivant : *a writable disk* (Meyer, non publié). Ce patron grammatical prend forcément la forme d'un adjectif dérivé de verbe. Les patrons grammaticaux expriment un moins grand nombre de relations que les patrons lexicaux. Ils expriment généralement une relation de finalité.
- c) Enfin, les *patrons paralinguistiques* expriment des relations au moyen d'items paralinguistiques, c'est-à-dire non lexicaux ou non grammaticaux, comme les parenthèses (...) ou les deux points (:).

C'est au moyen de l'extraction automatique ou semi-automatique à partir de corpus que les auteurs ont testé les patrons. Leur méthodologie consiste d'abord à identifier un certain nombre de patrons de connaissances pour chacune des relations conceptuelles étudiées (hyponymie, méronymie, fonction¹⁷). La deuxième étape consiste à analyser les données obtenues afin de repérer les silences (contextes qui n'ont pas été repérés par le logiciel lors de la requête, mais qui auraient dû l'être). Si nécessaire, d'autres patrons ou des raffinements à ceux

¹⁷ Cette relation correspond à ce que nous appelons *finalité*.

qui ont déjà été identifiés sont ajoutés. Ensuite, les auteurs apportent des modifications dans un outil appelé *Text Analyzer*¹⁸. Enfin, il faut répéter les deuxième et troisième étapes aussi souvent que nécessaire.

Les meilleurs résultats ont été enregistrés pour la relation d'hyponymie; c'est-à-dire que peu de bruit et peu de silence ont été observés. Une des difficultés soulevées par cette approche consiste à trouver le juste équilibre entre les avantages et les inconvénients que provoque l'addition de patrons de connaissances. En effet, un ajout peut engendrer beaucoup de bruit (contextes qui ne répondent pas aux critères de la requête et qui n'auraient donc pas dû être repérés par le logiciel).

Les difficultés liées à la recherche de patrons de connaissances sont de deux ordres : 1) leur complexité; et 2) leurs limites intrinsèques.

- 1) Complexité des patrons
 - a) Patrons imprévisibles : La liste des patrons lexicaux présents dans le corpus est difficile à dresser de manière exhaustive. Alors que certains patrons sont prévisibles, comme par exemple *est un* pour l'hyponymie ou *contient* pour la méronymie, d'autres sont imprévisibles. Par exemple, hors contexte, on n'associerait pas forcément le mot *recolonize* à de la méronymie. Cependant, dans la phrase « The numbers and types of mesophilic microbes that recolonize compost as it matures... », le patron **X recolonizes Y** exprime une relation méronymique. Certains patrons sont donc spécifiques, c'est-à-dire qu'ils sont rattachés à un corpus ou à un domaine.

¹⁸ Outil d'analyse générale de corpus qui sert entre autres de concordancier; il suffit à l'utilisateur de faire une requête sur une relation donnée pour que le *Text Analyzer* recherche dans le corpus les patrons qui se trouvent, par exemple, à une certaine distance à gauche ou à droite du mot clé; cet outil permet, de plus, d'ajouter des restrictions grammaticales aux patrons lors d'une requête.

- b) Patrons polysémiques : Un patron donné peut exprimer plus qu'une relation conceptuelle. Par exemple, le patron lexical **X is defined as Y**, dans « Compost **is** perhaps best **defined as** a pile of organic materials deliberately assembled for fast decomposition », exprime une relation hyperonymique. Cependant, en informatique, *define* veut souvent dire « créer » (**define the columns**). Plus un patron est polysémique, plus il risque d'exprimer des relations autres que celles auxquelles on s'attend. La préposition **in** est un bon exemple de ce problème. En effet, bien qu'elle serve souvent à exprimer une relation méronymique (« *If your compost mix is too low **in** nitrogen...* »), elle génère également beaucoup de bruit (« *as temperatures rise **in** the compost...* »).
- c) Caractéristiques inhérentes aux relations conceptuelles : Certaines relations conceptuelles sont plus complexes que d'autres. La méronymie, par exemple, comprend une liste étendue de relations différentes (voir la section 2.2.2). Ces types de méronymie sont associés à des patrons de connaissances différents. Par exemple, lorsque Davidson *et al.* (1998) et Meyer *et al.* (1999) ont utilisé des patrons méronymiques pour le terme *composting*, le pourcentage de silence a augmenté de façon vertigineuse, car il leur manquait des patrons qui se sont avérés essentiels pour ce concept. Puisque *composting* désigne un processus, et non un objet concret, ce terme a des parties « temporelles » plutôt que « physiques ». Ces parties temporelles sont désignées par des patrons tels que *X is a stage of Y*, *X is a phase of Y*, patrons qui n'étaient pas pertinents du tout pour le terme *compost*. Rappelons que, dans notre typologie, cette relation correspond à une relation méronymique de type phase – activité.
- d) Variabilité et spécificité des patrons : La forme de beaucoup de patrons varie grandement. Par exemple, la patron hyperonymique **X is a Y** a des variantes telles que **X is the Y**, **Xs are Ys**, etc. Enfin, parfois, pour réduire le bruit, il peut être nécessaire d'imposer des restrictions quant à l'utilisation d'un patron : il faudra qu'on le restreigne aux cas où il est

précédé ou suivi de mots bien choisis. Par exemple, on trouve la préposition **in** dans un grand nombre de contextes méronymiques (ex. : « *If your compost mix is too low **in** nitrogen...* »). Cependant, si elle est utilisée seule comme patron, elle produit beaucoup de bruit (« *Compost improves vitamin and mineral content in food...* »). Cette préposition devrait être restreinte aux cas où elle est précédée de *rich, high, low, etc.*

2) Limites intrinsèques

Les relations conceptuelles s'expriment par d'autres moyens que les patrons linguistiques. Nous avons reproduit trois exemples ci-dessous.

- a) **Éléments typographiques** : Les relations conceptuelles sont souvent introduites par des éléments typographiques ou des indices. Les auteurs ont remarqué que les éléments typographiques, les crochets, les parenthèses, les tirets et les virgules expriment souvent des relations hyperonymiques. Dans d'autres cas, d'excellentes définitions et contextes méronymiques sont précédés de questions qui constituent de bons indices de relations conceptuelles, comme « Qu'est-ce que le compost? » ou « De quoi se compose le compost? ».
- b) **Style** : Les différents styles et l'emploi de métaphores compliquent le repérage de patrons pour un outil informatique. Même dans les textes techniques et didactiques, les contextes métaphoriques sont très répandus.
- c) **Fenêtre de recherche** : La recherche de ces auteurs tient compte de la distance (à droite ou à gauche) entre le patron et le mot clé. Cette approche, appelée *search window* (fenêtre de recherche), doit être spécifiée au logiciel d'extraction de connaissances. Mais, dans les textes écrits, l'emploi de pronoms (liens anaphoriques) est fréquent pour éviter les répétitions; ces contextes sont donc malheureusement perdus.

Pour conclure, Meyer *et al.* ont classé les structures linguistiques identifiées en trois catégories : celles qui expriment la relation générique –

spécifique, celles qui expriment la relation partie – tout, et celles qui expriment la fonction. Nous reproduisons dans le Tableau 9 les patrons utilisés en français.

Tableau 9. Résultats de Meyer *et al.* (2001 : page Internet)¹⁹

Relations	Structures linguistiques
Relation générique – spécifique	Est un/sont des, type* d*, est l*/sont les, comme + terme, tel* qu*, classe* (d*), sorte* d*, appelé* + terme. Il s'agit d*, catégorie*, nommé* + terme, forme* d*, y compris.
Relation partie – tout	Contien*/conten*, compos* d*, constitu* d*, partie (de), présent en/dans, compren*, comport*, possèd*, dans + terme/terme + dans, composant*, dont, dispos* d*, équipé* d*, formé* d*/par.
Relation fonctionnelle	Verbes, permet* d*, permet* + syntagme nominal, (s')utilis* pour/comme, capable* d*, pour + verbe, ser* à/d*, destiné* à, par + terme, nécessaire à/pour, fonction*, rôle*, grâce à + terme, au moyen d*, par l'intermédiaire d*, consist* à.

2.3.2.5

Marshman *et al.* (2002)

Étant donné que peu de listes de patrons de connaissances sur le français sont disponibles, les auteures, membres du groupe de recherche de Meyer *et al.* (2001), ont rédigé cet article à partir de l'étude qui a servi à dresser la liste de structures linguistiques présentée dans le Tableau 9. L'article fournit un inventaire des patrons de connaissances lexicaux généralement utilisés en français, et plus spécifiquement des patrons lexicaux *génériques*. On entend par patron *générique*, un patron qui est susceptible d'apparaître dans différents corpus spécialisés. Les

¹⁹ L'astérisque peut remplacer un ou plusieurs caractères, s'il y a lieu. Quant à la barre oblique, elle signifie *ou*.

patrons étudiés expriment trois relations : l'hyperonymie, la méronymie et la fonction.

Pour cette recherche, deux corpus appartenant à deux langues de spécialité ont été utilisés : un corpus d'informatique (concepts de base de l'informatique) et un corpus de génétique (incluant la biotechnologie). Quatre raisons ont motivé le choix de ces langues de spécialité : 1) les relations conceptuelles étudiées sont bien représentées; 2) elles sont également bien représentées dans TERMIUM® et dans d'autres grandes banques de terminologie; 3) des textes « naturels »²⁰ des deux domaines sont facilement accessibles dans Internet; et 4) puisque les deux domaines sont différents, les auteures ont pensé qu'il serait possible de faire des comparaisons intéressantes entre les patrons, et que les patrons *génériques* étaient en effet partagés par différents domaines.

C'est le terme qui a été utilisé comme point de départ pour découvrir des patrons de connaissances. Sur les fiches de TERMIUM®, les champs du contexte et de la définition ont été dépouillés manuellement pour y retrouver des patrons. Quant aux textes du corpus (textes puisés dans Internet), les auteures ont utilisé la méthode KWIC (key-word-in-context concordance) pour chacun des termes, et chacune de ces concordances a été dépouillée manuellement.

Les auteures ont fait des observations sur les problèmes différents posés par la recherche de patrons dans des corpus de langue française et l'extraction de patrons dans les corpus de langue anglaise qui ont fait l'objet de leurs recherches précédentes.

Les principales différences observées entre les patrons anglais et français sont d'abord la très grande productivité de certains patrons français tels que

²⁰ « Les données linguistiques en question doivent apparaître dans un environnement « naturel » (des mots sont combinés à d'autres, sont utilisés dans des phrases, les phrases s'agencent dans un texte, etc.); sur ce plan, le corpus se distingue d'ouvrages de référence comme les dictionnaires qui sont le résultat d'analyses faites par des spécialistes et qui reflètent certains choix faits par eux (L'Homme 2004 : 118) ».

*constitu** et *désign**, contrairement à leurs équivalents anglais *constitute** et *designate** qui sont beaucoup moins utilisés en anglais. Les auteures pensent que le français serait réticent à employer des verbes tels que *avoir* et *être*, alors qu'en anglais, on n'hésite pas à employer des verbes comme *to be* et *to have*.

De plus, certains patrons anglais produisent beaucoup plus de bruit que leurs équivalents français. Par exemple, bien que *chez* puisse être utilisé comme patron en français, son équivalent anglais *in*, par contre, produirait beaucoup trop de bruit.

En français, les patrons sont également sujets à une plus grande variation. En effet, l'élision, l'accord en genre et en nombre, ainsi que la plus grande variété des formes verbales sont responsables de cette variation. Il est donc important de considérer toutes les variations des patrons, incluant le genre, le nombre et la variation morphologique, afin de ne pas perdre des contextes intéressants. Par exemple, *type of* varie peu en anglais, mais son équivalent français, *type* d**, se présente sous plusieurs formes : *type de*, *type d'*, *types de*, *types d'*, *de type* et *de types*.

Les différences générales entre l'anglais et le français concernent aussi l'ordre des mots, qui est plus souple en français. De plus, les patrons sont généralement plus longs en français qu'en anglais. Ceux-ci sont donc plus sujets aux variations et présentent une plus grande complexité. Les exemples 41 et 42 montrent cette différence entre l'anglais et le français.

41. Adaptor: a *device* for connecting two units which would not otherwise be directly connected... (Marshman *et al.* 2002 : 14).
42. Adaptateur : *Dispositif de connexion* conçu pour servir d'intermédiaire entre deux unités...(Marshman *et al.* 2002 : 15).

Ces différences rendent nécessaires l'utilisation de fenêtres de recherche plus grandes pour l'extraction semi-automatisée à partir des textes français, ce qui augmente ainsi le bruit.

Enfin, une autre difficulté en français concerne l'utilisation d'un pronom mettant en relief un élément de la phrase. Dans l'exemple 43, l'utilisation du pronom *c'* pourrait empêcher le repérage du contexte par un outil de repérage semi-automatique, à moins d'avoir au préalable programmé l'outil afin que celui-ci tienne compte des variations éventuelles.

43. Le premier périphérique d'un PC, *c'*est son moniteur ... (Marshman *et al.* 2002 :14).

2.3.2.6 Marshman (2002)

L'étude de Marshman (2002) a porté sur l'extraction semi-automatique d'informations conceptuelles au moyen de marqueurs des relations causales. L'auteure avait pour objectif d'identifier des marqueurs de la relation causale dans des textes médicaux anglais et français. Elle s'est inspirée principalement des travaux de Talmy (1985, 1988, 2000) et de ceux de Barrière (2001, 2002). Marshman tenait pour acquis que cette relation est caractéristique du domaine de spécialité étudié, soit la pharmaceutique.

L'étude de Marshman comportait les étapes suivantes :

- 1) L'identification de marqueurs linguistiques de la relation causale en français et en anglais dans le domaine des produits biopharmaceutiques.
- 2) L'étude de l'efficacité des patrons dans l'extraction de connaissances à partir de corpus, et la mesure de la précision de ces derniers afin d'identifier ceux qui sont les plus prometteurs.
- 3) La comparaison interlinguistique des similitudes et des différences entre les patrons anglais et français.

Un grand nombre de patrons indiquant une relation causale ont été étudiés, tant en anglais qu'en français. Quatre-vingt-dix pour cent d'entre eux ont montré une bonne précision (plus de 50 %). Ils deviennent, par le fait même, de très bons

candidats pour la recherche dans le domaine de l'analyse de textes semi-automatique.

Tant pour l'anglais que pour le français, un certain nombre de patrons recueillis au cours de la recherche étaient liés à la langue de spécialité pharmaceutique, alors que d'autres étaient liés au domaine médical en général.

Pour l'auteure, il a été surprenant de constater que les mots grammaticaux polysémiques (ex. : *after, by, après, par*) sont très fréquents dans les contextes exprimant une relation causale, et ce tant en anglais qu'en français. Les patrons formés de ces mots grammaticaux polysémiques posent problème dans l'extraction semi-automatique des connaissances puisqu'ils sont source de bruit. Pour cette raison, aucun de ces patrons n'a obtenu une précision supérieure à 50 %.

Les résultats de Marshman (2002) pourraient maintenant être utilisés sur d'autres corpus afin de raffiner les outils, et pourraient permettre à des programmeurs de mettre au point des systèmes d'extraction de connaissances plus perfectionnés.

2.3.2.7 Condamines *et al.* (1999)

Les travaux de Condamines *et al.* (1999) visaient à constituer un référentiel terminologique dans un contexte défini afin d'identifier les similitudes et les convergences entre quatre différents corpus, tant sur le plan de leur organisation conceptuelle que sur le plan linguistique.

La méthode utilisée a été élaborée par l'Équipe de recherche en syntaxe et sémantique (ERSS) pour la construction de Bases de Connaissances Terminologiques (BCT). Les chercheurs n'ont pas construit une BCT pour chacun des corpus. Ils ont plutôt privilégié certains thèmes qui paraissaient essentiels pour les différents domaines étudiés. Quatre classes ont fait l'objet d'une analyse, compte tenu du rôle particulièrement important qu'elles semblaient jouer. Ces

classes sont les suivantes : a) humains; b) lieux; c) communication; et d) surveillance.

Les auteurs ont retenu et défini sept relations principales : X est un Y; X est partie de Y; X émet Y (X est un humain ou un système, Y est un type de message); X envoie à Y (X est un humain ou non, Y est un récepteur humain ou non); X médiatise Y (X est un système, un objet, ou un moyen de communication, Y est un message); X reçoit Y (X est un humain ou non, Y est un message); X surveille Y (X ou Y sont un humain ou un objet). De même, sept autres relations, propres à la classe des humains, ont été identifiées.

En outre, les auteurs se sont intéressés aux similitudes et différences entre les corpus, en examinant chacun des quatre thèmes dans les corpus ainsi que toutes les relations. Pour systématiser la recherche, ils ont eu recours aux marqueurs linguistiques, comme l'illustre l'exemple suivant :

*« Chaque subdivision transmet au CIGT une fiche relative aux chantiers qui se terminent. Dans cet exemple, où le verbe est un verbe de communication, la relation « envoie à » se met en place entre le sujet (subdivision) et le complément d'objet indirect (CIGT), la relation « émet » se met en place entre le sujet et le complément d'objet direct (fiche relative aux chantiers qui se terminent) enfin, la relation « reçoit » se met en place entre le complément d'objet indirect (CIGT) et le complément d'objet direct (fiche relative aux chantiers qui se terminent) » (Condamines *et al.* 1999 : 6).*

Le logiciel d'analyse de textes SATO a été mis à profit pour assister les chercheurs dans cette exploration. Ils ont également eu recours à l'outil GEDITERM²¹ pour saisir l'ensemble des données terminologiques. Cet outil permet non seulement de disposer des résultats dans une base de données, mais aussi de dessiner des graphes de relations à partir d'un concept ou de dessiner tout le graphe d'une relation.

²¹ Outil de Gestion et d'Édition de Termes, construit par l'équipe de Nathalie Aussenac de l'IRIT (Institut de Recherche en Informatique de Toulouse).

Pour conclure, les auteurs ont fait un important travail de dépouillement des corpus. En effet, ils ont traité plus de huit cent cinquante termes et plus de huit cent cinquante concepts, et quatorze relations ont été retenues. Les termes traités ont été intégrés à la Base de Connaissances Terminologiques. Enfin, les auteurs ont dessiné (à l'aide de l'outil GEDITERM) des graphes de relations à partir des quatre classes principales. Ces relations ont été étiquetées séparément par corpus.

Au terme de ce travail, les auteurs ont proposé deux pistes de recherche à explorer, à savoir : 1) constituer un référentiel terminologique et un dictionnaire avec définitions en intégrant d'autres relations afin d'obtenir une modélisation complète des corpus; et, 2) travailler directement à la rédaction de définitions en s'appuyant sur les relations existantes. Cette seconde piste rejoint en partie l'un des objectifs de notre propre recherche.

2.3.2.8 Séguéla (1999)

L'auteur propose une méthode d'extraction de relations sémantiques à partir de marqueurs lexico-syntaxiques (ce que nous appelons *PRS*), qui vise à distinguer les connaissances linguistiques générales des connaissances linguistiques spécialisées. Sa méthode d'extraction de relations entre concepts s'appuie sur des techniques de TAL (traitement automatique de la langue) d'application de marqueurs linguistiques sur des corpus. L'auteur définit le marqueur comme un patron lexico-syntaxique qui désigne dans le discours une relation sémantique entre des termes, et ce, avec une certaine précision pour un corpus donné.

Deux types de marqueurs sont pris en compte par le système de gestion : les marqueurs génériques et les marqueurs spécifiques. « Un marqueur générique (ex. : « Y ÊTRE ARTICLE_INDÉFINI X » ou « X ADVERBE_DE_SPÉCIFICATION Y », etc.) est un patron lexico-syntaxique qui désigne une relation de façon plus ou moins stable à travers différents corpus et domaines techniques » (Séguéla 1999 : 55). Alors qu'un marqueur spécifique

(ex. : « calculs? DE Y (avec|à l'aide de|par X » ou le marqueur typographique « X/Y » pour la relation structurante liant une structure (CEA) à une sous-structure (DAM) dans l'organigramme de l'entreprise dans CEA/DAM) « est un patron lexico-syntaxique qui désigne une relation avec une certaine précision relativement à un corpus » (Séguéla 1999 : 55).

Dans le système conçu par Séguéla, appelé *Caméléon*, les marqueurs dits *génériques* sont structurés dans une première base de marqueurs; leur précision est évaluée pour chaque nouveau corpus (1). Puis, suit une phase d'acquisition de marqueurs spécifiques pour les relations de la base générique et d'autres relations jugées intéressantes pour un objectif applicatif donné (2). Une fois les marqueurs validés (3), ils sont projetés sur le corpus préalablement indexé par l'ensemble des termes du domaine (4) afin de proposer des relations sémantiques pour l'enrichissement des modèles du domaine.

La base de marqueurs génériques comprend deux types de relations sémantiques : les relations d'hyponymie et de méronymie. Ces relations sont présentes dans un grand nombre de domaines. De plus, elles sont structurelles puisqu'elles organisent les connaissances en hiérarchies. La relation de méronymie étant plus complexe que celle d'hyponymie, il a été nécessaire de la décomposer en un ensemble de six sous-catégories d'après la typologie de Winston *et al.* (1987), que nous avons présentée à la section 2.2.2.

D'abord, on projette sur le corpus la base de marqueurs génériques. Quatre étapes sont nécessaires à la validation :

« (1) Un sous-ensemble significatif d'occurrences du marqueur est présenté à l'utilisateur pour sa validation [...]; (2) Suite à la validation du sous-ensemble d'occurrences, un taux de précision est associé au marqueur; (3) Si le taux de précision est jugé valable, le marqueur est validé et associé à la base de marqueurs du corpus; (4) Si le taux de précision du marqueur est trop faible, il peut être restreint, c'est-à-dire rendu plus spécifique » (Séguéla 1999 : 56).

Si le marqueur a été rendu plus spécifique, il est ramené en (1).

La méthode a été évaluée sur un corpus de deux millions de mots contenant 12 000 textes d'une page environ qui sont des entrevues d'experts ou encore des extraits de documents de référence sur le nucléaire. Ces textes sont récents, spécialisés, visent à informer et sont hétérogènes.

Dix occurrences de marqueurs ont été choisies pour la projection de la base générique dans l'outil *Caméléon*. Pour les relations d'hyponymie, tous les marqueurs ont désigné cette relation avec une bonne précision (+ 75 %). Ces marqueurs, qui sont composés surtout de formes adverbiales, de structures d'énumération et de construction avec le verbe *être* sont assez fiables sur un nouveau corpus. Quant aux marqueurs de méronymie de la base générique, ils sont construits sur des patrons qui portent essentiellement sur des syntagmes nominaux et des formes verbales ainsi que leurs formes dérivées. La validation de ces marqueurs a montré que, dans ce domaine technique, les marqueurs tels que *groupe*, *assemblage et mélange*, correspondent à des termes du domaine tels que *groupe turbo-alternateur*, *assemblage combustible et mélange turbulent*. Les marqueurs contenant ces formes ont donc dû être restreints afin de ne pas générer trop de bruit. Une majorité de marqueurs de méronymie ont été validés, mais avec une précision plus faible que celle obtenue pour l'hyponymie, étant donné que 50 % à 75 % d'entre eux ont été validés.

Pour chacun des couples de concepts en relation dans les modèles, *Caméléon* repère les contextes de cooccurrence de termes dans lesquels ils se trouvent. L'étude de ces contextes a permis d'extraire de nouveaux candidats marqueurs. Pour la relation d'hyponymie, la projection de 60 couples de concepts a permis de recueillir 180 contextes. L'étude de ces derniers a permis de retrouver les marqueurs les plus fréquents définis dans la base générique. Trois marqueurs ont finalement été validés. Quant à la relation d'hyponymie, l'auteur a analysé 300 contextes à partir de la projection de 180 couples de termes désignant des concepts reliés par cette relation. Trente-sept marqueurs, productifs et précis, ont été ajoutés à partir de 12 formes verbales et leurs transformations. Condamines (2002, 2003).

Condamines (2002, 2003) fait une synthèse de ses travaux sur la recherche d'informations sémantiques dans les corpus. Elle fait un certain nombre d'observations pertinentes pour notre propre recherche, observations que nous résumons dans cette section.

a) Discursivité des marqueurs

Les marqueurs sont des signes linguistiques, mais on leur attribue un fonctionnement linguistique en fonction d'un objectif particulier. Une structure linguistique est appelée *marqueur* après avoir constaté qu'elle est souvent associée à des éléments que l'on peut représenter sous une même forme relationnelle. L'auteure a remarqué que les marqueurs sont généralement moins spécialisés que les termes qu'ils réunissent. Particulièrement en Traitement automatique des langues (TAL), on considère les marqueurs comme des indices dans les cas où la dépendance avec le corpus est faible. Cependant, ces structures ne peuvent être complètement désémantisées et peuvent donc subir des variations contextuelles.

b) Les marqueurs ne donnent pas l'interprétation de la relation

Les marqueurs permettent au linguiste d'interpréter une portion de corpus sous la forme d'une relation binaire ou n-aire. C'est seulement lorsque leur utilisation est devenue très fréquente qu'on peut associer directement une forme à un contenu. En outre, cette association n'est possible que grâce à une interprétation humaine.

c) Variation des marqueurs en fonction du genre textuel

Le genre du corpus pourrait avoir une influence sur la relation qu'exprime un marqueur, comme le mentionne Condamines :

« Il est donc possible d'envisager une linguistique de corpus qui s'attache non seulement à décrire les structures au plus juste de leur pertinence pour permettre de repérer une relation mais aussi d'associer cette possibilité à des genres de corpus » (Condamines 2003 : 150).

Les études sur les rapports entre marqueurs et corpus ont amené l'auteure à considérer trois types de fonctionnement : 1) dans les cas extrêmes, il n'y a presque pas de dépendance; ou inversement, 2) cette dépendance est très forte; 3) dans les cas intermédiaires, la dépendance n'est pas liée à un corpus en particulier, mais plutôt au genre du corpus.

La dépendance entre marqueur et relation n'existe qu'avec des marqueurs très généraux. En effet, Séguéla (2001) a montré que des patrons connus d'hyponymie et de méronymie ne sont pas répartis de façon égale dans tous les corpus. Rebeyrolle et Tanguy (2000) arrivent aux mêmes conclusions avec les marqueurs définitoires. Enfin, selon Condamines (2003), les cas d'indépendance totale entre corpus et marqueur sont rares.

Les marqueurs qui ne dépendent pas d'un corpus en particulier ainsi que les relations qu'ils expriment, sont moins spécialisés que les termes qu'ils relient. La plupart des relations sont très générales et ne dépendent donc pas du domaine. Ce sont donc surtout les termes liés par ces marqueurs qui précisent le domaine :

« Il s'agit en effet de rassembler sous une même relation des structures linguistiques qui peuvent être très différentes. [...] C'est d'ailleurs cette simplification qui permet de stabiliser certains marqueurs et qui autorise à penser que certains marqueurs traversent les domaines et les époques sans dommage (si l'on pense par exemple aux contextes définitoires qui sont déjà inscrits dans Aristote). C'est en tout cas, la seule hypothèse qui permette aux linguistes d'essayer de systématiser l'analyse des marques linguistiques pour accéder à tel ou tel contenu. Il y a simplification (abstraction diraient certains) et donc à la fois perte de sens et gain en efficacité » (Condamines 2003 : 132-133).

Cependant, dans certains énoncés, le lien entre le marqueur et la relation est très dépendant du corpus. Certains marqueurs ne sont pas propres à un genre de corpus et ne sont pertinents que pour un corpus en particulier; ces marqueurs ne sont donc pas prédictibles.

Le fonctionnement même de certains marqueurs peut être lié au genre du corpus. L'auteure en arrive à cette conclusion à la suite de l'étude de trois corpus de genres différents, soit les catalogues de jouets, les petites annonces

immobilières et les descriptions d'itinéraires. Les résultats de son étude montrent que le marqueur *avec* fonctionne surtout comme marqueur de méronymie et quasiment comme le seul marqueur de méronymie. Cependant, dans une œuvre littéraire dépouillée par l'auteure, ce même type de fonctionnement est présent à peine 15 fois sur un total de 616 occurrences.

Pour cette raison, il serait important de caractériser les genres textuels, c'est-à-dire de déterminer les genres de corpus pertinents pour décrire tel ou tel phénomène linguistique. Cette réflexion permettra, par exemple, d'associer un système de marquage à un genre de corpus.

d) Genres textuels et genres interprétatifs

Condamines aborde également la problématique de la non-coïncidence entre genres textuels intuitifs et régularités linguistiques que l'on repère en fonction du point de vue que l'on adopte :

« Ainsi, l'objectif de l'interprétation fait que ce qui apparaissait comme genre textuel stable peut éclater. Par exemple, pour ce qui est du marquage de la méronymie par *avec*, une description d'itinéraires a un fonctionnement plus proche des petites annonces immobilières que les petites annonces immobilières n'en ont avec les petites annonces de vente de voitures » (Condamines 2003 : 148).

Pour conclure, l'auteure suggère des pistes de réflexion pour améliorer le passage d'un corpus à une modélisation de ce corpus :

- 1) Dans le cas des réseaux relationnels, pouvoir définir des marqueurs en lien avec un genre. La recherche d'une représentation sous forme de relation permet d'étudier un certain nombre de phénomènes d'après leur capacité à marquer des relations. De plus, l'hypothèse de la stabilité des marqueurs permet de maîtriser les interprétations possibles et de les expliquer.
- 2) Explorer la possibilité de fonder les notions de genre, de genre textuel et de genre interprétatif sur l'observation de régularités linguistiques. L'auteure suggère que les genres textuels ne soient plus seulement étudiés

sur des bases introspectives. Il faudrait développer une réflexion sur la réalité linguistique des genres textuels et des genres interprétatifs.

2.3.2.9 Synthèse

Tous les travaux dont il a été question dans les sections précédentes ont été menés dans le but de repérer des informations sémantiques de façon automatique ou semi-automatique dans les corpus. Dans cette section, nous résumons ce que nous retenons des travaux que nous avons cités, ce qui nous en distingue, et, enfin, ce que nous croyons apporter à la recherche dans ce domaine.

Certains travaux, dont ceux de Davidson *et al.* (1998) et de Meyer *et al.* (2001), ont cherché à faire l'inventaire, non seulement des patrons lexicaux, mais aussi des patrons paralinguistiques et des patrons grammaticaux qui expriment des relations conceptuelles. Ils les regroupent sous le générique *patron riche en connaissances* ou *patron de connaissances*. Quant à nous, nous avons choisi de concentrer notre recherche sur les patrons lexicaux, c'est-à-dire des mots ou des groupes de mots (comprenant leurs variantes morphologiques).

Séguéla (1999) distingue les marqueurs génériques des marqueurs spécifiques. Les premiers expriment des relations sémantiques de façon plus ou moins stables à travers différents corpus et domaines techniques. Quant aux seconds, ils expriment des relations sémantiques de façon assez précise dans un corpus donné.

Dans ce travail, nous faisons l'hypothèse que les patrons lexicaux sont ceux qui sont susceptibles d'être génériques et, donc, qu'ils devraient être présents dans nos deux corpus.

La majorité des recherches sur les patrons de relations sémantiques ont porté sur des corpus de langue anglaise. Quelques recherches, dont celles de Condamines *et al.* (1999), Meyer *et al.* (1999), Rebeyrolle (2000) et Séguéla (1999) portent sur le français. Nous avons donc choisi d'étudier deux corpus de

langue française puisque seulement un certain nombre de chercheurs s'y sont intéressés avant nous. Nos analyses permettront d'identifier de nouveaux PRS pour cette langue.

Comme l'a clairement démontré Condamines (2003), la définition des genres textuels qui entrent dans la composition du corpus en fonction des structures à décrire est déterminante. Nous avons tenu compte, pour nos travaux, de la nature des textes de nos corpus. En effet, nos corpus de médecine et d'informatique sont tous deux composés de textes didactiques et de vulgarisation, car ces types de corpus étaient tous deux susceptibles de comporter un grand nombre de PRS, comme le suggèrent les études de Loffler-Laurian (1983, 1994), de Pearson (1998), et de Rebeyrolle (2000). Notre recherche vise toutefois à élaborer un modèle qui transcende les différences entre les domaines de spécialité et, par conséquent, entre les corpus. Nous nous distinguons donc, en ce sens, des travaux de Condamines (2003) pour qui la notion de « marqueur » est fortement liée au genre textuel.

En conclusion, nous croyons que notre recherche sur l'intégration d'information sémantique améliorerait grandement le travail d'extraction des contextes contenant l'expression d'une relation sémantique. Ce qui nous distingue des autres auteurs est la vérification de l'hypothèse selon laquelle il est possible de mettre en relation des PRS et des TS. Notre modèle de dépendance entre le TS et les PRS serait éventuellement utilisé en reconnaissance ou en repérage automatisé : une fois repérés, ces PRS pourraient servir de guide pour trouver rapidement – à l'aide d'un concordancier ou d'un logiciel d'extraction automatique ou semi-automatique de connaissances – des éléments descriptifs sur les termes, comme une définition, une description, etc.

3. MÉTHODOLOGIE

Le présent chapitre, expose, dans l'ordre, les démarches qui nous ont permis de mener à terme notre étude et d'effectuer une synthèse de nos résultats. Bien qu'informatisée, notre recherche terminologique respecte le modèle de recherche terminologique standard, c'est-à-dire que les principes de base reconnus en terminologie ont été respectés, du moins dans les premières étapes.

Nous décrivons d'abord le type de textes choisis pour notre étude et la méthode de mise en forme des corpus (3.1). Dans cette section, nous donnons le détail de nos corpus (leur taille, les textes choisis et les niveaux de spécialisation). Puis, nous expliquons les critères utilisés dans la sélection des termes (3.2) : nous donnons notre définition de « terme », nous expliquons comment s'est fait le choix des termes clés à partir desquels les contextes ont été recherchés, et nous traitons de la fréquence des termes et des contextes pertinents dans le contexte de notre étude.

Dans la section 3.3, nous traitons de l'extraction des contextes et des critères qui ont présidé à leur sélection. Puis, nous abordons le thème du traitement des patrons de relations sémantiques (PRS) (3.4). Ensuite, nous traitons du regroupement des termes par type sémantique (TS) qui s'inspire de la typologie de WordNet (3.5). Cette section justifiera le choix des grands génériques de WordNet dans notre travail.

Enfin, nous décrivons les données recensées dans notre base de données (3.6). Dans cette section, nous expliquons la façon dont les contextes sont traités. Enfin, nous décrivons comment nous associons les PRS aux TS (3.7).

3.1 Choix des textes et mise en forme des corpus spécialisés

Cette section traite du choix des textes et de la mise en forme de nos corpus spécialisés. Comme nous l'avons vu au chapitre précédent, le terme *corpus* doit comprendre, pour les fins de notre étude, les notions de « collection », de « langue » de « spécialité », d'« échantillon », de « représentativité » et d'« objectifs de recherche ».

3.1.1 Choix des textes

Pour le choix des textes, nous nous sommes inspirée des principes évoqués dans la section 2.1. D'abord, nous avons sélectionné des textes qui ont été rédigés par des spécialistes du domaine afin d'avoir accès à la terminologie la plus représentative de chacune des langues de spécialité étudiées. Le corpus de l'étude est composé de textes appartenant à deux domaines de spécialité, à savoir la médecine (et plus précisément la cardiologie, l'imagerie, l'optométrie et l'ophtalmologie) et l'informatique (plus précisément l'équipement, Internet, les logiciels, les réseaux, les systèmes d'exploitation et les virus).

Ensuite, seuls les textes récents ont été retenus, c'est-à-dire ceux qui ont été publiés (en fonction de la date d'édition et non d'une date de réimpression) après le début des années 1990. Précisons que la majorité des textes, c'est-à-dire ceux puisés dans le Web, datent de la fin des années 1990 et du début des années 2000.

Les textes que nous avons sélectionnés proviennent de sources variées, comme en fait foi le nombre de documents qui composent chacun des corpus. En effet, nous avons déjà souligné l'importance de diversifier les auteurs ainsi que les collectivités d'appartenance, pour assurer une meilleure représentativité des usages. De plus, nous avons prêté une attention toute spéciale au choix des contextes, c'est-à-dire que nous nous sommes assurée d'avoir un nombre suffisant de contextes rédigés par des auteurs différents pour appuyer notre analyse. (Se

reporter à l'annexe A qui présente la liste des textes ainsi que des détails sur chacun d'eux : le nom de fichier, le domaine de spécialité, le contenu général, le niveau de spécialisation et le nombre de mots.)

3.1.2 *Taille des corpus*

Comme la majorité des grands corpus utilisés en terminologie, nos corpus d'informatique et de médecine contiennent à eux deux plus d'un million de mots. Nous nous sommes également assurée de travailler avec des corpus comportant à peu près le même nombre de mots pour chacune des langues de spécialité étudiée, soit un peu plus d'un demi million de mots chacun, afin de les équilibrer.

Au total, les deux corpus sont composés de 189 documents et contiennent 1 251 592 mots, répartis comme suit :

Corpus de médecine : 117 documents, 558 349 mots.

Corpus d'informatique : 72 documents, 693 243 mots.

3.1.2.1 *Disponibilité des textes*

Les textes de médecine et d'informatique qui composent chacun de nos corpus d'étude font partie de la base textuelle du groupe de recherche « Observatoire de linguistique sens-texte » (OLST) du Département de linguistique et de traduction de l'Université de Montréal.

Les textes ont été, pour la très grande majorité, puisés dans le *Web* et, puisqu'ils étaient déjà en format électronique, ils ont tout simplement été téléchargés pour être traités. D'autres textes (ceux diffusés sur du matériel d'impression comme un livre ou une revue) ont été numérisés à partir d'ouvrages spécialisés. Les textes numérisés ont été corrigés au moyen d'un vérificateur d'orthographe.

3.1.3 Niveaux de spécialisation

Puisque nous étudions les relations entre les TS et les PRS, les textes choisis devaient, idéalement, renfermer le plus grand nombre d'éléments descriptifs possibles sur des concepts de la discipline, c'est-à-dire des définitions et des explications. Ainsi, rappelons que nous nous sommes inspirée de la typologie de Rebeyrolle (2000).

Nous avons retenu, principalement, les textes que Rebeyrolle (2000) appelle *textes didactiques* et *textes de vulgarisation*. Les textes appartenant au premier groupe sont rédigés par des spécialistes du domaine et s'adressent à des étudiants de niveau universitaire. Les textes appartenant au deuxième groupe sont rédigés par des spécialistes et s'adressent au grand public, mais non à des spécialistes ni à des spécialistes en devenir. Ces textes ont bien entendu été préférés à d'autres, car nous tenons pour acquis qu'ils contiennent un grand nombre d'éléments explicatifs sur les termes.

Comme nous l'avons dit à la section 2.1.3.3, Rebeyrolle (2000) a démontré que les textes didactiques comportent plus d'énoncés définitoires (constituant une sous-catégorie des contextes que nous étudions) que les textes appartenant à d'autres niveaux de spécialisation. Voilà pourquoi près de la moitié des textes choisis pour notre propre étude sont de nature didactique. En effet, les textes de médecine (sauf ceux d'optométrie) s'adressent principalement à des étudiants de niveau universitaire et sont donc plus spécialisés que ceux d'informatique ou d'optométrie.

Rappelons également que nous avons choisi d'inclure des textes de vulgarisation dans nos corpus. Bien que nos textes de vulgarisation soient destinés au grand public, il ne s'agit pas d'articles encyclopédiques ni d'articles de journaux, mais d'ouvrages d'informatique, d'articles de revues spécialisées en informatique, ou, encore, de sites spécialisés en optométrie. En ce qui concerne les textes d'informatique, les textes de vulgarisation retenus s'adressent à des lecteurs qui ont un intérêt pour l'informatique, sans être des experts du domaine.

Quant aux textes de vulgarisation choisis en optométrie, ils ont pour objectif de renseigner un patient ou un membre de sa famille sur un problème de santé ou un traitement. Ces textes de vulgarisation sont utilisés pour démystifier la terminologie qui pourrait paraître rébarbative sans les définitions qui l'accompagnent. Puisqu'un soin a été apporté à la spécialisation des textes, les contextes contenant l'explication d'une relation sémantique devraient donc, en principe, être bien représentés.

Le Tableau 10 montre la répartition des textes composant nos corpus en fonction des deux domaines de spécialité choisis, le nombre de mots et de documents pour chacun des domaines, ainsi que les sous-domaines et les niveaux de spécialisation auxquels les textes sont associés.

Tableau 10. Tableau récapitulatif du contenu du corpus de l'étude

Nombre de mots	Nombre de documents	Langues de spécialité	Sous-domaines	Niveaux de spécialisation des textes
558 349 mots	117	Médecine	Cardiologie	Les deux corpus comportent des textes didactiques et des textes de vulgarisation (pour plus de détails, consulter l'annexe A)
			Imagerie	
			Ophthalmologie	
			Optométrie	
693 243 mots	72	Informatique	Équipement	
			Internet	
			Logiciel	
			Réseau	
			Système d'exploitation	
			Virus	

3.2 Sélection des termes

Cette section traite de la sélection des termes qui serviront de points d'entrée aux corpus et aux contextes contenant un patron de relation sémantique. Nous définissons d'abord ce que nous entendons par « terme » (3.2.1), puis, nous expliquons comment s'est fait le choix des termes clés (3.2.2), enfin, nous traitons de la fréquence des termes et des contextes pertinents que nous avons retenus (3.2.3).

3.2.1 *Notion de terme*

Le terme est une unité lexicale signifiante qui représente une notion délimitée au sein d'un domaine de spécialité, l'informatique et la médecine en l'occurrence. C'est cette définition de *terme* que nous retiendrons pour notre étude. Autrement dit, nous écartons les unités lexicales dont le sens n'est pas lié aux domaines qui nous intéressent et les sens non spécialisés que peuvent avoir les termes choisis au préalable. Nous retenons également une délimitation technique qui fait l'objet du paragraphe suivant.

Le terme est une unité lexicale spécialisée qui est définie dans le contexte et que nous avons choisi préalablement pour faire une requête (ex. : *ordinateur*, *maladie*, etc.). Nous appelons également *terme*, l'hyponyme d'un terme choisi pour la requête, tels *ordinateur portable* pour *ordinateur*, ou encore *maladie de Binswanger* pour *maladie*. Nous reviendrons sur l'application de cette précision technique un peu plus loin et sur son implication en regard de l'extraction des contextes.

Les exemples 44 et 45 contiennent des termes associés à des contextes extraits de nos corpus.

Informatique :

- 44. Une *imprimante* **sert à** imprimer des documents suivant diverses qualités (antho).
- 45. Un *ordinateur* **est** une machine permettant de manipuler des données (conres).

Médecine :

- 46. Le *scanner* **sert à** guider une biopsie percutanée et peut influencer sur le traitement suggérant la nécessité d'un drainage précoce d'un abcès, d'un débridement pour ostéomyélite récurrente ou d'une fusion osseuse (rach).
- 47. Le *génie logiciel* **est** une science de l'ingénieur [...] (genilo).

Imprimante (44) et *ordinateur* (45) sont des termes d'informatique et *scanner* (46) et *génie logiciel* (47) sont des termes de médecine. En effet, ces unités désignent une notion à l'intérieur des domaines qui nous intéressent.

3.2.2 *Choix des termes clés*

C'est à partir des termes que nous recherchons les PRS dans les corpus. Pour les choisir, nous avons eu recours, en premier lieu, à des thésaurus, à des lexiques, à des vocabulaires terminologiques et à des dictionnaires spécialisés afin de déterminer quels sont les termes les plus représentatifs de chacun des domaines. Nous avons tenu pour acquis que des termes qui étaient définis ou expliqués dans des ouvrages de base étaient plus susceptibles d'être jumelés à des PRS. Nous augmentions aussi nos chances de trouver un nombre élevé de contextes pertinents dans nos corpus.

Nous avons également choisi, lorsque c'était possible, 2 termes par corpus et par TS. Au total, nous comptons 17 termes en médecine et 16 termes en informatique. La question des types sémantiques fait l'objet de la section 3.5.

Nous y expliquons également la raison pour laquelle le nombre de termes n'est pas le même dans les deux corpus.

3.2.3 *Fréquence des contextes pertinents dans le corpus et diversité des auteurs*

Une fois les termes choisis, il a fallu s'assurer que leur fréquence dans les corpus était suffisante pour appuyer notre analyse. Nous avons fixé un seuil de fréquence minimale à 20 occurrences²² du terme simple ou du terme complexe. De plus, afin de tenir compte des critères de représentativité, dont il a été question à la section 2.1.2, il était nécessaire que le nombre de contextes étudiés aient été rédigés par au moins deux auteurs²³ différents pour chacun des termes. Ces fréquences minimums peuvent paraître peu élevées, mais en moyenne, le nombre de contextes total étudié ainsi que la diversité des auteurs pour chaque terme dépasse largement cette fréquence minimum.

3.2.4 *Liste des termes clés retenus*

Le Tableau 11 présente la liste des termes clés retenus, classés par domaine. À l'annexe B, nous présentons cette même liste de termes clés accompagnés de leurs définitions, car c'est dans leurs sens techniques que les termes devraient être employés dans les contextes retenus.

²² Nous avons fait une exception pour un seul terme, à savoir *vieillesse*, qui apparaît seulement 7 fois dans le corpus de médecine. Nous l'avons conservé, car nous avons tout de même pu retenir 3 contextes pertinents au total. Nous sommes cependant consciente qu'il sera difficile de tirer des conclusions solides pour ce terme.

²³ Nous avons fait une exception pour un seul terme, à savoir *médecine*, pour lequel les deux contextes enregistrés proviennent de la même source. Nous l'avons conservé puisque les termes du TS auquel appartient ce terme, [PSYCHOLOGICAL FEATURE (FIELD)] en l'occurrence, sont peu souvent jumelés à une relation sémantique.

Tableau 11. Termes retenus pour chacun des corpus

Termes de médecine	Termes d'informatique
1. Angiographie	1. Bogue (bug) ²⁴
2. Cathéter	2. Codage
3. Chirurgie	3. Développement
4. Décès	4. Erreur
5. Développement	5. État
6. Douleur	6. Évolution
7. Électrocardiogramme (ECG) ²⁵	7. Génie
8. État	8. Imprimante
9. Maladie	9. Informatique
10. Médecin	10. Langage
11. Médecine	11. Logiciel
12. Patient	12. Ordinateur
13. Scanner	13. Panne
14. Symptôme	14. Programmation
15. Systole	15. Programmeur
16. Traitement	16. Utilisateur
17. Vieillessement	

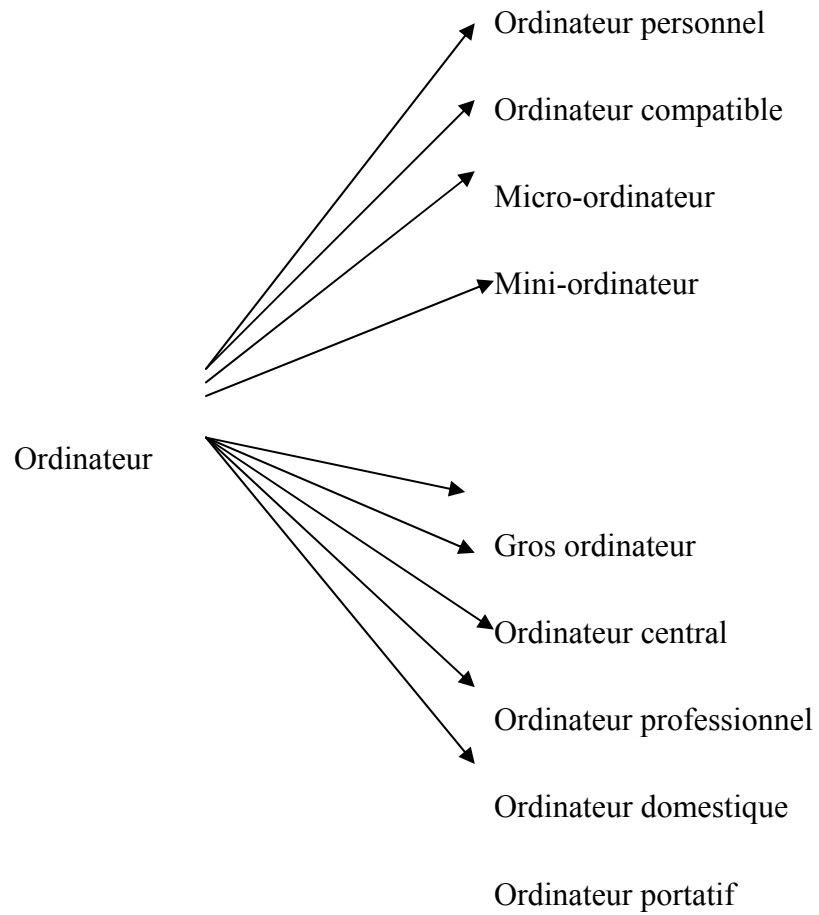
²⁴ Nous avons effectué la requête à partir de ces deux graphies, soit *bogue* et *bug*.

²⁵ Nous avons non seulement effectué la requête avec le terme *électrocardiogramme*, mais également avec l'abréviation de ce terme, soit ECG.

Chacun de ces termes simples constitue un point d'accès aux contextes pertinents des corpus et ils peuvent donner lieu à des termes complexes. Les termes complexes composés de l'un de ces termes simples sont également pris en compte pour notre étude si le terme simple en question constitue la tête du terme complexe. Par exemple, les termes *imprimante laser* et *imprimante matricielle* sont considérés puisqu'ils sont formés à partir du terme simple *imprimante*.

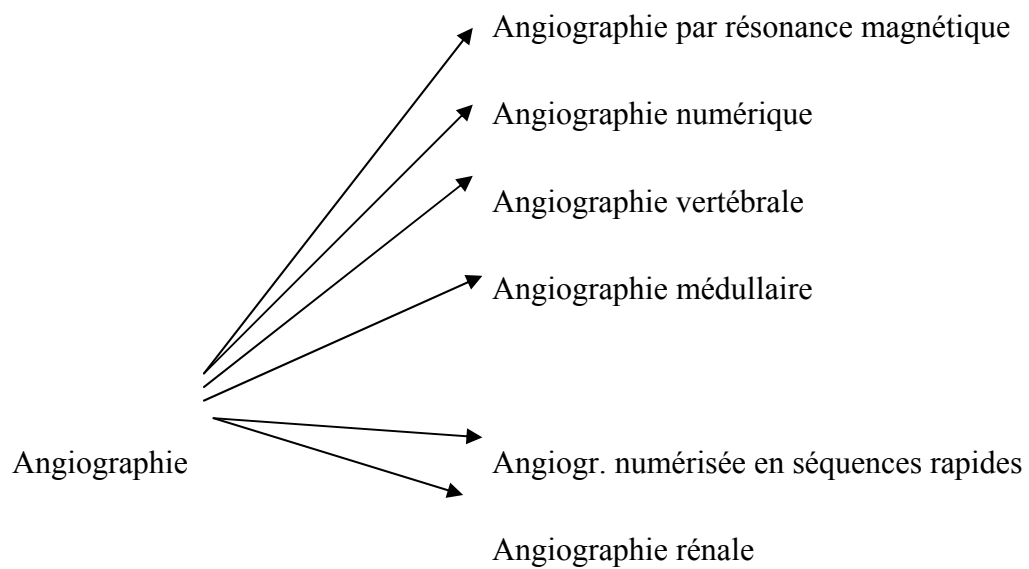
La Figure 5 montre 9 termes complexes formés à partir d'*ordinateur*. Ils ont été acceptés, car ils constituent des hyponymes du terme de la requête, soit *ordinateur*.

Figure 5 : Termes complexes repérés au moyen du terme clé *ordinateur*



La figure 6 montre 6 termes complexes, repérés dans le corpus de médecine, formés à partir d'*angiographie*. Ils ont également été retenus en vertu du principe expliqué plus haut.

Figure 6 : Termes complexes repérés au moyen du terme clé *angiographie*



Lors du dépouillement, nous considérons comme termes ceux qui ont préalablement été choisis pour une requête. Il arrive parfois qu'un contexte contienne le terme, un PRS et que l'information donnée sur le premier terme prenne la forme d'un terme également plutôt que celle d'une périphrase. Dans ce cas, nous appelons *terme* celui pour lequel une requête a été faite. C'est le cas d'une requête portant sur le terme *maladie* dans l'exemple 48.

48. La rétinopathie diabétique est potentiellement une *maladie oculaire* sérieuse **causée par** le diabète (ret_diab).

L'exemple 48 contient un contexte contenant le terme *maladie oculaire*, le PRS **Y cause X**, et l'information se rapportant au terme, soit le diabète. Ce contexte a donc été analysé du point de vue du terme *maladie oculaire*.

L'exemple 49 montre un contexte où le terme de la requête, *ordinateur* se trouve de part et d'autre du PRS. Nous considérons comme terme celui qui vient en premier à partir de la gauche dans le texte.

49. Nous nous intéresserons dans la suite qu'aux *ordinateurs de type PC*, **appelés** aussi ordinateurs compatibles IBM... (inordi).

L'exemple 49 contient un contexte contenant le terme *ordinateurs de type PC*, le PRS **X appelé Y**, et l'information se rapportant au terme, soit ordinateurs compatibles IBM. Ce contexte a donc été analysé du point de vue du terme *ordinateurs de type PC*.

3.3 Extraction des contextes

Un *contexte* est jugé *pertinent* lorsqu'il comprend trois éléments isolables : le terme, le PRS et l'information sémantique. Le contexte doit donc contenir un PRS de nature lexicale qui fait ressortir la relation sémantique, les éléments descriptifs (information) qui se rapportent au terme, et bien entendu le terme de la requête, comme dans l'exemple 50 :

50. L'*imprimante* (line printer) **permet** d'obtenir des résultats imprimés (zmspc1).

C'est lors du dépouillement du corpus que nous prenions la décision d'accepter ou de rejeter le contexte contenant le terme de la requête. Par exemple, nous devons nous assurer que l'emploi d'un PRS correspondait bien à notre propre définition. Nous avons fait intervenir un certain nombre de critères. Les premiers concernent le terme de la requête, alors que le second groupe concerne les patrons. Ces critères sont décrits dans les sections qui suivent.

3.3.1 *Extension de sens du terme*

Le premier critère consistait à s'assurer que le terme, dans le contexte où il se trouve, a le sens spécialisé que nous avons retenu au départ (voir l'annexe B). Par exemple, si le terme choisi est *langage* en informatique, nous nous sommes concentrée sur les contextes où ce terme a le sens de « langage de programmation » et non pas celui de « langage humain ». L'exemple 51 contient le terme *langage naturel*, type de langage qui ne correspond pas à notre définition. Ce contexte est donc rejeté. Cependant, l'exemple 52 contient le terme *langage de programmation*, terme qui appartient au domaine de l'informatique et qui a le sens que nous avons décrit préalablement.

51. Le *langage naturel* **est** le langage que les êtres humains utilisent pour communiquer, soit oralement, soit par écrit (linux3p3) .
52. ... il existe un *langage de programmation* qui **permet** de coder directement le binaire... (linux3p3).

3.3.2 *Reprise anaphorique*

Nous retenons également les contextes dans lesquels les termes sont repris sous forme anaphorique, comme dans les exemples 53 et 54 :

53. *Maladie de Binswanger* **C'est** une maladie de la substance blanche due à une artériosclérose des artères distales, qui s'observe habituellement chez les sujets âgés et hypertendus ou diabétiques (image1).
54. Le *scanner* permet aussi d'identifier la complication majeure du traumatisme crânien à cet âge, le gonflement cérébral malin; *il permet* enfin de documenter le diagnostic de mort cérébrale, encore qu'il n'a pas été prouvé que dans ce cas, il apporte la même certitude que l'angiographie (cran).

3.3.3 *Forme du PRS*

La deuxième série de critères concerne la forme du PRS. Il consiste à vérifier que le contexte contient bien un PRS pour le terme en question, c'est-à-dire un mot ou une suite de mots qui unissent un terme à des renseignements sur son sens. Nous nous concentrons sur les PRS lexicaux en nous appuyant sur l'hypothèse voulant qu'ils soient présents dans nos deux corpus.

Nous sommes consciente que, à cause de ce critère, nous perdons des verbes ou des adjectifs sémantiquement pleins (patrons grammaticaux). Toutefois, nous croyons que les relations sémantiques exprimées par ces verbes ou ces adjectifs sont liées à un corpus en particulier. De plus, ces patrons ne sont pas prévisibles. Enfin, nos corpus étant suffisamment volumineux, nous devrions obtenir un nombre de contextes pertinents satisfaisant, même en nous concentrant uniquement sur les patrons lexicaux.

Les exemples 55, 56, 57 et 58 rappellent la différence entre les PRS lexicaux et les autres PRS (appelés *grammaticaux* par Davidson *et al.* (1998)) et montrent de quelle manière nous avons appliqué cette distinction à nos propres textes.

55. Du plus petit calculateur de poche aux plus puissantes machines, les *ordinateurs* **sont** universellement **utilisés** **pour** calculer numériquement (access1).

56. Les *ordinateurs* **permettent** à l'utilisateur de traiter et de communiquer de l'information (adevim).
57. L'*ordinateur* **communique** avec l'écran grâce à une carte vidéo branchée sur sa carte maîtresse (delavi1).
58. ...l'*ordinateur* **transmet** ce qui doit être affiché à l'écran par le biais d'une interface appelée contrôleur vidéo (plaisen1).

Les exemples 55 et 56 montrent que les PRS **on utilise X pour Y** et **X permet Y** sont des PRS lexicaux qui relient un terme à une information sémantique exprimée de façon distincte. Ces contextes sont donc acceptés. Quant aux exemples 57 et 58, il s'agit de PRS grammaticaux (**X communique** et **X transmet**), c'est-à-dire que l'information sémantique est exprimée par le verbe lui-même. Ces contextes sont donc rejetés.

Enfin, rappelons que, dans la formulation des PRS, **X** représente le terme de la requête, et **Y** représente l'information qui se rapporte au terme. C'est la raison pour laquelle ces variables se trouvent tantôt à gauche, tantôt à droite du patron. Leur ordre n'est pas aléatoire et est pris en compte dans l'analyse des données.

3.3.4 *Difficultés liées à certains patrons*

Pour qu'un contexte soit pertinent et accepté, nous avons choisi de retenir tout contexte contenant des éléments descriptifs (de l'information) qui se rapportent au terme en question.

- a) Contrairement à Rebeyrolle (2000), nous avons choisi de retenir des contextes dans lesquels apparaît une **négation**, comme dans l'exemple 59 :
59. Les *imprimantes à jet d'encre* **ne sont pas conçues pour** de grosses impressions. (traum2)

- b) Nous acceptons également des contextes contenant des PRS accompagnés d’adverbes (ou de locutions adverbiales) du type : *généralement, souvent, à l’occasion, etc.*, comme dans l’exemple 60, où l’adverbe *souvent* est identifié par un soulignement double.

60. *Le traitement chirurgical* adapté à chaque cas **permet souvent** la normalisation spectaculaire et définitive des chiffres tensionnels (xgiroud).

Cependant, contrairement à la forme négative, lors de la neutralisation, l’adverbe n’est pas représenté dans le patron.

- c) En extraction de contextes semi-automatisée, la distance qui sépare un PRS du terme de la requête peut engendrer des silences. Si trop de mots séparent ces deux éléments, la relation risque de ne pas être repérée. Par exemple, un logiciel d’extraction de contextes pourrait ne pas pouvoir détecter la relation causale exprimée dans l’exemple 61 (20 mots séparent le terme de la requête et le patron).

61. Le point important est que toute *douleur* venant révéler, ou émailler l’évolution d’un anévrisme jusque-là quiescent, **doit** être considérée comme un signe d’alarme, et **faire craindre** – jusqu’à preuve du contraire – une fissuration de l’anévrisme (anevrism).

Ce type de contexte est accepté et conservé dans notre analyse, bien qu’en extraction semi-automatisée, il puisse être difficile à repérer.

- d) Plus d’un patron dans un même contexte :

Si, dans un même contexte, il y a présence de plusieurs patrons, nous relevons chacun de ces patrons afin qu’ils soient tous analysés, comme dans l’exemple 62 :

62. **Parmi** eux [*langages modernes*], **on compte** *Java, C++, Python, Perl*, ou **d’autres** *langages spécialisés* et moins connus **comme** Narval, dédié à la mise en œuvre d’assistants personnels intelligents (contrat).

Dans l'exemple 62, il y a présence de trois relations hyperonymiques avec les patrons **parmi X on compte Y**, **Y d'autres X** et **X comme Y**. Ce contexte donne donc lieu à trois entrées différentes dans la base de données.

- e) Rejet des contextes de relations causales dont le lien de cause à effet est ambigu

Un patron généralement associé à une relation sémantique peut ne pas exprimer de relation dans un contexte donné. Par exemple, bien que le patron **X lié à Y** exprime généralement une relation causale, dans l'exemple 63, on ne peut pas vraiment dire que le sexe soit la **cause** de la *maladie* en question. Ce contexte est classé dans la catégorie des patrons ambigus et est donc rejeté

63. L'adrénoleucodystrophie est une *maladie récessive* **liée au** sexe (image7).

Dans certains contextes de relations causales, il est difficile de s'assurer qu'il y a bien présence d'une relation de cause à effet. Il se peut, comme c'est le cas dans l'exemple 64, qu'il s'agisse de la provenance plutôt que de la cause comme telle.

64. ... de *douleurs d'origine* ovarienne qu'il s'agisse d'un simple kyste, d'une infection, d'une torsion annexielle ou plus rarement d'une authentique tumeur (doul).

Le contexte 64, tiré de notre corpus de médecine, est rejeté puisque ovarienne indique la provenance et non la cause des douleurs.

- f) Relations méronymiques inclassables selon la typologie de Winston *et al.* (1987)

Nous qualifions d'inclassables les relations méronymiques qui ne font pas partie de l'un des 6 types de la typologie de Winston *et al.* (1987) (voir la section 2.2.2). Nous avons rencontré des contextes de ce genre dans notre corpus d'informatique pour les termes clés *bogue* et *erreur*. Nous avons reproduit des exemples (65 et 66) pour chacun de ces termes.

65. Ce fabricant propose une ROM enfichable contenant le codage EPSON standard, mais des essais nous ont montré que cette ROM **comportait** beaucoup trop de *bugs*! (impri).
66. Malheureusement, les programmes écrits dans un tel langage **contenaient** souvent des *erreurs* qui n'étaient pas toujours faciles à corriger (pierre1).

Les contextes 65 et 66 sont acceptés et qualifiés d'inclassables puisque nous considérons que ces contextes expriment bien une relation méronymique, mais qu'ils ne correspondent à aucun des types proposés par Winston *et al.* (1987).

3.4 Traitement des PRS

Dans cette section, nous expliquons comment les patrons trouvés dans les contextes sont ramenés à une forme neutre pour faciliter l'analyse. D'une manière générale, les patrons sont représentés de la manière suivante :

X : terme choisi pour la requête (ex. : *ordinateur*, *vieillesse*, etc.);

Le patron sous forme neutralisée (les détails de cette neutralisation sont donnés ci-après);

Y : mots ou ensemble de mots associés sémantiquement au terme.

a) Voie active, voie passive

Les patrons contenant des verbes employés à la voix passive sont transformés à la voix active. Les seuls cas qui font exception sont ceux où un même verbe est toujours employé à la voix passive. Dans ce cas, nous conservons la voix passive puisque seul le passif est possible, comme dans les exemples 67 et 68.

67. [...] ces *maladies* [AFFECTIONS DYSMYÉLINISANTES OU LEUCODYSTROPHIE] **sont liées** à un trouble métabolique d'origine enzymatique (image 7).
68. C'est [maladie de Binswanger] une *maladie de la substance blanche* **due** à une artériosclérose des artères distales (image1).

Dans l'exemple 67, on ramène le patron à la forme **X lié à Y**, et non **on lie X à Y**, car ce patron est toujours employé à la voix passive dans nos corpus. Même chose en 68, où l'on ramène le patron à la forme **X dû à Y**, et non **on doit X à Y**. Il arrive aussi que des patrons apparaissent tantôt à la voix passive 69, tantôt à la voix active 70, et tantôt comme substantif 71.

69. La rétinopathie diabétique est potentiellement une *maladie oculaire* sérieuse **causée par** le diabète (ret_diab).
70. En effet une *panne* de ventilateur **peut** directement **causer** la mort de votre processeur (optrep).
71. ... remplacer l'aorte ascendante, afin de mettre le patient à l'abri de la **cause** première **de** *décès*, à savoir la rupture de l'aorte en intra-péricardique (dissecti);

Les exemples 69, 70 et 71 montrent un patron employé une fois à la voix passive, en 69, une fois à la voix active, en 70, et, une fois, sous forme de nominalisation en 71. Pour simplifier la notation, nous ramenons ce type de patron à la forme **cause de X Y** lorsque *cause* est substantif, et à la forme **X cause Y** lorsque *cause* est un verbe, qu'il soit à la voix active ou à la voix passive.

b) Genre et nombre

Les noms et adjectifs contenus dans les patrons sont ramenés au masculin, singulier. Le contexte de l'exemple 72 contenant un patron **Y telles X** (féminin pluriel) est ramené à **Y tel X** (masculin singulier).

72. Cette configuration, assez représentative d'un système relativement modeste, pourrait être complétée par des unités périphériques **telles** : une *imprimante*, une unité de disque rigide, une unité de cassettes de bandes magnétiques, etc. (zmspc1).

c) Temps et mode

Tous les verbes sont ramenés au présent de l'indicatif et au singulier. L'exemple 73 montre un contexte dans lequel le verbe à l'imparfait est ramené au présent de l'indicatif lors de la neutralisation.

73. L'*ordinateur* **rendait possible** l'interactivité (ecran).

Le patron **X rendait possible Y** (imparfait) est ramené à la forme **X rend possible Y** (présent de l'indicatif).

L'exemple 74, où le patron **X sont Y** (pluriel), est ramené à la forme **X est Y** (singulier).

74. *Mini-ordinateurs* et *micro-ordinateurs* **sont** donc des ordinateurs de taille restreinte comparables en dimension à un petit bureau. (pierre1).

d) Adverbes modifiant le patron

Les adverbes (*généralement, toujours, rarement, encore, etc.*) modifiant le patron sont éliminés lors de la neutralisation. Nous considérons que ces adverbes ne modifient que superficiellement la relation sémantique exprimée par un

patron²⁶. On ramène simplement le patron de l'exemple 75 à la forme : **X se compose de Y**.

75. Par contre, les *gros ordinateurs se composent* généralement **de** plusieurs unités physiques interconnectées en un ensemble dont la taille est souvent très grande (pierre1).

e) Verbes modaux

Les verbes modaux, comme *pouvoir* et *devoir* sont également écartés lors de la neutralisation. On ramène simplement le patron à la forme : **on utilise X pour Y**, comme dans l'exemple 76.

76. Dans cette section, nous allons nous attacher à comprendre, au niveau des principes et non de la technique, comment il se peut qu'un dispositif purement physique, un *ordinateur*, **puisse être utilisé pour** résoudre des problèmes abstraits, tels que ceux rapidement décrits dans la première section (aces1).

3.4.1 Exemples de neutralisation des patrons

Les Tableaux 12, 13, 14, 15 et 16 donnent des exemples de patrons neutralisés pour chacune des relations sémantiques étudiées, soit la relation hyperonymique, la relation méronymique, la relation causale, la relation de finalité ainsi que les patrons métalinguistiques.

²⁶ L'adverbe *jamais* aurait fait exception à la règle s'il avait été présent dans notre corpus. Nous ne l'aurions pas conservé comme tel lors de la neutralisation, mais le sens aurait été reflété dans le patron neutralisé grâce à la forme négative.

Tableau 12. Patrons neutralisés de la relation hyperonymique

Patron neutralisé	Patrons avant neutralisation
Autre Y X	Les autres Y X
Y comme X	Des Y comme la X
X est Y	La X est une Y, le X est un Y, X était le Y, X furent le Y
Classe de Y X	On peut subdiviser les Y en trois classes : X...
Exemple de X Y	X par exemple le Y, X par exemple Y
Catégorie de X Y	X se divise en deux catégories : Y..., X classé en deux catégories : Y...
Type de X Y	X sont de trois types : Y..., deux types de X : Y...
X comme Y	X comme Y, les X comme Y, des X comme Y
X en particulier Y	X en particulier le Y
X tel Y	X tel que Y, X tels que Y

Tableau 13. Patrons neutralisés de la relation méronymique

Patron neutralisé	Patrons avant neutralisation
X contient Y	Y contenue dans le X, Y contenu dans le X
Y dispose de X	Y disposant d'un X
X comporte Y	X comportent un Y, Y que comporte le X
Sous-ensemble de X Y	Y sous-ensemble d'un X, Y sous-ensemble du X
X dispose de Y	X disposant de Y, plusieurs X disposent de Y, un X dispose de Y, certains X disposent d'un Y
X comporte Y	Une X comporte un ensemble de Y
Y comporte X	Cette Y comportait X, X contenant le Y
X possède Y	X possèdent un Y, l'X possède une Y

Tableau 14. Patrons neutralisés de la relation causale

Patron neutralisé	Patrons avant neutralisation
Responsable de X Y	Y étant responsable de presque tous les X, Y responsables de X
Cause de X Y	La cause de X est l'Y, X à cause de l'Y
X dû à Y	Un X dû à ce Y, X due à une Y
X lié à Y	X étroitement liée aux Y, Y liés à la X, Y liées à la X, X sont liées à un Y
X conduit à Y	Une telle X conduit à des Y, les X qui conduisirent à l'Y
X provient de Y	X peuvent provenir de la Y, X proviennent d'une Y
X dépend de Y	X dépend de sa Y, X dépend de la Y, X dépendra du Y
en cas de X Y	Y en cas de X, en cas d'X l'Y
X entraîne Y	X entraînent des Y, X entraîne des Y
X provoque Y	X provoque parfois des Y, X provoque une Y

Tableau 15. Patrons neutralisés de la relation de finalité

Patron neutralisé	Patrons avant neutralisation
X sert à Y	X sert à Y, une X sert à Y, X qui vous sert à Y, X ne sert plus uniquement aux Y
X destiné à Y	X destiné lui à Y, X destiné à l'Y
X conçu pour Y	X a été conçue pour Y
But de X Y	X dont le but est de Y, le but de ce X est de Y, le but du X est de Y
On utilise X pour Y	Il est possible d'utiliser un X pour Y, X sont utilisés pour Y
Rôle de X Y	Le rôle de X est de Y, X dont le rôle est de Y
Objectif de X Y	L'objectif de X est de Y, l'objectif d'un X est de Y
X permet Y	X permettant le Y, X permet de Y, X a permis de Y, X permet la Y, X permettant de Y
On réserve Y à X	Y est réservé au X, Y sont réservés au X

Tableau 16. Neutralisations des patrons métalinguistiques

Patron neutralisé	Patrons avant neutralisation
On appelle X Y	Y s'appelle la X, on appelle X un Y
X appelé Y	X aussi appelée Y, X appelés aussi Y
X c'est-à-dire Y	X c'est-à-dire Y
On parle de X plutôt que de Y	On parle d'un X plutôt que d'un Y
Terme X s'oppose à Y	Ce terme s'opposant à celui de Y
X désigne Y	Le X désigne des Y

3.4.2 *Regroupement des patrons*

À cette étape, nous regroupons les patrons en fonction de la relation sémantique qu'ils établissent, en nous inspirant de la terminologie employée par Ahmad *et al.* (1992) et Loffler-Laurian (1994). Ce regroupement a fait l'objet de quelques aménagements particuliers qui sont abordés dans les sections qui suivent.

3.4.2.1 Problèmes d'ambiguïté de certains patrons

Afin de respecter le sens des PRS, leur neutralisation peut varier même si a priori ils semblent similaires. Les contextes 77 et 78 illustrent ce cas particulier :

77. UNIX, par exemple, est un système d'exploitation écrit en un langage de programmation évolué **appelé** C (zmspc2).
78. Le logiciel d'exploitation, **appelé** aussi système d'exploitation (operating system), affecte directement les opérations du micro-ordinateur (zmspc2).

Le patron neutralisé du contexte 77 est **on appelle X Y**, car *langage de programmation évolué* est un contexte définitoire qui n'est pas interchangeable avec C. Par contre, dans le contexte 78, le patron neutralisé est **X appelé Y**, car *logiciel d'exploitation* est synonyme de système d'exploitation. On ne ramène pas le patron **on appelle X Y** à **on appelle X Y** afin de distinguer des autres ce patron métalinguistique exprimant la synonymie.

Plutôt que d'exprimer une véritable relation sémantique, un patron peut établir un lien entre un nom et un adjectif qualificatif. Dans l'exemple 79, l'adjectif est souligné.

79. Le *traitement médical* **est** parfois suffisant mais il doit être mis en oeuvre le plus tôt possible (xabreg).

L'exemple 79 montre un patron **X est Y** qui exprime généralement une relation hyperonymique, mais qui, dans ce contexte, sert de copule. Ce type de contexte n'est tout simplement pas conservé, puisqu'il n'exprime pas l'une des relations sémantiques que nous étudions dans le cadre de la présente recherche.

3.4.2.2 Regroupement de certains patrons aux fins de l'interprétation des résultats

Dans le chapitre 4, nous avons regroupé les PRS similaires exprimant une même relation afin de présenter les patrons les plus fréquents et la répartition des PRS dans les corpus. Par *similaire*, nous entendons des patrons qui peuvent comporter quelques variantes même après la neutralisation, mais qui sont semblables du point de vue de la forme et expriment la même relation sémantique. En effet, deux patrons peuvent être similaires même si l'un d'eux est représenté par un substantif, alors que l'autre est représenté par un verbe, comme dans l'exemple 80.

80. **X cause Y** regroupe les PRS tels que : **cause de X Y** (substantif) et **X cause Y** (verbe).

L'exemple 81 montre que les PRS peuvent également faire partie du même regroupement et ce, même si des mots grammaticaux (prépositions, conjonctions, etc.) varient d'un PRS à l'autre.

81. **On utilise X pour Y** regroupe les PRS tels que : **Y utile dans X** (adjectif suivi de la charnière *dans*) et **Y utile comme X** (adjectif suivi de la charnière *comme*).

Pour chaque PRS, nous indiquerons entre parenthèses le nombre d'occurrences si celui-ci est supérieur à un (Nx), comme dans l'exemple 82.

82. cause de X Y (3x).

De même, les patrons similaires sont regroupés et ne sont comptés qu'une seule fois dans les relevés de fréquence sur la répartition des PRS dans les corpus²⁷. Ainsi, les diverses formes que peut prendre un patron sont comptées comme s'il s'agissait d'un seul et même patron, bien que dans le tableau, on indique, à titre indicatif seulement, le nombre d'occurrences pour chacun d'eux. Dans les colonnes réservées à un corpus (médecine ou informatique), le nombre entre parenthèses indique le nombre d'occurrences d'un patron dans le corpus. Dans la colonne réservée aux patrons communs aux deux corpus, les nombres entre parenthèses renvoient au nombre d'occurrences d'une forme donnée d'un patron selon le corpus : le premier nombre renvoie au corpus de médecine, et le second, au corpus d'informatique, comme dans l'exemple 83, où le patron **X permet Y** est présent 9 fois dans le corpus de médecine, et 43 fois dans le corpus d'informatique.

83. **X permet Y** (9x, 43x).

²⁷ Dans les statistiques sur les patrons de relations sémantiques les plus fréquents et les statistiques sur la répartition des PRS pour un TS donné, les patrons sont tous pris en compte dans les calculs, y compris les patrons similaires.

3.5 Regroupement des termes par type sémantique TS

Cette section porte sur le regroupement des termes par type sémantique (TS). Nous nous sommes inspirée, pour faire ce regroupement, de la typologie des noms de WordNet. Ainsi, nous aborderons la définition des grands génériques WordNet (3.5.1). Nous justifierons le choix de ces grands génériques (3.5.2). Enfin, nous traiterons du classement des termes (3.5.3).

3.5.1 Grands génériques de WordNet

Afin de nous aider à classer nos termes en fonction des TS auxquels ils appartiennent, nous nous sommes inspirée, dans un premier temps, du dictionnaire électronique WordNet et de ses onze grands génériques (*entity, abstraction, psychological feature, natural phenomenon, activity, event, group, location, possession, shape, state*) auxquels tous les noms de ce dictionnaire sont rattachés.

Un grand générique constitue le dernier maillon de la chaîne d'hyperonymes pour une unité lexicale.

Bien qu'il s'agisse d'un dictionnaire général, le choix de WordNet s'est imposé à nous pour les raisons suivantes. D'abord, il renferme environ 200 000 entrées²⁸. De plus, sa typologie des grands génériques a servi de canevas de base dans la version multilingue de ce dictionnaire, à savoir EuroWordNet. Cette nouvelle typologie sert en effet d'index interlangue (Vossen 1998).

De plus, un certain nombre de ressources spécialisées sont mal adaptées à l'utilisation que nous voulons en faire ici. Par exemple, nous aurions pu retenir l'arbre de domaines utilisé par l'Office québécois de la langue française, mais ce dernier ne nous aurait pas permis de trouver les TS auxquels sont rattachés les

²⁸ Selon WordNet (Page consultée en juin 2004). En ligne.

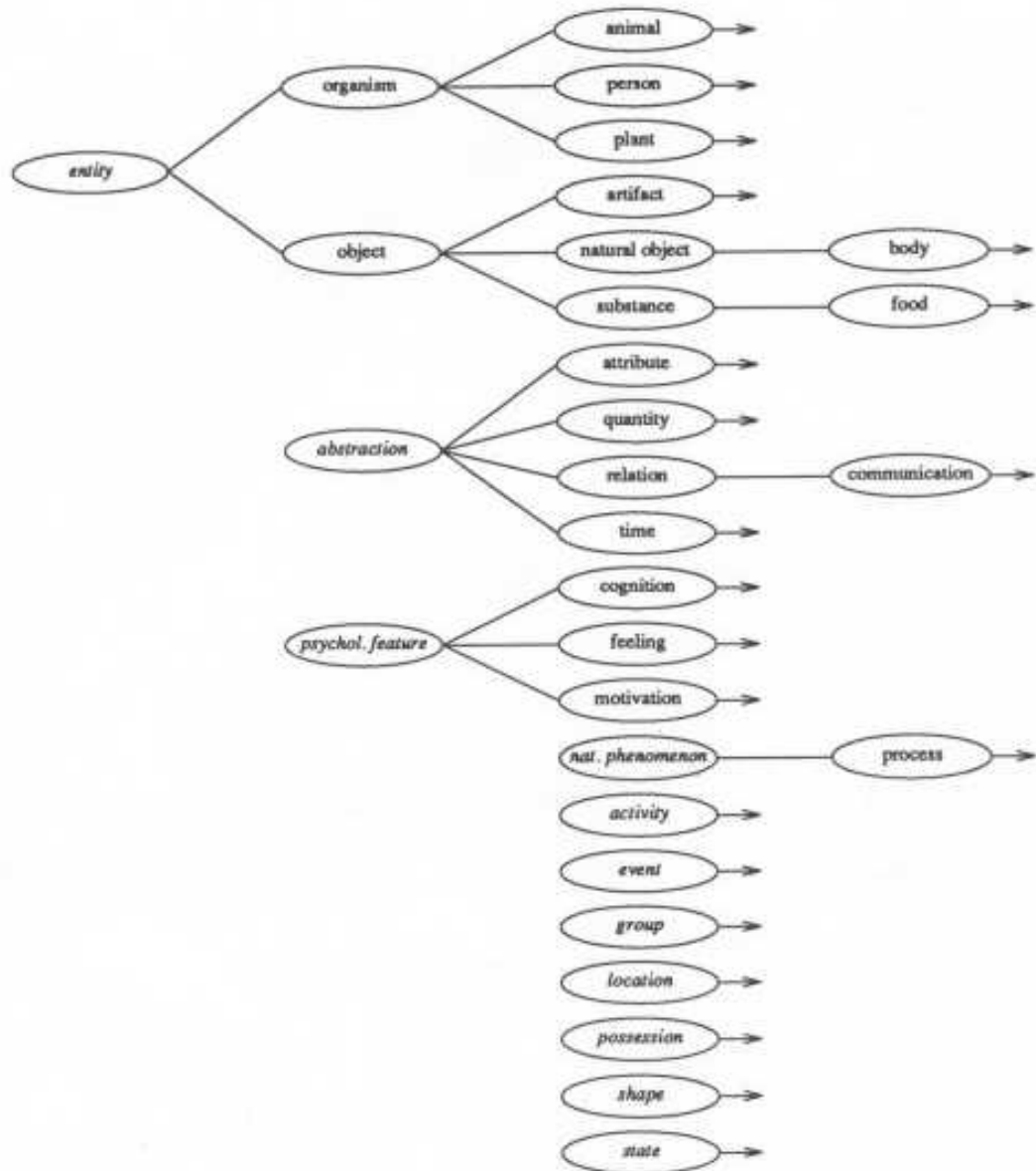
<http://www.cogsci.princeton.edu/~wn/man/wnstats.7WN.html>

termes de la hiérarchie, car les langues de spécialité sont toutes indépendantes les unes des autres : elles ne sont pas rattachées entre elles par des grands génériques.

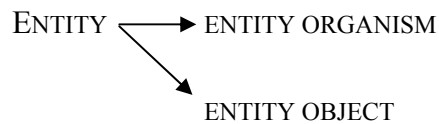
Quant au *UMLS (Unified Medical Language System) Semantic Network*, il constitue un réseau sémantique portant sur le domaine médical qui est en constante évolution. Il a été mis sur pied par la National Library of Medicine (États-Unis) et contient 134 TS reliés par 54 liens de nature différente. Ce réseau d'une très grande envergure aurait pu être utilisé pour la partie de notre recherche qui porte sur la médecine, mais n'aurait pu être étendue à l'informatique. De même pour la hiérarchie élaborée par Marie-Claude L'Homme de l'Université de Montréal (L'Homme 1995, 1996), également spécialisée, c'est-à-dire qu'elle porte sur une langue de spécialité en particulier, l'informatique en l'occurrence.

Les onze grands génériques de WordNet (en italique) sont représentés sous forme de hiérarchie à la figure 7.

Figure 7 : Grands génériques (Unique Beginners) de WordNet (Miller 1998 : 30)



Notre premier objectif consistait à sélectionner des termes appartenant à chacun de ces onze grands génériques pour chacun des corpus analysés (deux termes par grand générique et par corpus). Toutefois, nous nous sommes heurtée à de nombreuses difficultés. Certains types sémantiques correspondaient à des termes peu fréquents dans un corpus; certains types sémantiques n'avaient de correspondance que dans un seul corpus. Nous avons donc retenu neuf types sémantiques, qui sont les suivants²⁹ :



ABSTRACTION

PSYCHOLOGICAL FEATURE → PSYCHOLOGICAL FEATURE (FIELD)³⁰

PHENOMENON

ACTIVITY

EVENT

STATE

Pour chacun des types sémantiques suivants : [ACTIVITY], [ENTITY OBJECT], [ENTITY ORGANISM], [EVENT], [PHENOMENON], [PSYCHOLOGICAL FEATURE (FIELD)] et [STATE], nous avons retenu deux termes provenant de notre corpus d'informatique et deux termes provenant de notre corpus de médecine. Quant au TS [ABSTRACTION], nous avons seulement retenu un terme de médecine et deux termes d'informatique, alors que pour le TS [PSYCHOLOGICAL FEATURE],

²⁹ Les définitions de ces TS se trouvent à l'annexe C.

³⁰ Nous avons raffiné l'étiquette [PSYCHOLOGICAL FEATURE], que nous avons appelée [PSYCHOLOGICAL FEATURE (FIELD)], et qui englobe seulement des termes relatifs à un domaine du savoir, en l'occurrence *médecine, chirurgie, informatique et génie*.

deux termes de médecine ont été retenus, mais il n'a pas été possible de trouver des correspondances intéressantes dans le corpus d'informatique.

3.5.2 *Classement des termes*

Nous avons finalement retenu 17 termes en médecine et 16 termes en informatique, pour un total de 33 termes. Ci-dessous, nous avons dressé la liste des termes retenus (accompagnés de leur TS), que nous avons classés par domaine.

MÉDECINE

- 1) *angiographie* [ACTIVITY];
- 2) *cathéter* [ENTITY (OBJECT)];
- 3) *chirurgie* [PSYCHOLOGICAL FEATURE (FIELD)];
- 4) *décès* [EVENT];
- 5) *développement* [PHENOMENON];
- 6) *douleur* [PSYCHOLOGICAL FEATURE];
- 7) *état* [STATE];
- 8) *électrocardiogramme (ECG)* [ABSTRACTION];
- 9) *maladie* [STATE];
- 10) *médecin* [ENTITY (ORGANISM)];
- 11) *médecine* [PSYCHOLOGICAL FEATURE (FIELD)];
- 12) *patient* [ENTITY (ORGANISM)];
- 13) *scanner* [ENTITY (OBJECT)];
- 14) *symptôme* [PSYCHOLOGICAL FEATURE];
- 15) *systole* [EVENT];
- 16) *traitement* [ACTIVITY];
- 17) *vieillesse* [PHENOMENON].

INFORMATIQUE

- 1) *bug (bogue)* [STATE];
- 2) *codage* [ACTIVITY];
- 3) *développement* [PHENOMENON];
- 4) *erreur* [EVENT];
- 5) *état* [STATE];
- 6) *évolution* [PHENOMENON];
- 7) *génie* [PSYCHOLOGICAL FEATURE (FIELD)];
- 8) *imprimante* [ENTITY (OBJECT)];
- 9) *informatique* [PSYCHOLOGICAL FEATURE (FIELD)];
- 10) *langage* [ABSTRACTION];
- 11) *logiciel* [ABSTRACTION];
- 12) *ordinateur* [ENTITY (OBJECT)];
- 13) *panne* [EVENT];
- 14) *programmation* [ACTIVITY];
- 15) *programmeur* [ENTITY (ORGANISM)];
- 16) *utilisateur* [ENTITY (ORGANISM)].

3.6 Description de la base de données

Dans cette section, nous expliquons la façon dont les contextes sont traités dans la base de données.

Les informations sont d'abord classées de la manière suivante : terme, PRS et information se rapportant au terme. À ces informations, nous ajoutons le TS auquel appartient le terme, ainsi que la relation sémantique exprimée par un patron, soit hyperonymie, méronymie, causalité, finalité ou patron métalinguistique.

Nous présentons ci-dessous la description de chacun des champs du Tableau 17 :

Domaine : Ce champ indique le corpus (médecine ou informatique) d'où provient le contexte.

Terme : Ce champ indique l'un des termes clés choisis pour la requête.

TS du terme : Ce champ comprend les types sémantiques auxquels renvoient les termes clés dans WordNet, soit [ENTITY ORGANISM], [ENTITY OBJECT], [ABSTRACTION], [PSYCHOLOGICAL FEATURE], [PSYCHOLOGICAL FEATURE (FIELD)], [PHENOMENON], [ACTIVITY], [EVENT] et [STATE].

PRS : Ce champ indique le patron de relation sémantique neutralisé.

Relation sémantique exprimée par le PRS : Ce champ indique le lien sémantique qui lie le terme à l'information se rapportant à ce dernier (relation de finalité, relation causale, relation méronymique, relation hyperonymique ou patron métalinguistique).

Contexte : Ce champ contient le segment de texte où figurent le terme, le patron de relation sémantique et l'information qui se rapporte à ce terme.

Le Tableau 17 montre, par quelques exemples, les rubriques étudiées.

Tableau 17. Exemples d'entrées dans la base de données

DOMAINE	TERME	TS DU TERME	PRS	RELATION SÉMANTIQUE
Informatique	Logiciel	[ENTITY]	X destiné à Y	Finalité
... le logiciel NETPRINT destiné lui à l'impression sur matériel APPLE (impri).				
Informatique	Logiciel	[ENTITY]	Au moyen de X Y	Finalité
Autrement dit, dans 90 % des cas, la récupération d'un fichier mp3 doit se faire au moyen d'un logiciel FTP (formp3).				
Informatique	Langage	[ACTIVITY]	X permet Y	Finalité
Le langage HTML permet la programmation de pages qui pourront être vues par le Web (ceveil).				
Informatique	Logiciel	[ABSTRACTION]	Y contient X	Méronymique
Les réseaux sophistiqués centrés sur un micro « serveur » sont à la merci de la santé de ce « serveur » (il contient en outre tous les logiciels) (impri).				
Informatique	Logiciel	[ABSTRACTION]	Terme X fait référence à Y	Patron métalinguistique
Dans le jargon informatique, ce terme [logiciel], [...], fait référence à la construction d'une machine abstraite, composée d'instructions logiquement reliées et rédigées dans un langage compréhensible par la machine concrète qu'est l'ordinateur (begupa).				
Informatique	Logiciel	[ABSTRACTION]	On appelle X Y	Patron métalinguistique
Ces programmes [cet ensemble d'instructions constitue un programme dont la fonction est, en règle générale, spécifique : comptabilité, dessin, gestion de données] sont appelés logiciels (begupa).				
Médecine	Maladie	[STATE]	Y est X	Hyperonymique
Les leucodystrophies sont des maladies métaboliques [...] (image7).				

Informatique	Imprimante	[ENTITY OBJECT]	Y tel X	Hyperonymique
Or, on constate que le débit, c'est-à-dire le nombre d'informations élémentaires par seconde, des organes périphériques tels que les imprimantes et les lecteurs de cartes est beaucoup plus faible que celui de l'unité centrale (pierre1).				
Médecine	Maladie	[STATE]	Y consécutif à X	Causale
Ce [cataractes survenant dans le cadre d'une atteinte oculaire isolée] sont des cataractes consécutives à un traumatisme oculaire (contusion, plaie perforante), à une autre maladie (uvéite, myopie forte...), à l'action d'agents physiques (chaleur, radiations, electrocution), à l'administration d'agents toxiques locaux (corticoïdes) (opht-122).				
Médecine	Maladie	[STATE]	X est Y	Hyperonymique
La maladie de Lyme est une borréliose (patho3).				
Médecine	Maladie	[STATE]	Responsable de X Y	Causale
Le virus varicelle-zona (VZV) est responsable de la maladie [zona ophtalmique] réalisant une ganglioradiculite postérieure par atteinte du ganglion de Gasser (oph-mg).				
Médecine	Douleur	[PSYCHOLOGICAL FEATURE]	Type de X Y	Hyperonymique
... le type de la douleur (aiguë ou chronique) (doul).				
Médecine	Douleur	[PSYCHOLOGICAL FEATURE]	Y cause X	Causale
Il y a plusieurs petites glandes et conduits de glandes dans nos paupières. Elles peuvent s'infecter ou tout simplement se bloquer et enfler, causant de la douleur ou de l'inconfort (orgelet).				

3.7 Conclusion

Nous avons vu, dans ce chapitre, les démarches qui nous ont permis de mener à terme notre étude. Nous avons d'abord traité des corpus et décrit avec force détails leurs contenus. Nous en avons précisé les différentes caractéristiques dont la taille et les niveaux de spécialisation.

Par la suite, nous avons traité des termes retenus comme clés d'accès aux contextes. Nous avons vu quels étaient les critères d'acceptation des termes étudiés. Nous avons ensuite traité de l'extraction des contextes et des critères qui ont présidé à leur sélection.

Puis, nous avons étudié la question du traitement des PRS, du regroupement des termes par TS, et de la notion de grand générique (*unique beginners*) de WordNet.

Enfin, nous avons décrit le traitement des contextes dans notre base de données. Nous avons aussi expliqué l'association des PRS aux TS afin de dégager des généralisations sur les associations de patrons à des TS.

Dans le prochain chapitre, nous présenterons les résultats de la recherche.

4. RÉSULTATS

L'objet de ce chapitre est la présentation des résultats de notre recherche. D'abord, nous allons examiner quelques généralités sur l'analyse des contextes : le nombre de contextes retenus ainsi que le patron de relation sémantique le plus fréquent pour chaque relation étudiée, le nombre de contextes retenus par rapport au nombre de contextes étudiés pour chacun des termes ainsi que la répartition des relations sémantiques dans les corpus.

Étant donné que la sélection des contextes joue un rôle fondamental dans notre étude, nous rappellerons ici les explications concernant le choix des contextes. D'abord, pour être retenu, un contexte devait contenir le terme de la requête, ou, si c'était un terme complexe, la tête de ce terme devait être composée du terme clé. De plus, ce terme devait obligatoirement correspondre au concept décrit par WordNet et à la définition qui se trouve à l'annexe B. Les contextes devaient également contenir un patron lexical, puisque les patrons propres à un domaine n'étaient pas repérés. Quant à la relation qui liait le PRS au terme, elle devait être l'une des quatre relations sémantiques étudiées (causale, de finalité, méronymique ou hyperonymique). Enfin, nous avons également étudié les contextes qui contenaient un patron métalinguistique.

Par la suite, nous présentons les résultats obtenus, regroupés en fonction de chacun des types sémantiques auxquels appartiennent les termes sélectionnés : 4.2 [ENTITY OBJECT], 4.3 [EVENT], 4.4 [ACTIVITY], 4.5 [PSYCHOLOGICAL FEATURE (FIELD)], 4.6 [PSYCHOLOGICAL FEATURE], 4.7 [ENTITY ORGANISM], 4.8 [PHENOMENON], 4.9 [STATE] et 4.10 [ABSTRACTION].

Pour chacun de ces TS, nous donnons le détail du nombre de contextes et des termes retenus pour chacun des corpus. Ensuite, nous dressons un tableau présentant, par corpus, les résultats obtenus, c'est-à-dire les relations sémantiques accompagnées des PRS qui leur sont associés. Nous nous attardons ensuite sur les

relations sémantiques les plus représentées dans chacun des corpus. Enfin, nous étudions les PRS eux-mêmes et leur distribution par corpus.

La section 4.11 fera un récapitulatif des principaux résultats dégagés par notre analyse.

4.1 Remarques générales sur les contextes extraits

Parmi les quelque 8500 contextes contenant les termes choisis, 709 contenaient un patron exprimant des relations de finalité, de causalité, d'hyponymie ou de méronymie ou, encore, un patron métalinguistique. En informatique, nous avons retenu 468 contextes sur un total de 5765; huit pour cent des contextes analysés ont donc été retenus. Quant au corpus de médecine, nous avons retenu 241 contextes sur un total de 2653, c'est-à-dire que neuf pour cent des contextes dans lesquels apparaissent les termes clés ont été retenus.

Deux raisons peuvent expliquer la raison pour laquelle les contextes proviennent en plus grand nombre du corpus d'informatique. D'abord, notre corpus d'informatique est plus volumineux que notre corpus de médecine (693 243 mots en informatique comparativement à 558 349 mots en médecine). De plus, presque tous les textes de notre corpus de médecine sont des textes didactiques (ceux qui restent sont des textes de vulgarisation), alors que notre corpus d'informatique contient principalement des textes de vulgarisation. Il semble que ces derniers contiennent davantage de contextes intéressants que les textes didactiques. Ces observations sont contraires aux conclusions de Rebeyrolle (2000), mais rappelons que l'étude de Rebeyrolle ne portait que sur les énoncés définitoires qui correspondent à une partie des contextes que nous avons étudiés.

4.1.1 *Résultats obtenus pour chacune des relations*

Pour la relation hyperonymique, nous avons retenu, au total, 203 contextes, qui sont répartis comme suit : 152 contextes provenant du corpus d'informatique et 51 contextes provenant du corpus de médecine. Pour le corpus de médecine, c'est le patron **X est Y** qui est le plus important pour cette relation tous TS confondus, qui regroupe les formes suivantes : **X est Y (11x)**, **Y est X (14x)**, **Y c'est X**, **X qu'est Y**, **X c'est Y**. Ce patron est présent dans 55 % des contextes (28 patrons sur un total de 51). Pour le corpus d'informatique, le patron le plus important pour cette relation tous TS confondus est également **X est Y**, patron qui regroupe les formes suivantes : **X est Y (27x)**, **Y est X (19x)**, **Y qu'est X**, **Y c'est X (5x)**, **X c'est Y (3x)**. Ce patron représente 36 % des cas (55 patrons sur un total de 152).

En ce qui concerne la relation de finalité, nous avons retenu, au total, 227 contextes, qui sont répartis comme suit : 86 contextes provenant du corpus de médecine, et 141 contextes provenant du corpus d'informatique. Il s'agit de la relation sémantique pour laquelle nous avons obtenu le plus grand nombre de contextes. Pour le corpus de médecine, le patron le plus important pour cette relation tous TS confondus est : **X permet Y**, qui regroupe les formes suivantes : **X permet Y (24x)** et **X ne permet pas Y**. Ce patron est présent dans 29 % des contextes (25 patrons sur un total de 86). Pour le corpus d'informatique, le patron le plus important pour cette relation tous TS confondus est : **X permet Y**, qui regroupe les formes suivantes : **X permet Y (50x)**, **Y permet X (3x)**, **Y ne permet pas X** et **favoriser X permet Y**. Ce patron représente 39 % des cas (55 patrons sur un total de 141).

Pour la relation causale, nous avons retenu, au total, 143 contextes, qui sont répartis comme suit : 93 contextes provenant du corpus de médecine, et 50 contextes provenant du corpus d'informatique. Pour le corpus de médecine, le patron le plus important pour cette relation tous TS confondus est : **X cause Y**, qui regroupe les formes suivantes : **cause de X Y (5x)**, **Y cause X (3x)**, **Y ne**

cause pas X (2x), cause de Y X (5x). Ce patron représente 16 % des cas (15 patrons sur un total de 93). Pour le corpus d'informatique, le patron le plus important pour cette relation tous TS confondus est : **X entraîne Y**, qui regroupe les formes suivantes : **X entraîne Y (3x), Y entraîne X** et **Y entraîne conséquence sur X**. Ces patrons arrivent ex-aequo avec un résultat de 10 % (5 patrons sur un total de 50).

C'est la relation méronymique qui est la moins souvent exprimée dans les contextes étudiés. Nous avons retenu, au total, 91 contextes, qui sont répartis comme suit : 4 contextes provenant du corpus de médecine, et 87 contextes provenant du corpus d'informatique. Dans le corpus de médecine, le patron le plus important pour cette relation tous TS confondus est : **X comprend Y**, qui regroupe les formes suivantes : **X comprend Y, Y comprend X**. Ce patron est présent dans 2 des 4 contextes. Pour le corpus d'informatique, le patron le plus important pour cette relation tous TS confondus est : **X comporte Y**, qui regroupe les formes suivantes : **X comporte Y (6x) et Y comporte X (5x)**. Ce patron représente 13 % des cas (11 patrons sur un total de 87).

Quant aux patrons métalinguistiques, nous avons retenu, au total 45 contextes, qui sont répartis comme suit : 7 contextes provenant du corpus de médecine, et 38 contextes provenant du corpus d'informatique. Pour le corpus de médecine, le patron le plus important pour cette relation tous TS confondus est : **X ou Y**, qui regroupe les formes suivantes : **X ou Y, Y ou X (2x)**. Ce patron apparaît dans 3 des 7 contextes. Pour le corpus d'informatique, le patron le plus important pour cette relation tous TS confondus est : **on appelle X Y**, qui regroupe les formes suivantes : **on appelle X Y (4x), on appelle Y X, appellation de X Y, X appelé Y (6x)**. Ce patron représente 32 % des cas (12 patrons sur un total de 38).

On constate un déséquilibre entre les corpus en ce qui concerne les relations sémantiques. En effet, les contextes provenant du corpus d'informatique sont beaucoup plus nombreux que ceux provenant du corpus de médecine pour

quatre des cinq relations. Une seule exception toutefois concerne la relation causale, où près de deux fois plus de contextes sont tirés du corpus de médecine.

De plus, l'on constate que pour deux des cinq relations, soit celles qui expriment l'hyponymie et la finalité, le patron le plus important est le même dans nos deux corpus.

4.1.2 *Proportion des contextes retenus par rapport au nombre de contextes dépouillés pour chacun des termes clés*

Comme nous l'avons mentionné au début de ce chapitre, le nombre de contextes contenant à la fois un terme clé et l'expression d'une relation sémantique est nettement plus réduit que le nombre de contextes contenant le terme uniquement.

Le Tableau 18 présente, pour chacun des termes clés, la proportion de contextes retenus en fonction du nombre de contextes dépouillés. Dans la première colonne, nous indiquons le terme, suivi, dans la deuxième colonne, du corpus d'où le terme est tiré, et dans la troisième colonne, du type sémantique auquel appartient le terme. La quatrième colonne indique, quant à elle, le nombre de contextes retenus, suivi, dans l'avant-dernière colonne, du nombre de contextes dépouillés. Enfin, la dernière colonne indique le pourcentage de contextes retenus par rapport au nombre de contextes dépouillés.

Tableau 18. Nombre de contextes retenus pour chacun des termes clés/Nombre de contextes dépouillés

TERME	CORPUS	TYPE SÉMANTIQUE	NOMBRE DE CONTEXTES RETENUS	NOMBRE DE CONTEXTES DÉPOUILLÉS	%
Informatique	Informatique	[PSYCHOLOGICAL FEATURE (FIELD)]	3	407	1 %
Utilisateur	Informatique	[ENTITY ORGANISM]	7	923	1 %
Chirurgie	Médecine	[PSYCHOLOGICAL FEATURE (FIELD)]	2	137	1 %
Patient	Médecine	[ENTITY ORGANISM]	2	266	1 %
Scanner	Médecine	[ENTITY OBJECT]	6	246	2 %
État	Informatique	[STATE]	4	121	3 %
Programmeur	Informatique	[ENTITY ORGANISM]	4	108	4 %
Erreur	Informatique	[EVENT]	12	283	4 %
Systole	Médecine	[EVENT]	4	59	7 %
Médecine	Médecine	[PSYCHOLOGICAL FEATURE (FIELD)]	2	30	7 %
Ordinateur	Informatique	[ENTITY OBJECT]	81	1118	7 %
Traitement	Médecine	[ACTIVITY]	53	807	7 %
Programmation	Informatique	[ACTIVITY]	14	214	7 %
Développement	Médecine	[PHENOMENON]	11	143	8 %
Développement	Informatique	[PHENOMENON]	23	273	8 %
Imprimante	Informatique	[ENTITY OBJECT]	55	582	9 %
Codage	Informatique	[ACTIVITY]	6	60	10 %
Électrocardiogramme (ECG)	Médecine	[ABSTRACTION]	17	148	11 %
Évolution	Informatique	[PHENOMENON]	14	123	11 %
Médecin	Médecine	[ENTITY ORGANISM]	3	23	13 %
État	Médecine	[STATE]	13	104	13 %
Angiographie	Médecine	[ACTIVITY]	14	99	14 %
Cathéter	Médecine	[ENTITY OBJECT]	4	28	14 %
Logiciel	Informatique	[ABSTRACTION]	133	985	14 %
Panne	Informatique	[EVENT]	7	48	15 %
Maladie	Médecine	[STATE]	44	270	16 %
Douleur	Médecine	[PSYCHOLOGICAL FEATURE]	34	194	18 %

TERME	CORPUS	TYPE SÉMANTIQUE	NOMBRE DE CONTEXTES RETENUS	NOMBRE DE CONTEXTES DÉPOUILLÉS	%
Langage	Informatique	[ABSTRACTION]	85	441	19 %
Bogue (bug)	Informatique	[STATE]	11	49	22 %
Génie	Informatique	[PSYCHOLOGICAL FEATURE (FIELD)]	9	30	30 %
Symptôme	Médecine	[PSYCHOLOGICAL FEATURE]	22	72	31 %
Décès	Médecine	[EVENT]	7	20	35 %
Vieillessement	Médecine	[PHENOMENON]	3	7	3/7
Total			709	8418	8 %

Comme le montre le Tableau 18, 8 % des contextes étudiés en moyenne ont été retenus pour chacun des termes étudiés. Le TS [PSYCHOLOGICAL FEATURE] est celui dont la proportion de contextes retenus par rapport aux contextes étudiés est la plus forte, car pour les deux termes appartenant à ce type sémantique, soit *douleur* et *symptôme*, ce sont respectivement 18 % et 31 % des contextes qui ont été retenus. Ce TS est suivi de près par le TS [ABSTRACTION] puisque pour les trois termes appartenant à ce TS, soit *langage*, *logiciel* et *électrocardiogramme (ECG)*, on enregistre respectivement 19 %, 14 % et 11 % de contextes retenus. Enfin, pour trois des quatre termes appartenant au TS [STATE], soit *bogue*, *maladie* et *état* (médecine), on trouve respectivement 22 %, 16 % et 13 % de contextes retenus.

Cependant, on constate que certains types sémantiques sont peu souvent jumelés à une relation sémantique. C'est le cas du TS [ENTITY ORGANISM] qui est rarement associé à des PRS. En effet, en moyenne, pour l'ensemble de ce TS, seulement un contexte sur vingt a été retenu. Pour les termes *utilisateur*, *patient* et *programmeur*, ce sont respectivement 1 %, 1 % et 4 % des contextes qui ont été retenus. Le terme *médecin* se trouve, quant à lui, autour de la moyenne avec 13%.

On remarque également que pour trois des quatre termes appartenant au TS [PSYCHOLOGICAL FEATURE (FIELD)], une faible proportion de contextes a été retenue. En effet, les termes *informatique*, *chirurgie* et *médecine* n'apparaissent que dans 1 %, 1 % et 7 % respectivement de contextes porteurs d'une relation sémantique. Le terme *génie* fait exception puisque 30 % des contextes étudiés ont été retenus.

La proportion des contextes retenus pour le TS [PHENOMENON] est, quant à elle, moyennement élevée pour trois des quatre termes de ce TS. En effet, pour les termes *développement* (médecine) et *développement* (informatique) et *évolution*, ce sont 8 %, 8 % et 11 % des contextes qui ont été retenus. Pour le terme *vieillesse* cependant, la proportion de contextes intéressants grimpe à 3/7, mais le nombre de contextes étudiés pour ce terme est peu élevé (7 contextes).

Pour le TS [EVENT], on remarque un déséquilibre entre les termes qu'il regroupe. En effet, pour deux des quatre termes appartenant à ce type sémantique, 4 % et 7 % des contextes ont été retenus respectivement pour les termes *erreur* et *systole*. Cependant, pour le terme *panne*, 15 % des contextes ont été retenus, alors que pour le terme *décès*, 35 % des contextes se sont révélés intéressants pour notre étude.

En résumé, ces chiffres montrent que la proportion de contextes retenus par rapport au nombre de contextes dépouillés est, en moyenne, de 8 %. Seuls trois TS, soit [PSYCHOLOGICAL FEATURE], [ABSTRACTION] et [STATE], sont représentés par des termes dont la proportion de contextes retenus est plus élevée (moyenne de 18 % pour ces trois types sémantiques, soit 25 % pour [PSYCHOLOGICAL FEATURE], 15 % pour [ABSTRACTION] et 14 % pour [STATE]). Cependant, deux TS, soit [ENTITY ORGANISM] et [PSYCHOLOGICAL FEATURE (FIELD)] sont surtout représentés par des termes dont la proportion de contextes retenus est peu élevée. Ces chiffres tendent à montrer que certains types sémantiques ou certains termes sont plus susceptibles d'être associés à l'expression d'une relation sémantique dans les corpus. Toutefois, il faudrait mener une étude sur un plus grand nombre de termes pour pouvoir le conclure de façon certaine.

4.1.3 Répartition des relations sémantiques dans les corpus

Nous allons maintenant examiner la répartition des relations sémantiques étudiées – hyperonymie, méronymie, causale, finalité – et celle des patrons métalinguistiques dans nos corpus. Le Tableau 19 présente, dans les colonnes du centre, la proportion des contextes pour chacune des relations sémantiques par rapport au total des contextes retenus pour chacun des corpus. Les chiffres indiquent le nombre de contextes retenus pour une relation sémantique donnée par rapport au nombre total de contextes retenus pour chacun des corpus et, enfin, le pourcentage auquel cette proportion correspond.

Tableau 19. Répartition des relations sémantiques

Relation sémantique	Corpus d'informatique	Corpus de médecine	Nombre total et pourcentage
Causale	50/468 contextes = 11 %	93/241 contextes = 39 %	143 contextes = 20 %
Patron métalinguistique	38/468 contextes = 8 %	7/241 contextes = 3 %	45 contextes = 6 %
Hyperonymique	152/468 contextes = 32 %	51/241 contextes = 21 %	203 contextes = 29 %
Méronymique	87/468 contextes = 19 %	4/241 contextes = 2 %	91 contextes = 13 %
Finalité	141/468 contextes = 30 %	86/241 contextes = 36 %	227 contextes = 32 %
Total	468 contextes = 100 %	241 contextes = 100 %	709 contextes = 100 %

Comme le montre le Tableau 19, la relation la plus importante dans le corpus d'informatique est la relation hyperonymique (32 %), qui est suivie de très près par la relation de finalité (30 %). Dans le corpus de médecine, les relations les plus importantes sont la relation causale (39 %) et la relation de finalité (36 %). À l'opposé, pour les patrons métalinguistiques, seulement 5 % d'écart sépare les deux corpus. En effet, le corpus d'informatique contient 8 % de patrons métalinguistiques, contre 3 % pour le corpus de médecine. Il s'agit, dans les deux corpus, de patrons peu nombreux.

L'écart le plus important constaté entre les corpus est celui des relations causales puisque le corpus d'informatique contient 11 % de relations causales contre 39 % pour le corpus de médecine (ce qui représente un écart de 28 %). L'écart entre les deux corpus est également important en ce qui concerne la relation méronymique puisque 17 % les sépare. Cette dernière relation est beaucoup plus importante dans le corpus d'informatique que dans le corpus de

médecine. On constate également des écarts sensibles pour d'autres relations sémantiques : 11 % pour la relation hyperonymique et 6 % pour la relation de finalité.

Dans le Tableau 19, la colonne de droite, qui indique le nombre total de chacune des relations étudiées ainsi que le pourcentage de chacune d'elles pour l'ensemble des corpus, montre que la relation de finalité est la plus représentée (32 % des contextes). Cette relation est suivie par les relations hyperonymique (29 %), causale (20 %) et méronymique (13 %). Les patrons métalinguistiques, quant à eux, sont ceux qui sont le moins représentés, puisqu'ils ne sont présents que dans 6 % des contextes.

4.1.4 *Répartition des relations sémantiques dans les corpus pour lesquelles deux termes par corpus ont été étudiés*

Rappelons que, pour une partie des TS, nous n'avons pu étudier un nombre équivalent des termes dans chacun des corpus. Ces TS sont [ABSTRACTION] et [PSYCHOLOGICAL FEATURE].

Le Tableau 20 présente la répartition des relations sémantiques par corpus uniquement des relations pour lesquelles deux termes par corpus ont été étudiés. Il présente aussi leur fréquence indépendamment des corpus.

Tableau 20. Répartition des relations sémantiques pour lesquelles deux termes par corpus ont été étudiés

Relation sémantique	Corpus d'informatique	Corpus de médecine	Nombre total et pourcentage
Causale	49 contextes/20 %	62 contextes/37 %	111 contextes = 27 %
Patron métalinguistique	17 contextes/7 %	6 contextes/4 %	23 contextes = 6 %
Hyperonymique	59 contextes /24 %	26 contextes/15 %	85 contextes = 20 %
Méronymique	55 contextes/22 %	3 contextes/2 %	58 contextes = 14 %
Finalité	70 contextes/28 %	71 contextes/42 %	141 contextes = 34 %
Total	250 contextes = 100 %	168 contextes = 100 %	418 contextes = 100 %

Comme le montre le Tableau 20, pour le corpus d'informatique et le corpus de médecine, c'est toujours la relation de finalité qui est la plus importante, même lorsque les résultats sont analysés pour les TS pour lesquels nous avons étudié deux termes par corpus. L'écart le plus important constaté entre les corpus est celui des relations méronymiques, avec 20 % d'écart entre les corpus. Suivent de près les relations causales. En effet, 17 % d'écart sépare les deux corpus, puisque le corpus d'informatique contient seulement 20 % de contextes exprimant une relation causale contre 37 % pour le corpus de médecine. À l'opposé, pour les patrons métalinguistiques, seulement 3 % d'écart sépare les deux corpus. En effet, le corpus d'informatique contient 7 % de patrons métalinguistiques, contre 4 % pour le corpus de médecine. Quant aux autres relations, il y a 14 % d'écart entre les corpus pour la relation de finalité et 9 % d'écart pour la relation hyperonymique.

Le Tableau 20 montre également, dans la colonne de droite, que la relation la plus importante tous corpus confondus est la relation de finalité, représentée dans 34 % des contextes. Cette relation est suivie par les relations causale, hyperonymique et méronymique, relations représentées dans 27 %, 20 % et 14 % des contextes respectivement. Quant aux patrons métalinguistiques, ils ne sont représentés, quant à eux, que dans seulement 6 % des contextes.

Ces chiffres sont très semblables à ceux cités dans le Tableau 19, qui montre la fréquence des relations pour toutes les relations (y compris celles pour lesquelles deux termes par corpus n'ont pas été sélectionnés). La seule différence significative concerne les relations hyperonymique et causale, où, dans le Tableau 19, elles arrivent en deuxième (29 %) et troisième (20 %) positions respectivement. Dans le Tableau 20, c'est l'inverse qui se produit. En effet, c'est la relation causale qui est la deuxième relation la plus importante, présente dans 27 % des contextes, alors que la relation hyperonymique est la troisième relation la plus importante, puisqu'elle apparaît dans 20 % des contextes.

L'objectif de la section 4.1 était de dégager quelques généralités sur l'analyse des contextes. Dans les sections 4.2 à 4.10, nous présentons les résultats obtenus, regroupés en fonction de chacun des types sémantiques.

4.2 TS : [ENTITY OBJECT]

Pour ce TS, nous avons retenu 146 contextes au total. En médecine, nous avons sélectionné les termes *cathéter* et *scanner*, et en informatique, les termes *imprimante* et *ordinateur*. Le Tableau 21 présente, par domaine, les résultats obtenus pour le TS [ENTITY OBJECT] : les relations sémantiques accompagnées des PRS³¹ qui leur sont associés ainsi que le nombre de contextes relevés. Ces données sont présentées séparément pour chacun des corpus.

³¹ Pour chacun des tableaux de répartition des PRS du chapitre 4, le PRS le plus fréquent pour un type sémantique donné est souligné.

Tableau 21. Répartition des PRS pour le TS [ENTITY OBJECT]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
FINALITÉ	43	9	X permet Y (4x) ³² , on utilise X pour Y, rôle de X Y, au moyen de X Y, par l'intermédiaire de X Y, X peut Y	Médecine
		34	(pour) Y il faut X, X sert à Y (5x), X offre Y (2x), rôle de X Y (2x), on utilise X pour Y (5x), on utilise X à des fins Y, on confie Y à X, X n'est pas conçu pour Y, au moyen de X Y, X prend en charge Y, intérêt de X Y, X sait faire Y (2x), X pour Y, X peut Y (3x), X permet Y (2x), X rend possible Y, grâce à X Y, X orienté vers Y, X inventé pour Y, X destiné à Y	Informatique
PATRON MÉTALINGUISTIQUE	14	14	X dit Y, X on dit Y, on appelle X Y (2x), X appelé Y (3x), X c'est-à-dire Y, on qualifie Y de X, définition de X Y (2x), on définit X comme Y, dénomination de X Y, X ou Y	Informatique

³² La fréquence d'une forme de PRS est indiquée entre parenthèses sous la forme Nx.

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
HYPERONYMIQUE	50	49	Type de X Y (10x), catégorie de X Y (3x), X comme Y (6x), Y comme X, Y tel X (4x), X notamment Y, X se subdivise en groupes Y, classe de X Y, classe de Y X, le plus courant Y X, principal Y X, <u>X est Y</u> (9x), <u>Y est X</u> (3x), <u>Y c'est X</u> , <u>Y qu'est X</u> , tous les X à l'exception de Y, Y il s'agit de X, entre autres Y X, d'autres Y X, X vu comme Y	Informatique
		1	Type de X Y	Médecine
MÉRONYMIQUE	39	39	à l'intérieur de X Y, X avec Y, X comporte Y, X compose Y, Y compose X (2x), X se compose de Y (2x), composant de X Y, X contient Y (2x), X dispose de Y (6x), X doté de Y, X équipé de Y, X fait partie de Y, X fonctionne à l'aide de Y, X intègre Y, X muni de Y (3x), X n'est pas muni de Y, X n'a pas de Y, X possède Y (5x), Y complète X, X constitué de Y (2x), Y constitué de X, Y dans X, Y inclut X, Y où se trouve X	Informatique

Le Tableau 22 présente la répartition dans les deux corpus des quatre relations sémantiques relevées pour ce type sémantique. Dans la colonne du centre, les nombres indiquent la proportion d'une relation sémantique donnée en fonction du nombre total de contextes retenus pour ce TS. La colonne de droite indique, quant à elle, les pourcentages correspondants.

Tableau 22. Répartition des contextes par relation sémantique par ordre décroissant pour le TS [ENTITY OBJECT]

Relation	Nombre/Total	Pourcentage
Relation hyperonymique	50/146	34 %
Relation de finalité	43/146	29 %
Relation méronymique	39/146	27 %
Patron métalinguistique	14/146	10 %

La relation la plus importante pour le TS [ENTITY OBJECT] est la relation d'hyperonymie, qui est présente dans 34 % des contextes observés. Quant à la relation causale, elle est complètement absente pour ce type sémantique. Les contextes 84 et 85 présentent des exemples de la relation hyperonymique tirés respectivement de nos corpus de médecine et d'informatique.

84. La version la plus accomplie de ce matériel comprend un ballonnet en latex ou en silicone, fixé sur un *cathéter à souplesse progressive et à double voie* de **type** « cathéter de Moret » (image9).
85. ... les *micro-ordinateurs*, **comme** le IBM-PC, qui sont de très petits ordinateurs, de taille et de coût faibles (pierre1);

Le PRS le plus fréquent³³, toutes relations confondues, est le patron hyperonymique **X est Y**, qui regroupe les formes suivantes : **X est Y (9x)**, **Y est X (3x)**, **Y c'est X**, **Y qu'est X**. Il est présent dans 10 % des contextes étudiés (14 occurrences sur 146) si on considère l'ensemble des contextes associés à ce TS, et dans 28 % des contextes (14 occurrences sur 50) exprimant une relation hyperonymique.

Le TS [ENTITY OBJECT] est également associé à un nombre important de relations de finalité, à savoir 29 %. Les exemples 86 et 87, tirés respectivement de nos corpus de médecine et d'informatique, illustrent cette relation.

- 86. Technique : Injection sélective dans l'artère pulmonaire **au moyen d'un cathéter** introduit par une veine du bras ou la veine fémorale (embolie).
- 87. Une *imprimante sert à* imprimer des documents suivant diverses qualités (antho).

Le TS [ENTITY OBJECT] est également associé à un certain nombre de relations méronymiques (27 %), qui sont présentes uniquement dans notre corpus d'informatique. Elles sont presque toutes de nature partie fonctionnelle – toutes, à l'exception d'une seule, qui exprime une relation lieu – zone. Voici des exemples tirés de notre corpus d'informatique. Nous avons reproduit un exemple de relation partie fonctionnelle – tout en 88, et un exemple de relation lieu – zone en 89.

- 88. Dans les pages qui précèdent, nous avons vu qu'un *ordinateur est doté d'un* organe de mémoire, appelé mémoire principale (zmspc1).
- 89. Prenons le cas d'une pièce **où se trouvent** deux à quatre micros et une *imprimante* partagée (impri).

Le TS [ENTITY OBJECT] est également associé à 10 % de patrons métalinguistiques. Nous avons reproduit un exemple en 90.

³³ Pour plus de détails, voir la section 3.4.2.1.

90. En 1981, l'avènement de l'*ordinateur personnel*, **c'est-à-dire** de l'ordinateur dédié à un utilisateur unique marque un point de rupture technologique (ecran).

4.2.1 Répartition des relations sémantiques par corpus

Puisque 136 contextes sur 146 (93 %) proviennent du corpus d'informatique, contre seulement 10 contextes (7 %) pour le corpus de médecine, il vaut la peine de nous attarder sur les résultats obtenus pour chacun des corpus et tenter de fournir une explication à ce déséquilibre. Les Tableaux 23 et 24 présentent, par corpus, la répartition des relations sémantiques observées par ordre décroissant.

Tableau 23. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ENTITY OBJECT]

Relation	Nombre/Total	Pourcentage
Relation de finalité	9/10	90 %
Relation hyperonymique	1/10	10 %
Relation méronymique	0/10	0 %
Patron métalinguistique	0/10	0 %

Tableau 24. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [ENTITY OBJECT]

Relation	Nombre/Total	Pourcentage
Relation hyperonymique	49/136	36 %
Relation méronymique	39/136	29 %
Relation de finalité	34/136	25 %
Patron métalinguistique	14/136	10 %

Comme le montrent les Tableaux 23 et 24, bien que la relation de finalité soit bien représentée dans les deux corpus, il n'en reste pas moins que l'écart constaté entre les corpus est très important. En effet, le corpus de médecine compte 90 % de relations de finalité contre 25 % pour le corpus d'informatique. De plus, les relations méronymiques comptent pour 29 % dans le corpus d'informatique, alors qu'on n'en trouve aucune dans le corpus de médecine. Quant aux relations hyperonymiques, elles comptent pour 36 % dans le corpus d'informatique, alors que le corpus de médecine n'en contient que 10 %. Cependant, l'absence de relations causales est commune aux deux corpus. Enfin, on constate que les relations sont mieux réparties dans le corpus d'informatique.

Le Tableau 25 reprend la répartition des relations sémantiques pour le TS [ENTITY OBJECT], cette fois-ci par terme clé.

Tableau 25. Répartition des relations sémantiques par terme clé pour le TS [ENTITY OBJECT]

	RELATION CAUSALE	PATRON MÉTALINGUISTIQUE	RELATION DE FINALITÉ	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Cathéter (médecine)	0	0	3	1	0
Scanner (médecine)	0	0	6	0	0
Imprimante (informatique)	0	4	12	28	11 (10 partie fonctionnelle – tout, 1 lieu – zone)
Ordinateur (informatique)	0	10	22	21	28 (28 partie fonctionnelle – tout)

Comme le montre le Tableau 25, les relations sémantiques sont plus variées et plus nombreuses pour les termes *imprimante* et *ordinateur* que pour les termes *scanner* et *cathéter*. En effet, seulement deux relations sont associées à ces termes (finalité et hyperonymie), alors que quatre relations sont liées aux termes d'informatique (PM, hyperonymie, méronymie et finalité). De plus, pour le terme *ordinateur* les 28 relations méronymiques qui lui sont associées sont toutes de nature partie fonctionnelle – tout. Quant au terme *imprimante*, 10 des 11 relations méronymiques sont également de nature partie fonctionnelle – tout.

4.2.2 Répartition des PRS pour le TS [ENTITY OBJECT]

Cette sous-section examine la forme des patrons relevés pour le TS [ENTITY OBJECT] dans chacun des corpus et celle de patrons présents dans les deux corpus. Le Tableau 26 montre la répartition des patrons selon les corpus, et met au jour ceux qui sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 26. Répartition des patrons par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Finalité	Par l'intermédiaire de X Y	(pour) Y il faut X, X offre Y (2x), on confie Y à X, X n'est pas conçu pour Y, X prend en charge Y, X pour Y, X peut Y (2x), rôle de X Y, X rend possible Y, grâce à X Y, X orienté vers Y, X inventé pour Y, X destiné à Y, intérêt de X Y, X sait faire Y, X sert à Y (5x)	X permet Y (4x, 2x) ³⁴ , on utilise X pour Y (1x, 4x) (on utilise X à des fins Y (0x, 1x)), rôle de X Y (1x, 1x), au moyen de X Y (1x, 1x), X peut Y (1x, 1x)
Patron métalinguistique	*	X dit Y (X on dit Y), X appelé Y (3x) (on appelle X Y (2x)), X c'est-à-dire Y, on qualifie Y de X, définition de X Y (2x) (on définit X comme Y), dénomination de X Y, X ou Y	*
Hyperonymique	*	Catégorie de X Y (3x), X comme Y (6x) (Y comme X), X notamment Y, X se subdivise en groupes Y, classe de X Y (classe de Y X), le plus courant Y X, principal Y X, <u>X est Y (9x) (Y est X (3x), Y qu'est X, Y c'est X)</u> , tous les X à l'exception de Y, Y il s'agit de X, entre autres Y X (d'autres Y X), X vu comme Y, Y tel X (4x)	type de X Y (1x, 9x)

³⁴ Les nombres entre parenthèses indiquent le nombre d'occurrences pour chacun des corpus : celui de gauche renvoie au corpus de médecine, alors que celui de droite renvoie au corpus d'informatique.

Relation	Médecine	Informatique	Médecine et Informatique
Méronymique	*	À l'intérieur de X Y, X avec Y, X comporte Y, X compose Y (Y compose X (2x) X se compose de Y (2x), composant de X Y), X contient Y (2x), X dispose de Y (6x), X doté de Y, X équipé de Y, X fait partie de Y, X fonctionne à l'aide de Y, X intègre Y, X muni de Y (3x) (X n'est pas muni de Y), X n'a pas de Y, X possède Y (5x), Y complète X, X constitué de Y (2x) (Y constitué de X), Y dans X, Y inclut X, Y où se trouve X	*

Comme le montre le Tableau 26Tableau 26, la majorité des PRS sont propres au domaine de l'informatique ou de la médecine. En effet, seulement 6 patrons (9 %) sont partagés par les deux corpus. La plupart des patrons présents dans les deux corpus sont des patrons de finalité. Par ailleurs, 57 patrons (91 %) sont propres soit au corpus d'informatique, soit au corpus de médecine.

4.2.3 Remarques générales sur le TS [ENTITY OBJECT]

En résumé, en ce qui concerne le TS [ENTITY OBJECT], la relation hyperonymique est la plus importante. En effet, elle est exprimée dans 34 % des contextes étudiés. La relation la plus importante diffère entre les corpus, puisque, dans le corpus de médecine, c'est la relation de finalité qui est la mieux représentée pour ce TS alors que, dans le corpus d'informatique, c'est la relation hyperonymique qui est la plus importante. Cependant, aucune relation causale n'a été répertoriée ni dans le corpus de médecine, ni dans le corpus d'informatique.

Le PRS le plus fréquent, toutes relations confondues, est le patron hyperonymique **X est Y**, qui regroupe les formes suivantes : **X est Y (9x)**, **Y est X (3x)**, **Y c'est X**, **Y qu'est X**. Il est présent dans 10 % des contextes étudiés (14 occurrences sur 146) pour l'ensemble des relations de ce TS, et dans 28 % des contextes (14 occurrences sur 50) exprimant une relation hyperonymique. En outre, il n'est présent que dans le corpus d'informatique.

Enfin, la majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, seulement 9 % des patrons sont partagés par les deux corpus, alors que 91 % des patrons sont exclusifs à l'informatique ou à la médecine.

4.3 TS : [EVENT]

Pour le TS [EVENT], nous avons retenu, au total, 30 contextes. En médecine, nous avons sélectionné les termes *décès* et *systole* et, en informatique,

les termes et *erreur* et *panne*. Le Tableau 27Tableau 27 présente, par domaine, les résultats obtenus pour le TS [EVENT] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 28 présente la répartition dans les deux corpus des quatre relations sémantiques observées pour ce type sémantique. La répartition est présentée par ordre décroissant.

Tableau 27. Répartition des relations sémantiques pour le TS [EVENT]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
CAUSALE	25	16	X cause Y, en cas de X Y (3x), en raison de X Y, origine de X Y, X dû à Y, X entraîne Y, X ne provient pas de Y, X provient de Y (3x), X résulte de Y, générateur de X Y, Y provoque X, source de X Y	Informatique
		9	Cause de X Y (3x), X déclenche Y, X lié à Y, responsable de X Y (2x), Y risque X (2x)	Médecine
PATRON MÉTALINGUISTIQUE	1	1	On appelle X Y	Médecine
MÉRONYMIQUE	1	1	Y contient X	Informatique
HYPERONYMIQUE	3	2	Type de X Y (2x)	Informatique
		1	Il s'agit de X Y	Médecine

Tableau 28. Répartition des relations sémantiques par ordre décroissant pour le TS [EVENT]

Relation	Nombre/Total	Pourcentage
Relation causale	25/30	83 %
Relation hyperonymique	3/30	10 %
Relation méronymique	1/30	3 %
Patron métalinguistique	1/30	3 %

D'abord, on constate que la relation de finalité est absente pour le TS [EVENT]. La relation la plus importante pour ce TS est la relation causale, qui est présente dans 83 % des contextes étudiés. Les contextes 91, 92, 93, 94, 95 et 96 montrent que les relations causales pour le TS [EVENT] sont variées. Ces contextes sont extraits de nos corpus de médecine et d'informatique.

Augmentation :

91. À vous de prendre en compte les délais éventuels de livraison et d'immobilisation **en cas de panne** (pcprem).

Création :

92. La barrette SIM est la première de la génération actuelle des barrettes mémoire. Elle était présentée sous forme de SIP (single inline package) dont les contacts étaient réalisés par des broches semblables à des circuits intégrés. Elles ont été abandonnées au profit de barrettes SIMM car leurs contacts étaient **sources de pannes** (antho).
93. ...l'existence de *telles systoles [systoles ectopiques ventriculaires]* **liées à l'insuffisance cardiaque** (xrenal1)...

Destruction :

94. Ces embolies systémiques ont un mauvais pronostic; **responsables de 15 à 30 % des décès** (retr_mit).
95. En effet, une *panne* de ventilateur **peut** directement **causer** la mort de votre processeur (optrep).

Diminution :

96. La copie se trouvant sur la disquette cible peut être incomplète, **en raison d'une erreur** irrémédiable d'entrée-sortie (zmspc2).

Les PRS les plus fréquents sont les patrons de relation causale **X cause Y** et **X provient de Y**. Le patron **X cause Y** regroupe les formes suivantes : **cause de X Y (3x)** et **X cause Y**. Quant au patron **X provient de Y**, il regroupe les

formes suivantes : **X ne provient pas de Y** et **X provient de Y** (3x). Ces patrons sont tous deux présents dans 13 % des contextes étudiés (4 occurrences sur 30) pour ce TS, et dans 16 % des contextes exprimant une relation causale (4 occurrences sur 25) .

Dans les patrons de relation causale relevés, le X (c'est-à-dire le terme de la requête) représente l'effet dans 72 % des contextes. Les patrons suivants sont des illustrations de ce cas de figure : **responsable de X Y**, **cause de X Y**, **X résulte de Y**, etc. Dans 28 % des cas, le X renvoie à la cause, comme dans les patrons suivants : **X cause Y**, **en raison de X Y**, etc.

La seule relation méronymique relevée (exemple 97) provient du corpus d'informatique. Nous la qualifions d'*inclassable*, étant donné qu'elle n'appartient à aucune des subdivisions proposées par Winston *et al.* (1987).

97. Malheureusement, les programmes écrits dans un tel langage **contenaient** souvent des *erreurs* qui n'étaient pas toujours faciles à corriger (pierre1).

4.3.1 *Répartition des relations sémantiques par corpus*

Dans cette section, nous comparons les résultats obtenus pour chacun des deux corpus. Les Tableaux 29 et 30 présentent, pour chacun des corpus, la répartition des quatre relations sémantiques par ordre décroissant.

Tableau 29. Répartition des relations sémantiques du corpus de médecine pour le TS [EVENT]

Relation	Nombre/Total	Pourcentage
Relation causale	9/11	82 %
Patron métalinguistique	1/11	9 %
Relation hyperonymique	1/11	9 %
Relation méronymique	0/11	0 %

Tableau 30. Répartition des relations sémantiques du corpus d'informatique pour le TS [EVENT]

Relation	Nombre/Total	Pourcentage
Relation causale	16/19	84 %
Relation hyperonymique	2/19	11 %
Relation méronymique	1/19	5 %
Patron métalinguistique	0/19	0 %

Les écarts entre les deux corpus sont moins marqués ici que pour le TS [ENTITY OBJECT] qui a fait l'objet de la section précédente. On remarquera que la relation causale est, dans les deux corpus, la relation la plus importante. En effet, le corpus de médecine en contient 82 %, et le corpus d'informatique en contient 84 %. Le corpus de médecine compte aussi une relation hyperonymique et un patron métalinguistique. Dans le corpus d'informatique, la relation causale est suivie par la relation hyperonymique (11 %) et la relation méronymique (5 %). La relation méronymique est absente du corpus de médecine pour ce TS, alors que la relation méronymique, bien que peu fréquente, est présente dans le corpus d'informatique (5 %). Cependant, le corpus d'informatique ne contient aucun patron métalinguistique. Ces légers écarts peuvent être attribuables aux caractéristiques distinctes des termes ou à la taille des corpus.

En ce qui concerne la relation causale, il y a une divergence entre les données obtenues pour le corpus d'informatique et celui de médecine. En effet, seulement 9 relations causales sur 25 (36 %) sont associées aux termes de médecine contre 16 relations sur 25 (64 %) pour le corpus d'informatique. Le corpus d'informatique contient donc presque deux fois plus de relations causales que le corpus de médecine pour ce TS.

De plus, des 15 relations causales de nature création, 13 proviennent du corpus d'informatique et 2 du corpus de médecine. Des 8 relations causales de nature destruction, 7 proviennent du corpus de médecine et une seule du corpus d'informatique. Quant aux relations causales de nature augmentation et diminution, elles proviennent du corpus d'informatique.

Le Tableau 31 reprend la répartition des relations sémantiques pour le TS [EVENT], cette fois-ci par terme clé.

Tableau 31. Répartition des relations sémantiques par terme clé pour le TS [EVENT]

	RELATION CAUSALE	PATRON MÉTALIN-GUISTIQUE	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Décès (médecine)	7 (7 destruction)	0	0	0
Systole (médecine)	2 (2 création)	1	1	0
Panne (informatique)	7 (5 création, 1 destruction, 1 augmentation)	0	0	0
Erreur (informatique)	9 (8 création, 1 diminution)	0	2	1 (inclassable)

Comme le montre le Tableau 31, c'est pour le terme d'informatique *erreur* que les relations sémantiques sont les plus variées et les plus nombreuses. En effet, pour ce terme, on dénombre 12 relations réparties entre les relations causale, hyperonymique et méronymique. Huit des neuf relations causales expriment une création. Le terme *panne*, quant à lui, n'est associé qu'à la relation causale, exprimant surtout la création (5 contextes sur 7). Le terme *décès* est également associé à une seule relation sémantique, soit la relation causale, exprimant exclusivement la destruction. Enfin, le terme *systole*, est le terme qui est le moins fréquent pour ce TS. Il est cependant associé à trois relations sémantiques (causale, hyperonymique et patron métalinguistique).

4.3.2 Répartition des PRS pour le TS [EVENT]

Cette sous-section examine la forme des patrons relevés pour le TS [EVENT] dans chacun des corpus et dans les deux corpus. Le Tableau 320 montre

la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 32. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Causale	X déclenche Y, X lié à Y, Y induit X, responsable de X Y (2x), Y risque X (2x)	En cas de X Y (3x), en raison de X Y, origine de X Y, X dû à Y, X entraîne Y, X provient de Y (3x) (X ne provient pas de Y), X résulte de Y, générateur de X Y, Y provoque X, source de X Y	<u>Cause de X Y (3x, 0x) (X cause Y (0x, 1x))</u>
Patron métalinguistique	On appelle X Y		*
Hyperonymique	Il s'agit de X Y		*
Méronymique	*	Type de X Y (2x), Y contient X	*

Comme le montre le Tableau 32, très peu de PRS sont communs aux deux corpus. En effet, un seul patron exprimant une relation causale est présent dans le corpus de médecine et d'informatique (rappelons qu'il s'agissait de la relation sémantique la plus importante pour ce TS). Par ailleurs, 19 patrons (95 %), sont propres soit à l'informatique, soit à la médecine.

4.3.3 Remarques générales sur le TS [EVENT]

En conclusion, en ce qui concerne le TS [EVENT], toutes les relations sémantiques sont représentées à l'exception de la relation de finalité qui n'a été répertoriée ni dans le corpus de médecine, ni dans le corpus d'informatique. La relation causale est la relation sémantique la plus importante. En effet, elle est exprimée dans 83 % des contextes étudiés.

Les PRS les plus fréquents sont les patrons de relation causale **X cause Y** et **X provient de Y**. Le patron **X cause Y** regroupe les formes suivantes : **cause de X Y (3x)** et **X cause Y**. Quant au patron **X provient de Y**, il regroupe les formes suivantes : **X provient de Y (3x)** et **X ne provient pas de Y**. Les patrons **X cause Y** et **X provient de Y** représentent chacun 16 % des relations causales, et 13 % de l'ensemble de ce TS. Mais cela ne représente que quatre contextes par patron; il est donc difficile de parler de patrons productifs dans ce cas.

La majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, seulement 5 % des patrons sont partagés par les deux corpus alors que 95 % des patrons sont exclusifs de l'informatique ou de la médecine.

4.4 TS : [ACTIVITY]

Pour le TS [ACTIVITY], nous avons retenu 87 contextes au total. En médecine, nous avons sélectionné les termes *angiographie* et *traitement*, et en informatique, les termes *codage* et *programmation*. Le Tableau 33 présente, par

domaine, les résultats obtenus pour le TS [ACTIVITY] : les relations sémantiques, le nombre total de contextes relevés pour chacun des corpus, ainsi que les PRS relevés.

Le 0 présente la répartition dans les deux corpus des 5 relations sémantiques par ordre décroissant.

Tableau 33. Répartition des PRS pour le TS [ACTIVITY]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
FINALITÉ	62	52	Objectif de X Y (2x), on réserve X à Y, on réserve Y à X (2x), on réserve X pour Y, rôle de X Y, Y suffit pour X, X suffit à Y, Y indiqué dans X (3x), X indiqué lorsque Y, X vise à Y (2x), <u>Y utile dans X (2x)</u> , <u>Y utile comme X</u> , on utilise Y dans X (9x), on utilise X à titre Y, on utilise X comme Y, indispensable pour X Y, on emploie Y pour X, on emploie Y dans X (2x), base de X Y, <u>X permet Y (11x)</u> , but de X Y (6x), X nécessaire pour Y	Médecine
		10	<u>X permet Y (2x)</u> , <u>Y permet X</u> , <u>Y ne permet pas X</u> , Y facilite X, X offre Y, X assuré à l'aide de Y, Y rend plus aisé X, rôle de X Y, objectif de X Y	Informatique
CAUSALE	8	3	X dépend de Y (3x)	Informatique
		5	On rencontre Y lors de X, complication de X Y, arrêt de X Y, arrêt de X entraîne Y, X provoque Y	Médecine

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
PATRON MÉTALINGUISTIQUE	1	1	On appelle X Y	Médecine
HYPERONYMIQUE	9	6	Exemple de X Y, Y représente X, X est Y (3x), Y il s'agit de X	Médecine
		3	Y est X, exemple de X Y, Y comme X	Informatique
MÉRONYMIQUE	7	4	Étape de X Y, Y inhérent à X, Y aspect du processus de X, phase de Y X	Informatique
		3	X comporte Y, X comprend Y, étape de X Y	Médecine

Tableau 34. Répartition des relations sémantiques par ordre décroissant pour le TS [ACTIVITY]

Relation	Nombre/Total	Pourcentage
Relation de finalité	62/87	71 %
Relation hyperonymique	9/87	10 %
Relation causale	8/87	9 %
Relation méronymique	7/87	8 %
Patron métalinguistique	1/87	1 %

Pour le TS [ACTIVITY], toutes les relations sémantiques sont présentes. La relation sémantique la plus importante est la relation de finalité, exprimée dans 71 % des contextes retenus. Les exemples 98 et 99, tirés respectivement de nos corpus de médecine et d'informatique, sont des contextes explicitant une relation de finalité.

98. La quinidine **est employée dans** le *traitement* des arythmies auriculaires (fibrillation et flutter) et ventriculaires (digitaline, infarctus) (zschmit).
99. ... la *programmation* **est assurée à l'aide de** bibliothèques de communication par échange de messages (cosnar).

Le patron le plus fréquent est le patron de finalité **X permet Y**. Il regroupe les formes suivantes : **X permet Y (13x)**, **Y permet X**, **Y ne permet pas X**. Ces patrons, tirés des deux corpus, sont présents dans 17 % des cas (15 sur 87) pour l'ensemble de ce TS, et dans 24 % (15 sur 62) pour la relation de finalité. Suit de très près un autre patron de finalité, soit le patron **on utilise X pour Y**, qui regroupe les formes suivantes : **Y utile dans X (2x)**, **Y utile comme X**, **on utilise Y dans X (9x)**, **on utilise X à titre Y**, **on utilise X comme Y**. N'apparaissant que dans le corpus de médecine, ce patron est présent dans 16 % des contextes étudiés (14 occurrences sur 87) pour l'ensemble de ce TS, et dans 23 % (14 occurrences sur 62) des contextes exprimant une relation de finalité.

Pour le TS [ACTIVITY], on dénombre également 8 % de relations méronymiques, toutes de nature phase – activité. Les exemples 100 et 101, tirés respectivement de nos corpus de médecine et d'informatique, illustrent cette relation méronymique.

100. Le *traitement* des accidents hémorragiques **comporte**, bien entendu, l'arrêt du traitement anticoagulant et un traitement antagoniste (ztouitou).
101. Le paramétrage est un **aspect** essentiel **du processus de programmation** (genilo).

Pour le TS [ACTIVITY], on dénombre aussi 10 % de relations hyperonymiques. Le contexte 102 en présente un exemple en médecine, alors que le contexte 103 en présente un exemple en informatique.

102. Le *traitement* de la cardiopathie sousjacent (par **exemple** le traitement anti-angineux)... (lechat).
103. En ce qui concerne ses **exemples** de *programmations* l'auteur a opté pour le BASIC STANDARD parce qu'il est connu de tous les programmeurs amateurs ou professionnels, et de fait plus facilement traduisible en d'autres langages (impri).

Seulement 9 % des relations de ce TS sont des relations causales. Dans le corpus de médecine, 4 contextes expriment la création et un contexte exprime l'augmentation, alors que les trois contextes provenant du corpus d'informatique expriment la modification. Les contextes 104 et 105 sont tirés de notre corpus de médecine et montrent respectivement un exemple de relation causale de nature création et un exemple de relation causale de nature augmentation. Quant au contexte 106, il montre un exemple de relation causale de nature modification, provenant du corpus d'informatique.

104. Le *traitement* chronique par l'amiodarone **provoque** parfois des neuropathies à type le plus souvent de polyradiculonévrite sensitive ou motrice (symptomatologie prédominante: sensitive) parfois associée à des tremblements, à un syndrome extrapyramidal ou à un syndrome cérébelleux (xmedpat).
105. L'**arrêt du traitement entraîne** habituellement une réascension des chiffres tensionnels (xgiroud);
106. Un type, lequel est une précision sur la nature des valeurs en question; dans la mesure où, rappelons-le, le *codage* d'une donnée **dépend de** sa nature, la zone mémoire référencée par l'adresse a une taille et un format propres à chaque sorte de donnée; le type de la variable permet de les calculer; la notion de type sera précisée ultérieurement (access1).

Parmi les 8 relations causales, on dénombre 4 cas où le X est l'effet et 4 cas où le X est la cause de Y. Les exemples 107 et 108 illustrent ces deux cas de figure.

X qui est l'effet :

107. Il va tout d'abord de soi que le *codage* d'informations **dépend de** la nature et du degré de complexité de celles-ci ainsi que de la manière dont elles sont organisées (accés1).

X qui est la cause :

108. Les troubles digestifs (nausées, vomissements, diarrhées) **sont** fréquemment **rencontrés lors des** *traitements* par les sels de quinidine; ils s'observent également mais, à un moindre degré, avec les produits de classe Ib et Ic (zdjiane).

Enfin, pour le TS [ACTIVITY], on dénombre un seul patron métalinguistique. Il provient du corpus de médecine (109).

109. Ce *traitement*, **appelé** photocoagulation, consiste à focaliser la lumière du laser sur les vaisseaux qui laissent passer le fluide dans la macula (ret_diab).

Le Tableau 35 reprend la répartition des relations sémantiques pour le TS [ACTIVITY], cette fois-ci par terme clé.

Tableau 35. Répartition des relations sémantiques par terme clé pour le TS [ACTIVITY]

	RELATION CAUSALE	PATRON MÉTALINGUISTIQUE	RELATION DE FINALITÉ	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Angiographie (médecine)	0	0	12	2	0
Traitement (médecine)	5 (4 création, 1 augmentation)	1	40	4	3 (3 phase-activité)
Programmation (informatique)	0	0	8	3	3 (3 phase – activité)
Codage (informatique)	3 (3 modification)	0	2	0	1 (1 phase – activité)

Comme le montre le Tableau 35, c'est le terme de médecine *traitement* qui est le plus fréquemment associé à une relation sémantique pour ce TS. On retrouve, en effet, 53 contextes, qui sont répartis dans les cinq différentes relations sémantiques. Quatre des cinq relations causales expriment de la création, et les 3 relations méronymiques sont de nature phase – activité. Le terme d'informatique *programmation*, quant à lui, apparaît dans 14 contextes répartis entre les relations de finalité, l'hyponymie et la méronymie. Les trois relations méronymiques sont également toutes de nature phase – activité. Le terme *angiographie* figure également dans 14 contextes, et il exprime soit une relation de finalité, soit une relation hyperonymique. Toutefois, contrairement aux trois autres termes, il n'est jamais associé à une relation méronymique. Enfin, le terme le moins fréquent pour ce TS est *codage*, terme qui est associé aux relations causale, de finalité et

méronymique. La relation méronymique exprimée est, encore une fois, de nature phase – activité.

4.4.1 *Répartition des relations sémantiques par corpus*

Dans cette section, nous comparons les résultats obtenus pour le corpus de médecine et ceux livrés par le corpus d'informatique. Les Tableaux 36 et 37 présentent, par corpus, la répartition des 5 relations sémantiques par ordre décroissant.

Tableau 36. Répartition des relations sémantiques du corpus de médecine pour le TS [ACTIVITY]

Relation	Nombre/Total	Pourcentage
Relation de finalité	52/67	78 %
Relation hyperonymique	6/67	9 %
Relation causale	5/67	7 %
Relation méronymique	3/67	4 %
Patron métalinguistique	1/67	1 %

Tableau 37. Répartition des relations sémantiques du corpus d'informatique pour le TS [ACTIVITY]

Relation	Nombre/Total	Pourcentage
Relation de finalité	10/20	50 %
Relation méronymique	4/20	20 %
Relation hyperonymique	3/20	15 %
Relation causale	3/20	15 %
Patron métalinguistique	0/20	0 %

Comme pour le TS [ENTITY OBJECT], le nombre de contextes retenus est beaucoup plus important pour un corpus que pour l'autre. Pour le TS [ACTIVITY] cependant, ce sont les contextes tirés du corpus de médecine qui sont les plus nombreux. En effet, 77 % des contextes proviennent du corpus de médecine, contre seulement 23 % pour celui d'informatique.

On remarquera également que la relation de finalité est, dans les deux corpus, la relation la plus fréquente avec, respectivement, 78 % pour le corpus de médecine, et 50 % pour le corpus d'informatique. Cette relation est suivie de loin par la relation hyperonymique en médecine (9 %), et la relation méronymique en informatique (20 %). Quant à la relation causale, elle ne représente que 7 % des relations pour le corpus de médecine et 15 % pour celui d'informatique.

Il est intéressant de noter que 4 des 5 relations causales du corpus de médecine sont de nature création. Quant aux relations causales du corpus d'informatique, elles expriment exclusivement la modification.

4.4.2 *Répartition des PRS pour le TS [ACTIVITY]*

Cette sous-section examine la forme des patrons relevés pour le TS [ACTIVITY] dans chacun des corpus et dans les deux corpus. Le Tableau 38 montre la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 38. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Finalité	On réserve X à Y (on réserve Y à X (2x), on réserve X pour Y), Y suffit pour X (X suffit à Y), Y indiqué dans X (3x) (X indiqué lorsque Y), X vise à Y (2x), <u>Y utile dans X (2x) (Y utile comme X, on utilise Y dans X (9x), on utilise X à titre Y, on utilise X comme Y), indispensable pour X Y, on emploie Y pour X (on emploie Y dans X (2x)), base de X Y, but de X Y (6x), X nécessaire pour Y</u>	Y facilite X, X offre Y, X assuré à l'aide de Y, Y rend plus aisé X	X permet Y (11x, 2x) (Y permet X (0x, 1x), Y ne permet pas X (0x, 1x)), rôle de X Y (1x, 1x), objectif de X Y (2x, 1x)
Causale	On rencontre Y lors de X, complication de X Y, arrêt de X Y (arrêt de X entraîne Y), X provoque Y	X dépend de Y (3x)	*
Patron métalinguistique	On appelle X Y	*	*
Hyperonymique	Y représente X, Y il s'agit de X	Y comme X	Exemple de X Y (1x, 3x), Y est X (0x, 1x), (X est Y (3x, 0x))
Méronymique	X comporte Y, X comprend Y	Y inhérent à X, Y aspect du processus de X, phase de Y X	Étape de X Y (1x, 1x)

Comme le montre le Tableau 38, la majorité des PRS sont exclusifs à un corpus. En effet, 6 patrons (18 %) sont partagés par les deux corpus. La plupart des patrons présents dans les deux corpus sont des patrons exprimant la finalité (rappelons qu'il s'agissait de la relation sémantique la plus importante pour ce TS). Par ailleurs, 28 patrons 82 % sont propres soit à l'informatique, soit à la médecine. Il s'agit de l'un des deux TS pour lequel le plus grand nombre de patrons communs a été relevé, car il arrive ex-æquo avec le TS [PHENOMENON] qui fait l'objet de la section 4.7.

4.4.3 *Remarques générales sur le TS [ACTIVITY]*

En conclusion, en ce qui concerne TS [ACTIVITY], la relation de finalité est la plus importante. En effet, elle est exprimée dans 71 % des contextes étudiés. Bien que les données sur les relations diffèrent légèrement d'un corpus à l'autre, il n'en demeure pas moins que, pour les deux corpus, c'est la relation de finalité qui est la mieux représentée, surpassant de loin toutes les autres relations, surtout en ce qui a trait au corpus de médecine.

Le patron le plus fréquent est le patron de finalité **X permet Y**, qui regroupe les formes suivantes : **X permet Y (13x)**, **Y permet X**, **X ne permet pas Y**. Ce patron, apparaissant dans les deux corpus, est présent dans 17 % des contextes (15 sur 87) pour l'ensemble de ce TS, et dans 24 % (15 sur 62) pour la relation de finalité. Suit de très près un autre patron de finalité, soit **on utilise X pour Y**, qui regroupe les formes suivantes : **Y utile dans X (2x)**, **Y utile comme X**, **on utilise Y dans X (9x)**, **on utilise X à titre Y**, **on utilise X comme Y**. Ce patron, qui n'apparaît que dans le corpus de médecine est présent dans 16 % des contextes (14 occurrences sur 87) pour l'ensemble de ce TS, et dans 23 % (14 occurrences sur 62) sur le total des contextes exprimant une relation de finalité.

La majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, 18 % des patrons sont partagés par les deux corpus alors que 82 % des patrons sont exclusifs de l'informatique ou de la médecine. Tout

comme le TS [PHENOMENON], le TS [ACTIVITY] atteint le pourcentage le plus élevé de patrons partagés par les deux corpus dans notre étude.

4.5 TS : [PSYCHOLOGICAL FEATURE (FIELD)]

Pour le TS [PSYCHOLOGICAL FEATURE (FIELD)], 16 contextes ont été retenus. En médecine, nous avons sélectionné les termes *chirurgie* et *médecine*, et en informatique, les termes *génie* et *informatique*. Rappelons que pour ce TS, un grand nombre de contextes ont été étudiés, mais seulement un petit nombre d'entre eux contenaient effectivement l'expression d'une relation sémantique. En effet, pour trois des quatre termes étudiés pour ce TS, soit les termes *informatique*, *chirurgie* et *médecine*, la proportion des contextes retenus par rapport au nombre de contextes étudiés est respectivement de 1 %, 1 % et 7 %.

Le Tableau 39 présente, par domaine, les résultats obtenus pour le TS [PSYCHOLOGICAL FEATURE (FIELD)] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 40 montre la répartition dans les deux corpus des trois relations sémantiques observées. Elles sont présentées par ordre décroissant.

Tableau 39. Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
HYPERONYMIQUE	2	2	X est Y (2x)	Informatique
FINALITÉ	13	4	X permet Y, moyen de Y X, on emploie Y en X, on utilise Y en X	Médecine
		9	<u>Objectif de X Y (4x)</u> , X consiste à Y (2x), défi de X Y, but de X Y, X permet Y	Informatique
MÉRONYMIQUE	1	1	Dimension de X Y	Informatique

Tableau 40. Répartition des relations sémantiques par ordre décroissant pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]

Relation	Nombre/Total	Pourcentage
Relation de finalité	13/16	81 %
Relation hyperonymique	2/16	13 %
Relation méronymique	1/16	6 %

La relation la plus importante pour le TS [PSYCHOLOGICAL FEATURE (FIELD)] est la relation de finalité, qui est exprimée dans 81 % des contextes observés. Nous avons reproduit des exemples en 110 et en 111, qui sont tirés respectivement de nos corpus d'informatique et de médecine.

110. L'un des **buts du génie logiciel** est de fournir les méthodes et les outils pour mieux gérer cette complexité, en particulier les tests (genilo).
111. La *chirurgie endoculaire* Elle **permet** le traitement des complications (hémorragie du vitré – décollement de rétine) (opht126).

La relation de finalité est suivie, de loin, par les relations hyperonymique (13 %) et méronymique (6 %). Quant aux relations causales et aux patrons métalinguistiques, ils sont complètement absents de nos corpus pour ce TS.

Le patron le plus fréquent est le patron de finalité : **objectif de X Y**. Il est présent dans 25 % des contextes étudiés pour l'ensemble des relations (4 patrons sur 16), et dans 31 % des contextes exprimant une relation de finalité (4 patrons sur 13). Ce patron n'est présent que dans le corpus d'informatique.

Pour le TS [PSYCHOLOGICAL FEATURE (FIELD)], on dénombre aussi 13 % de contextes contenant l'expression d'une relation hyperonymique. Le contexte 112 provient de notre corpus d'informatique et présente une illustration de cette relation.

112. Le *génie logiciel* **est** une métadiscipline qui fixe des règles et des garde-fous qui vont agir comme autant de conditions nécessaires au bon déroulement du processus de fabrication (genilo).

Enfin, le TS [PSYCHOLOGICAL FEATURE (FIELD)] est associé à une seule relation méronymique, qui est de nature partie fonctionnelle – tout. Le contexte 113 provient de notre corpus d'informatique.

113. L'ingénierie des données est une **dimension** très importante **du génie logiciel** (*genilo*).

Le Tableau 41 reprend la répartition des relations sémantiques pour le TS [PSYCHOLOGICAL FEATURE (FIELD)], cette fois-ci par terme clé.

Tableau 41. Répartition des relations sémantiques par terme clé pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]

	RELATION CAUSALE	RELATION DE FINALITÉ	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Médecine (médecine)	0	2	0	0
Chirurgie (médecine)	0	2	0	0
Informatique (informatique)	0	3	0	0
Génie (informatique)	0	6	2	1 (partie fonctionnelle – tout)

Comme le montre le Tableau 41, les relations sémantiques sont plus nombreuses et plus variées pour le terme *génie*. En outre, seul le terme *génie* est associé à des relations sémantiques autres que la relation de finalité, soit les relations hyperonymique et méronymique. Toutefois, comme les contextes dans lesquels on retrouve une relation sémantique sont rares pour ce type sémantique, il est extrêmement difficile d'en tirer une conclusion générale.

4.5.1 *Répartition des relations sémantiques par corpus*

Douze contextes sur 16 (75 %) proviennent du corpus d'informatique, contre seulement 4 contextes (25 %) pour le corpus de médecine. Dans cette section, nous tenterons de fournir une explication à ce déséquilibre en comparant les résultats obtenus pour chacun des deux corpus. Les Tableaux 42 et 43 présentent, par corpus, la répartition des relations sémantiques observées. Celles-ci sont réparties par ordre décroissant.

Tableau 42. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]³⁵

Relation	Nombre/Total
Relation de finalité	4/4
Relation hyperonymique	0/4
Relation méronymique	0/4

Tableau 43. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]

Relation	Nombre/Total	Pourcentage
Relation de finalité	9/12	75 %
Relation hyperonymique	2/12	17 %
Relation méronymique	1/12	8 %

³⁵ Dans les cas où le nombre d'occurrences est inférieur à 10, nous avons omis d'inscrire le pourcentage correspondant afin de tenir compte des principes de la statistique.

Comme le montrent les Tableaux 42 et 43, la relation de finalité est la principale relation sémantique dans chacun des deux corpus. Elle compte en effet pour la totalité des relations du corpus de médecine, et pour 75 % des relations du corpus d'informatique. Ce dernier compte également 17 % de relations hyperonymiques et 8 % de relations méronymiques. Le déséquilibre entre les deux corpus provient surtout des caractéristiques distinctes des termes. En effet, le terme génie est le seul terme à être associé aux relations hyperonymique et méronymique.

4.5.2 Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE (FIELD)]

Dans cette sous-section, nous traitons des différentes formes de patrons relevés pour le TS [PSYCHOLOGICAL FEATURE (FIELD)] dans chacun des corpus et dans les deux corpus. Le Tableau 44 présente la répartition des patrons dans chacun des corpus, et ceux qui sont communs au corpus de médecine et au corpus d'informatique.

Tableau 44. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Hyperonymique	*	X est Y (2x)	*
Méronymique	*	Dimension de X Y	*
Finalité	Moyen de Y X, on emploie Y en X, on utilise Y en X	Objectif de X Y (4x), X consiste à Y (2x), défi de X Y, but de X Y	X permet Y (1x, 1x)

Comme le montre le Tableau 44, la majorité des PRS sont propres au domaine de l'informatique ou à celui de la médecine. En effet, seulement un patron (10 %), exprimant une relation de finalité, est partagé par les deux corpus. La relation de finalité est précisément la relation la plus importante pour ce TS. Par ailleurs, 9 patrons (90 %) sont propres à l'informatique ou à la médecine. Ce résultat n'est guère surprenant puisque peu de contextes ont été recueillis.

4.5.3 *Remarques générales sur le TS [PSYCHOLOGICAL FEATURE (FIELD)]*

En conclusion, en ce qui concerne le TS [PSYCHOLOGICAL FEATURE (FIELD)], il semble rarement associé à des relations sémantiques (du moins, celles que nous avons étudiées). Seules les relations hyperonymique, méronymique et de finalité sont représentées pour ce TS, alors que la relation causale et les patrons métalinguistiques sont absents de nos corpus. C'est la relation de finalité qui est la plus importante pour les deux corpus. En effet, elle est exprimée dans 81 % des contextes étudiés. Elle est la seule relation exprimée dans le corpus de médecine, alors que dans le corpus d'informatique, elle apparaît dans 75 % des contextes. Le corpus d'informatique compte également 17 % de relations hyperonymiques et 8 % de relations méronymiques. Enfin, un seul patron sur 10 est partagé par les deux corpus, alors que 90 % des patrons sont exclusifs de l'informatique ou de la médecine.

Le patron le plus fréquent, toutes relations confondues, est le patron **objectif de X Y**, patron exprimant une relation de finalité. Il est présent dans 25 % des cas pour l'ensemble des relations, et dans 31 % des cas pour les relations de finalité.

4.6 **TS : [ENTITY ORGANISM]**

Nous avons retenu, au total, 16 contextes pour le TS [ENTITY ORGANISM]. En médecine, nous avons sélectionné les termes : *médecin* et *patient*, et en

informatique, les termes *programmeur et utilisateur*. Pour ce TS, comme pour le type sémantique [PSYCHOLOGICAL FEATURE (FIELD)], qui a fait l'objet de la section 4.5, un grand nombre de contextes ont été étudiés, mais seulement un petit nombre d'entre eux contenaient effectivement l'expression d'une relation sémantique. En effet, pour les quatre termes étudiés pour ce TS, soit les termes *utilisateur, patient, programmeur et médecin*, la proportion des contextes retenus par rapport au nombre de contextes étudiés est respectivement de 1 %, 1 %, 4 % et 13 %.

Le Tableau 45 présente, par domaine, les résultats obtenus pour le TS [ENTITY ORGANISM] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 46 présente la répartition dans les deux corpus des 5 relations sémantiques par ordre décroissant.

Tableau 45. Répartition des PRS pour le TS [ENTITY ORGANISM]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
CAUSALE	2	1	On attribue Y à X	Médecine
		1	X provoque Y	Informatique
HYPERONYMIQUE	2	2	X comme Y, X et autre Y	Informatique
MÉRONYMIQUE	3	3	X appartient à Y, <u>X fait partie de Y (2x)</u>	Informatique
PATRON MÉTALINGUISTIQUE	1	1	X c'est-à-dire Y	Médecine
FINALITÉ	8	3	<u>Rôle de X Y</u> , X doit Y, Y s'adresse à X	Médecine
		5	<u>Rôle de X Y</u> , X tenu de Y, X se consacre à Y, X capable de Y, Y au service de X	Informatique

Tableau 46. Répartition des relations sémantiques par ordre décroissant pour le TS [ENTITY ORGANISM]

Relation	Nombre/Total	Pourcentage
Relation de finalité	8/16	50 %
Relation méronymique	3/16	19 %
Relation causale	2/16	13 %
Relation hyperonymique	2/16	13 %
Patron métalinguistique	1/16	6 %

Comme le montre le Tableau 46, toutes les relations sont représentées pour ce TS, même si un petit nombre de contextes contient l'expression d'une relation sémantique. La relation de finalité est la relation la plus importante avec 50 % pour l'ensemble de ce TS. Des exemples tirés respectivement de nos corpus de médecine et d'informatique sont présentés en 114 et en 115.

114. Cela revient à souligner l'importance du **rôle du médecin traitant** [...] non seulement prescrire, mais assurer, pour un traitement de longue durée, l'observation de la prescription (xgiroud).
115. ... les machines pourront tirer parti des informations pour effectuer les tâches répétitives **au service de l'utilisateur** (missio).

Pour ce TS, nous n'avons relevé qu'un seul patron métalinguistique que nous reproduisons en 116.

116. Il faut donc avoir recours à une présélection qui permet d'éliminer : après un entretien préliminaire, les *patients à « risque élevé »*, **c'est-à-dire** les malades ayant des difficultés de liaison avec l'hôpital (alcooliques, psychopathes, etc.) (xgiroud).

Les patrons les plus fréquents pour ce TS sont les patrons : **X fait partie de Y** et **rôle de X Y**. Le premier est un patron méronymique provenant du corpus d'informatique. Il est présent dans 2 contextes sur 8 exprimant une relation méronymique, et dans 13 % des contextes étudiés (2 occurrences sur 16) pour ce TS. Quant au second patron (**rôle de X Y**) exprimant la finalité, il est présent dans 2 contextes sur 8 exprimant une relation de finalité, et dans 13 % des contextes étudiés (2 occurrences sur 16) pour ce TS.

On dénombre également, pour le TS [ENTITY ORGANISM], deux relations causales, dont une est de nature modification (corpus d'informatique) et une est de nature création (corpus de médecine), illustrées en 117 et en 118.

Modification :

117. En déplaçant la souris sur le bureau, l'*utilisateur* **provoque** une réaction à l'écran sur lequel une flèche se déplace, comme en écho

du mouvement de la souris sur le bureau, dans la même direction et à la même vitesse (plaisen1).

Création :

118. La digitale a été la première plante reconnue comme renfermant des composés susceptibles d'atténuer les signes d'insuffisance cardiaque et c'est à un *médecin* de Birmingham, William Withering, que l'**on attribue** cette découverte en l'an 1785 (xrenal1).

Dans les deux contextes exprimant une relation causale associée au TS [ENTITY ORGANISM], c'est le X (le terme de la requête) qui constitue la cause.

Par ailleurs, on dénombre trois contextes contenant l'expression d'une relation méronymique. Ils proviennent tous trois du corpus d'informatique. Dans les trois contextes, la relation est exprimée entre un élément et un ensemble. Un exemple est présenté en 119.

119. Dans ce cas, on pourra résoudre ces problèmes en attribuant un groupe d'utilisateurs video, auquel **appartiendront** tous les *utilisateurs* ayant le droit d'utiliser la carte d'acquisition TV, et de changer le groupe des fichiers spéciaux de périphériques /dev/videoX (linux3p2).

Le Tableau 47 reprend la répartition des relations sémantiques pour le TS [ENTITY ORGANISM], cette fois-ci par terme clé.

**Tableau 47. Répartition des relations sémantiques par terme clé pour le TS
[ENTITY ORGANISM]**

	RELATION CAUSALE	PATRON MÉTALIN- GUISTIQUE	RELATION DE FINALITÉ	RELATION HYPERONY- MIQUE	RELATION MÉRONY- MIQUE
Médecin (médecine)	1 (création)	0	2	0	0
Patient (médecine)	0	1	1	0	0
Programmeur (informatique)	0	0	3	1	0
Utilisateur (informatique)	1 (modification)	0	2	1	3 (élément – ensemble)

Comme le montre le Tableau 47, c'est le terme *utilisateur* – terme pour lequel nous avons retenu le plus grand nombre de contextes – qui est associé au plus grand nombre de relations sémantiques (causale, de finalité, hyperonymique et méronymique). Quant aux termes *programmeur* et *médecin*, ils sont tous deux associés à seulement deux relations chacun. Le premier est associé aux relations de finalité et d'hyperonymie, tandis que le second est associé aux relations causale et de finalité. Enfin, le terme *patient* n'est jumelé qu'à une relation de finalité et à un patron métalinguistique.

4.6.1 *Répartition des relations sémantiques par corpus*

Onze contextes sur 16 (69 %) proviennent du corpus d'informatique, contre 5 contextes (31 %) pour le corpus de médecine. Dans cette section, nous comparerons les résultats obtenus pour chacun des corpus afin de tenter de fournir une explication à ce déséquilibre. Les Tableaux 48 et 49 présentent, par corpus, la répartition des relations sémantiques observées par ordre décroissant.

Tableau 48. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ENTITY ORGANISM]

Relation	Nombre/Total
Relation de finalité	3/5
Relation causale	1/5
Patron métalinguistique	1/5

Tableau 49. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [ENTITY ORGANISM]

Relation	Nombre/Total	Pourcentage
Relation de finalité	5/11	45 %
Relation méronymique	3/11	27 %
Relation hyperonymique	2/11	18 %
Relation causale	1/11	9 %
Patron métalinguistique	0/11	0 %

Comme le montrent les Tableaux 48 et 49, la relation de finalité est la relation la plus importante dans chacun des corpus. En effet, le corpus de médecine compte 3 relations de finalité sur un total de 5, et le corpus de médecine en compte 45 %. Pour les autres relations, l'écart entre le nombre de relations par corpus varie peu ou pas du tout. En ce qui a trait à la relation causale, par exemple, nos corpus de médecine et d'informatique comptent chacun une seule relation.

Enfin, les relations méronymique et hyperonymique sont liées aux termes d'informatique seulement, alors que le seul patron métalinguistique est lié à un terme de médecine.

4.6.2 *Répartition des PRS pour le TS [ENTITY ORGANISM]*

Cette sous-section examine la forme des patrons relevés pour le TS [ENTITY ORGANISM] dans chacun des corpus et dans les deux corpus. Le Tableau 50 montre la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 50. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Causale	On attribue Y à X	X provoque Y	*
Hyperonymique	*	X comme Y, X et autre Y	*
Méronymique	*	X appartient à Y, X fait partie de Y (2x)	*
Patron métalinguistique	X c'est-à-dire Y	*	*
Finalité	X doit Y, Y s'adresse à X	X tenu de Y, X se consacre à Y, X capable de Y, Y au service de X	Rôle de X Y (1x, 1x)

Comme le montre le Tableau 50, la majorité des PRS sont propres au domaine de l'informatique ou à celui de la médecine. En effet, un seul patron sur quatorze (7 %), exprimant une relation de finalité, est commun aux deux corpus (rappelons qu'il s'agissait de la relation sémantique la plus importante pour ce TS). Par ailleurs, treize patrons (93 %) sont propres à l'informatique ou à la médecine.

4.6.3 *Remarques générales sur le TS [ENTITY ORGANISM]*

En conclusion, en ce qui concerne le TS [ENTITY ORGANISM], toutes les relations sémantiques sont représentées, bien que le nombre de contextes exprimant chacune des relations soit très peu élevé. C'est la relation de finalité qui est la plus importante. En effet, elle est exprimée dans 50 % des contextes étudiés. Il s'agit également de la relation sémantique la plus importante dans chacun des

deux corpus (60 % des contextes étudiés en informatique et 45 % des contextes en médecine).

En outre, le corpus de médecine compte 20 % de relations causales. Le corpus d'informatique compte 27 % de relations méronymiques et 18 % de relations hyperonymiques. Quant au seul patron métalinguistique associé à ce TS, il apparaît dans notre corpus de médecine.

Vu le peu de contextes répertoriés dans le corpus de médecine, les résultats ne permettent pas de tirer une conclusion générale pour ce TS puisque aucune relation ne se démarque vraiment des autres.

Les patrons les plus fréquents pour ce TS sont **X fait partie de Y** et **rôle de X Y**. Le premier est un patron méronymique relevé uniquement dans le corpus d'informatique. Il est présent dans 2 contextes étudiés sur 8 parmi les patrons appartenant à la relation méronymique, et 13 % des contextes étudiés (2 patrons sur 16) pour l'ensemble du TS [ENTITY ORGANISM]. Quant au second patron (**rôle de X Y**), il s'agit d'un patron de finalité qui est présent dans nos deux corpus. Il est également présent dans 2 contextes étudiés sur 8 parmi les patrons appartenant à la relation de finalité, et 13 % des contextes étudiés (2 patrons sur 16) pour l'ensemble de ce TS.

La majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, un seul patron sur quatorze (7 %) est partagé par les deux corpus alors que 93 % des patrons sont exclusifs de l'informatique ou de la médecine.

4.7 TS : [PHENOMENON]

Au total, 51 contextes contenaient l'explication d'une relation sémantique pour le TS [PHENOMENON]. En médecine, nous avons sélectionné les termes *développement* et *vieillessement*, et en informatique, les termes *développement* et *évolution*. Le Tableau 51 présente, par domaine, les résultats obtenus pour le TS

[PHENOMENON] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 510 présente la répartition dans les deux corpus des 3 relations sémantiques par ordre décroissant.

Tableau 51. Répartition des PRS pour le TS [PHENOMENON]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
CAUSALE	38	14	Avec X Y, Y stimule X, <u>X entraîne Y</u> , <u>Y entraîne X</u> (2x), Y contribue à X, Y favorise X, X résulte de Y, conséquence de X Y, Y explique X, cause de Y X (2x) Y dû à X, X conduit à Y	Médecine
		24	X conduit à Y, Y condition nécessaire à X, facteur de Y X, facteur dans X Y, compte tenu de X Y, Y implique X, cause de X Y, conséquence de X Y, avec X Y, X lié à Y, clé essentielle de X Y, résultat de Y X, <u>Y entraîne X</u> , <u>X entraîne Y</u> (2x), <u>Y entraîne conséquence sur X</u> , X dépend de Y, Y repose sur X, Y encourage X, frein à X Y, Y freine X, Y existe grâce à X, Y influence X, rôle dans X Y	Informatique
HYPERONYMIQUE	1	1	X tel Y	Informatique
FINALITÉ	12	12	Objectif de Y X, à l'aide de Y X, pour X Y, Y permet X (2x), X permet Y (2x), favoriser X permet Y, Y favorise X, Y assure X, Y destiné à X, Y facilite X	Informatique

Tableau 52. Répartition des relations sémantiques par ordre décroissant pour le TS [PHENOMENON]

Relation	Nombre/Total	Pourcentage
Relation causale	38/51	74 %
Relation de finalité	12/51	24 %
Relation hyperonymique	1/51	2 %

La relation la plus importante pour le TS [PHENOMENON] est la relation causale, qui apparaît dans 74 % des contextes observés. Nous avons relevé 7 relations causales de nature différente : 17 relations augmentation, 9 relations création, 4 relations destruction, 2 relations modification, 3 relations diminution, 2 relations conservation et 1 relation maintien, illustrés par les exemples 120 à 128, tirés de nos corpus de médecine et d'informatique.

Augmentation :

120. **Avec** le *vieillissement* de la population, les problèmes neurologiques du sujet âgé sont de plus en plus fréquents et leur analyse justifie souvent la réalisation d'une imagerie (image1).
121. **Avec** les récents *développements* technologiques, la mémoire principale peut atteindre une capacité de 20 mégaoctets (pierre1).

Création :

122. Alors que les deux héli-moelle se rejoignent au-dessous d'un septum osseux ou fibroduremerien, la moelle ne peut pas assurer son ascension relative au cours de la croissance foetale, ce qui **peut entraîner** le *développement* d'une myélopathie secondaire par syndrome de la moelle fixée (voir ci-dessous)(snc).
123. Comme nous l'avons vu, l'*évolution* des ordinateurs **a entraîné** le besoin de disposer d'un système d'exploitation réellement multitâche capable de gérer des ordinateurs qui devenaient interconnectés et fonctionnaient en temps partagé pour fournir des services à des utilisateurs de plus en plus nombreux (internet).

Destruction :

124. Ce phénomène est particulièrement important chez l'enfant et l'adolescent : à cet âge, la principale **cause** [...] **de** mortalité [...] est le *développement* [...] d'une hypertension intracrânienne sans cause chirurgicalement traitable (cran).

Modification :

125. Depuis quelques années, aux États-Unis en particulier, la tendance s'est inversée et la brevetabilité des logiciels a été reconnue. L'Europe s'apprêterait à suivre la même voie. Ce changement a d'ores et déjà une **influence** sur l'*évolution* de l'Internet et du Web (missio).

Diminution :

126. Un des enjeux du domaine (qui **peut** parfois **être un frein au développement**) est la maîtrise de la complexité croissante : la réalisation d'une fonction donnée nécessite souvent l'intégration de composants logiciels hétérogènes et parfois des adaptations difficiles (voir par exemple la complexité des navigateurs Web) (mission).

Conservation :

127. L'*évolution* sur le long terme **dépendra** donc **du** besoin effectif en matière de mécanismes de qualité de service... (missio)

Maintien :

128. Au contraire, l'Internet est un standard simple, qui **n'existe** que **grâce au développement** collectif, effectué dans la transparence et la recherche de l'efficacité plutôt que par l'emballage commercial (societ).

Le PRS le plus fréquent, toutes relations confondues, est le patron causal **X entraîne Y**, qui regroupe les formes suivantes : **X entraîne Y (3x)**, **Y entraîne X (3x)** et **Y entraîne conséquence sur X**. Il apparaît dans 14 % des contextes étudiés (7 occurrences sur 51) pour ce TS, et dans 18 % des contextes (7 occurrences sur 38) exprimant une relation causale. Voici des exemples tirés respectivement de nos corpus de médecine et d'informatique (129 et 130).

129. Le *développement latéral* de la masse **entraîne** une déformation progressive de la pièce sacro-coccygienne avec aspect de sacrum « en cimenterre » correspondant à une déformation de l'os, plutôt qu'à son agénésie (snc).
130. L'*évolution* [...] de la technologie des semi-conducteurs, **a entraîné** une évolution concomitante des chemins de données (trajet).

Dans les patrons de relation causale relevés, le **X** (c'est-à-dire le terme de la requête) représente l'effet dans 53 % des contextes. Les patrons suivants sont des illustrations de ce cas de figure : **Y contribue à X**, **cause de X Y**, etc. Dans 47 % des cas, le **X** renvoie à la cause, comme dans les patrons suivants : **X entraîne Y**, **conséquence de X Y**, **facteur de Y X**, etc. Les résultats sont donc équilibrés pour ce type sémantique, puisqu'il y a presque autant de PRS où le X représente l'effet, que de PRS où le X renvoie à la cause.

Le TS [PHENOMENON] est également associé à 24 % de relations de finalité et 2 % de relations hyperonymiques. Aucune relation méronymique ni aucun patron métalinguistique n'a été relevé.

Les exemples 131 et 132 illustrent les autres relations observées pour ce type sémantique, soit les relations de finalité et d'hyperonymie.

Finalité (informatique) :

131. L'action "Globalisation des ressources informatiques et des données" (GRID) du ministère de la Recherche **a pour objectif** le *développement* des activités de recherche dans ce domaine pour inventer et mettre au point de nouveaux modèles (cosnar).

Hyperonymique (informatique) :

132. Ce modèle est toutefois en passe d'atteindre ses limites car les nouvelles *évolutions* de l'Internet **telles** que le passage à IPv6, nécessitent des couplages plus fins entre les mécanismes réseaux et les applications pour améliorer la sécurité et la fiabilité, la qualité de service et les performances (missio).

Le Tableau 53 reprend la répartition des relations sémantiques pour le TS

[PHENOMENON], cette fois-ci par terme clé.

Tableau 53. Répartition des relations sémantiques par terme clé pour le TS [PHENOMENON]

	RELATION CAUSALE	RELATION HYPERO-NYMIQUE	RELATION DE FINALITÉ	RELATION MÉRONYMIQUE
Vieillessement (médecine)	3 (1 augmentation, 1 destruction et 1 modification)	0	0	0
Développement (médecine)	11 (7 création, 2 augmentation, 2 destruction)	0	0	0
Évolution (informatique)	10 (7 augmentation, 1 création, 1 conservation et 1 modification)	1	3	0
Développement (informatique)	14 (1 conservation, 7 augmentation, 1 maintien et 1 création, 4 modification)	0	9	0

Comme le montre le Tableau 53, pour les termes *vieillissement* et *développement* (médecine), seule la relation causale est représentée, alors que pour le terme *évolution*, trois relations sont représentées, soit la relation causale, la finalité et l'hyponymie. Quant au terme *développement* (informatique), il est associé aux relations causale et de finalité. De plus, il est associé beaucoup plus fréquemment que les autres termes à la relation de finalité.

4.7.1 *Répartition des relations sémantiques par corpus*

Puisque 37 contextes sur 51 (73 %) proviennent du corpus d'informatique, contre seulement 14 contextes (27 %) pour le corpus de médecine, nous examinerons les résultats obtenus de plus près pour chacun des corpus et commenterons cet écart. Les Tableaux 54 et 55 présentent, par corpus, la répartition des relations sémantiques observées par ordre décroissant.

Tableau 54. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [PHENOMENON]

Relation	Nombre/Total	Pourcentage
Relation causale	14/14	100 %
Relation de finalité	0/14	0 %
Relation hyperonymique	0/14	0 %

Tableau 55. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [PHENOMENON]

Relation	Nombre/Total	Pourcentage
Relation causale	24/37	65 %
Relation de finalité	12/37	32 %
Relation hyperonymique	1/37	3 %

Comme le montrent les Tableaux 54 et 55, bien que la relation causale soit la principale relation observée dans les deux corpus, les écarts sont assez importants entre les corpus. En effet, dans le corpus de médecine le TS [PHENOMENON] est toujours associé à une relation causale. En revanche, il n'est jumelé à une relation causale que dans 65 % des contextes extraits du corpus d'informatique. La relation causale est suivie dans le corpus d'informatique par la relation de finalité, avec 32 %, et, loin derrière, par la relation hyperonymique (3 %). Comme on l'a vu, le terme *développement* (informatique) est souvent associé à la relation de finalité. Quant aux relations hyperonymique et de finalité, elles sont absentes du corpus de médecine pour ce TS.

4.7.2 *Répartition des PRS pour le TS [PHENOMENON]*

Cette sous-section examine la forme des patrons relevés pour le TS [PHENOMENON] dans chacun des corpus et dans les deux corpus. Le Tableau 56 montre la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 56. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Causale	Y stimule X, Y contribue à X, Y favorise X, Y explique X, Y dû à X	Y condition nécessaire à X, facteur de Y X (facteur dans X Y), compte tenu de X Y, Y implique X, X lié à Y, clé essentielle de X Y, Y entraîne X, X dépend de Y, Y repose sur X, Y encourage X, frein à X Y (Y freine X), Y existe grâce à X, Y influence X, rôle dans X Y	Avec X Y (1x, 1x), <u>X entraîne Y (1x, 2x)</u> (Y entraîne X (2x, 0x), Y entraîne conséquence sur X (0x, 1x), X résulte de Y (1x, 0x) (résultat de Y X), conséquence de X Y (1x, 1x), X conduit à Y (1x, 1x), cause de X Y (0x, 1x) (cause de Y X (2x, 0x))
Hyperonymique	*	X tel Y	*
Finalité	*	Objectif de Y X, à l'aide de Y X, pour X Y, Y permet X (2x) (X permet Y (2x), favoriser X permet Y), Y favorise X, Y assure X, Y destiné à X, Y facilite X	*

Comme le montre le Tableau 56, la majorité des PRS associés à ce TS, soit 28 patrons sur 34 (82 %), sont propres au domaine de l'informatique ou de la médecine. Toutefois, 6 (18 %) patrons causaux sont partagés par les deux corpus (rappelons qu'il s'agissait de la relation sémantique la plus importante pour ce TS). Il s'agit de l'un des deux TS pour lequel le plus grand nombre de patrons communs a été relevé, car il arrive ex-æquo avec le TS [ACTIVITY].

4.7.3 *Remarques générales sur le TS [PHENOMENON]*

En conclusion, en ce qui concerne le TS [PHENOMENON], la relation causale est la plus importante. En effet, elle est exprimée dans 74 % des contextes étudiés. Cette relation est exprimée dans la totalité des contextes extraits du corpus de médecine. En contrepartie, aucun patron métalinguistique ni aucune relation méronymique n'a été répertorié ni dans le corpus de médecine, ni dans le corpus d'informatique. En informatique, on a également observé les relations sémantiques suivantes : finalité et hyperonymie.

Le PRS le plus fréquent, toutes relations confondues, est le patron causal **X entraîne Y**, qui regroupe les formes suivantes : **X entraîne Y (3x)**, **Y entraîne X (3x)** et **Y entraîne conséquence sur X**, présent tant en médecine qu'en informatique. Ce patron causal est présent dans 14 % des cas pour l'ensemble des relations étudiées, et dans 18 % des cas pour l'ensemble des relations causales.

Enfin, 18 % des patrons sont partagés par les deux corpus alors que 82 % des patrons sont exclusifs de l'informatique ou de la médecine. Il s'agit de l'un des deux TS, avec le TS [ACTIVITY], pour lequel le plus grand nombre de patrons communs a été relevé, soit 18 % dans le cas des patrons partagés par les deux corpus (voir section 4.4).

4.8 TS : [STATE]

Nous avons retenu, au total, 72 contextes pour le TS [STATE]. En médecine, ces contextes ont été extraits à partir des termes *état* et *maladie*, et en informatique, à partir des termes *état* et *bogue* (ou *bug*). Le Tableau 57, présente, par domaine, les résultats obtenus pour le TS [STATE] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 57 présente la répartition dans les deux corpus des 5 relations sémantiques. La répartition est présentée par ordre décroissant.

Tableau 57. Répartition des PRS pour le TS [STATE]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
FINALITÉ	3	3	On utilise Y dans X (3x)	Médecine
CAUSALE	38	5	X dû à Y, X vient du fait Y, Y apporte X, Y provoque X, Y déterminé par X	Informatique
		33	Résultat de X Y, Y cause X, cause de Y X (3x), cause de X Y, Y consécutif à X, X donne Y, X dû à Y, X entraîne Y (3x), Y entraîne X, Y n'entraîne pas X, Y lié à X (2x), X lié à Y (2x), responsable de X Y (2x), responsable de Y X, générateur de X Y, secondaire à X Y (2x), Y précipite X, X favorise Y, origine de X Y, X entre en ligne de compte pour Y, Y influe sur X, Y aggrave X, Y modifie X, en fonction de X Y, Y dépend de X	Médecine
PATRON MÉTALINGUISTIQUE	6	3	X ou Y, Y ou X (2x)	Médecine
		3	X c'est-à-dire Y, Y c'est-à-dire X (2x)	Informatique
HYPERONYMIQUE	18	18	<u>Y est X (9x), X est Y, Y c'est X, X c'est Y</u> , X il s'agit de Y, Y en particulier X, X en particulier Y, type de X Y, autre Y X, X comme Y	Médecine
MÉRONYMIQUE	7	7	Y contient X, Y comporte X (5x), Y n'est pas exempt de X	Informatique

Tableau 58. Répartition des relations sémantiques par ordre décroissant pour le TS [STATE]

Relation	Nombre/Total	Pourcentage
Relation causale	38/72	53 %
Relation hyperonymique	18/72	25 %
Relation méronymique	7/72	10 %
Patrons métalinguistiques	6/72	8 %
Relation de finalité	3/72	4 %

Toutes les relations sémantiques sont représentées pour le TS [STATE], y compris les patrons métalinguistiques. La relation la plus importante est la relation causale, qui apparaît dans 53 % des contextes observés. La relation la moins importante est la relation de finalité attestée dans 4 % des contextes seulement.

Bien que la relation causale soit la relation la plus importante pour le TS [STATE], le patron le plus fréquent pour l'ensemble des contextes renfermant un terme appartenant à ce TS est un patron hyperonymique : **X est Y**. Ce patron, exclusif au corpus de médecine, regroupe les formes suivantes : **X est Y, Y est X (9x), Y c'est X, X c'est Y**. Il est présent dans 67 % des contextes exprimant une relation hyperonymique (12 occurrences sur 18), et dans 17 % des contextes étudiés (12 occurrences sur 72) pour ce TS.

Pour le TS [STATE], 8 % des contextes contiennent un patron métalinguistique, 25 %, l'expression d'une relation hyperonymique et 10 %, l'expression d'une relation méronymique.

Les sept relations méronymiques font partie de ce que nous avons appelé les *inclassables*, car elles n'entrent pas dans la classification de Winston *et al.* (1987). Elles proviennent toutes du corpus d'informatique et elles sont toutes associées au terme *bogue (bug)*. En voici un exemple (133) :

133. ...mais des essais nous ont montré que cette ROM **comportait** beaucoup trop de *bugs* (impri).

Quant aux 38 relations causales, elles se répartissent de la manière suivante : 28 expriment la création, 6, la modification, 2, l'augmentation, 1, le maintien et 1, la diminution. Des exemples tirés de nos corpus sont reproduits en (134 à 139).

Création :

134. Des carences (syndrome de Gayet-Wernicke) et la *maladie dégénérative* (paralysie supranucléaire progressive) **entraînent** des paralysies oculomotrices (opht121)
135. Signalons aussi un *bogue dû à* ce type de cartes (macosx).

Modification :

136. C'est ainsi que des accidents ischémiques, coronaires ou cérébraux, **peuvent influencer sur** l'*état cardiaque et neurologique* ou même *rénal* (xgiroud).
137. L'enchaînement des actions **est déterminé par** l'*état* d'un certain nombre de « registres » qui contiennent l'information nécessaire à l'enchaînement (c'est une généralisation de la notion de compteur ordinal, de registre interruption, de mot d'état... que l'on trouve dans tout ordinateur) (genilo).

Augmentation :

138. Les hypertensions neurotoniques, sujettes à des paroxysmes **favorisés par** les *états émotifs*, tirent donc bénéfice des propriétés neuroleptiques associées de la réserpine (xrenal1).

Diminution :

139. Près de 3 % vont souffrir d'une perte visuelle sévère à **cause de** cette *maladie* [*rétinopathie diabétique*] (ret_diab).

Maintien :

140. L'interruption des AVK est définitive quand le traitement anticoagulant n'apparaît plus indispensable. Son arrêt brutal est possible car aucune étude n'a démontré qu'il **entraînait** un *état d'hypercoagulabilité* avec un risque accru de thrombose (zdjiane).

Pour le TS [STATE], il y a presque autant de X qui expriment la cause que l'effet. Dans 45 % (17 contextes) des cas, le X exprime la cause, et dans 55 % (21 contextes) des cas, le X exprime l'effet. Nous présentons un exemple (141) dans lequel le X exprime la cause, et un exemple (142) où le X exprime l'effet. Ces exemples sont tirés respectivement de nos corpus de médecine et d'informatique.

141. Le **résultat de** cette *maladie* est la cécité nocturne (au début) puis la restriction des champs de vision (vision en tunnel) (ret_pig).
142. Le *bug* de l'an 2000 **vient du fait** que les dates sont sur de nombreux ordinateurs codées sur deux chiffres, c'est-à-dire qu'au lieu d'afficher (de calculer) 4 chiffres (comme 1974), ces ordinateurs affichent 2 chiffres (donc 74) (sysint).

Quant aux patrons métalinguistiques, ils font partie des relations les moins représentées pour le TS [STATE]. Nous en avons reproduit des exemples tirés respectivement de nos corpus de médecine (143) et d'informatique (144).

143. Le rétrécissement aortique dégénératif **ou** *maladie de Mönckeberg* (retr_aor).
144. La Table d'Allocation de Fichiers est une liste de valeurs numériques permettant de décrire l'allocation des clusters d'une partition, **c'est-à-dire** l'*état* de chaque cluster de la partition dont elle fait partie (optrep).

Les relations hyperonymiques apparaissent dans 25 % des contextes étudiés pour le TS [STATE]. Cette relation n'est présente que dans notre corpus de médecine. Nous en avons reproduit un exemple en 145.

145. Les leucodystrophies **sont** des *maladies métaboliques* d'origine enzymatique (image7).

Enfin, les relations de finalité comptent pour 4 % de l'ensemble du TS [STATE] et ne sont présentes, encore une fois, que dans le corpus de médecine. Nous avons reproduit un exemple en 146.

146. **On l'utilise** [isoprénaline] **dans** les *états de choc post-opératoires*, et au cours de défaillances aiguës : asphyxie, intoxication, prématuré anoxique... (zcarraz).

Bien que la proportion de contextes exprimant une relation de finalité soit peu élevée, nous sommes surpris de constater leur présence. Pour cette raison, nous les examinerons une à une (147 à 149) et tenterons d'expliquer pourquoi il arrive que cette relation puisse être associée à ce type sémantique.

147. L'ajmaline.– L'ajmaline est un alcaloïde d'une plante, le rauwolfia. Elle diminue l'excitabilité du myocarde et ralentit le rythme. Elle est employée dans l'hyperexcitabilité ventriculaire. **On l'utilise dans** la *maladie de Bouveret* (ztoutou).
148. L'effet de l'héparine est immédiat et dure 5 à 6 heures. L'héparine n'est pas une médication au long cours. **On l'utilise dans** la *maladie thromboembolique*, l'hémodialyse, les coagulopathies de consommation (ztoutou).
149. **On l'utilise** [isoprénaline] **dans** les *états de choc post-opératoires*, et au cours de défaillances aiguës: asphyxie, intoxication, prématuré anoxique... (zcarraz).

Les exemples 147, 148 et 149 montrent les 3 relations de finalité relevées pour le TS [STATE]. Ils montrent tous trois que le patron utilisé et qui exprime une relation de finalité est **on utilise Y dans X**, où Y représente un médicament (l'élément associé à la fonction), et X une maladie (le lieu d'intervention du Y).

Le Tableau 59 Tableau 59 reprend la répartition des relations sémantiques pour le TS [STATE], cette fois-ci par terme clé.

Tableau 59. Répartition des relations sémantiques par terme clé pour le TS [STATE]

	RELATION CAUSALE	PATRON MÉTALIN-GUISTIQUE	RELATION DE FINALITÉ	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Maladie (médecine)	21 (20 création, 1 diminution)	3	2	18	0
État (médecine)	12 (4 création, 6 modification et 2 augmentation)	0	1	0	0
Bogue (informatique)	4 (4 création)	0	0	0	7 (7 inclassables)
État (informatique)	1 (1 modification)	3	0	0	0

Comme le montre le Tableau 59, c'est pour le terme *maladie* que le plus grand nombre de contextes a été relevé. Ce terme est lié à un grand nombre de relations hyperonymiques, car les 18 relations hyperonymiques relevées sont liées à ce terme. De plus, les relations causales expriment presque exclusivement la création. Enfin, toutes les relations sont représentées pour ce terme à l'exception de la relation méronymique.

Le terme *état* (médecine) est le deuxième terme le plus souvent associé à un patron de relation sémantique. Deux des cinq relations sont représentées pour ce terme, mais seule la relation causale est fréquente : elle est exprimée dans douze contextes (contre un seul pour la relation de finalité).

Quant au terme *bogue* (informatique), seules les relations causale et méronymique sont représentées pour ce terme. Les relations causales expriment toutes la création. En outre, aucune des relations méronymiques n'a pu être classée selon le modèle de Winston *et al.* (1987). Enfin, le terme *état* (informatique) apparaît dans très peu de contextes. Les relations sémantiques auxquelles il est associé sont les relations causale et les patrons métalinguistiques.

4.8.1 Répartition des relations sémantiques par corpus

Étant donné que seulement 15 contextes sur 72 (21 %) proviennent du corpus d'informatique, contre 57 contextes (79 %) pour le corpus de médecine, nous examinerons plus en profondeur les résultats obtenus pour chacun des corpus. Les Tableaux 60 et 61 présentent, par corpus, la répartition des relations sémantiques observées par ordre décroissant.

Tableau 60. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [STATE]

Relation	Nombre/Total	Pourcentage
Relation causale	33/57	58 %
Relation hyperonymique	18/57	32 %
Patron métalinguistique	3/57	5 %
Relation de finalité	3/57	5 %
Relation méronymique	0/57	0 %

Tableau 61. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [STATE]

Relation	Nombre/Total	Pourcentage
Relation méronymique	7/15	47 %
Relation causale	5/15	33 %
Patron métalinguistique	3/15	20 %
Relation hyperonymique	0/15	0 %
Relation de finalité	0/15	0 %

Comme le montrent les Tableaux 60 et 61, la relation causale est la principale relation dans le corpus de médecine, puisqu'elle est exprimée dans 58 % des contextes, alors que c'est la relation méronymique qui est la plus importante dans le corpus d'informatique : elle est exprimée dans 47 % des contextes étudiés.

La relation causale compte tout de même pour 33 % dans le corpus d'informatique. Il se pourrait que le contraste entre les deux corpus en ce qui a trait à la relation méronymique (absente du corpus de médecine et principale relation en informatique) ait un lien avec le fait qu'aucun des contextes n'ait pu être classés selon l'un des six types du modèle de Winston *et al.* (1987) ou encore, que cette relation soit associée au terme *bogue*.

Il y a également déséquilibre entre les deux corpus en ce qui a trait à la relation hyperonymique puisque, dans le corpus de médecine, elle apparaît dans 32 % des contextes, alors que dans le corpus d'informatique, elle est complètement absente. Il semble donc que cette relation soit davantage liée au corpus de médecine. Enfin, en ce qui a trait à la relation de finalité, on n'en trouve aucune trace dans le corpus d'informatique, alors qu'elle n'apparaît que dans trois contextes (5 %) dans le corpus de médecine.

4.8.2 Répartition des PRS pour le TS [STATE]

Cette sous-section examine la forme des patrons relevés pour le TS [STATE] dans chacun des corpus et dans les deux corpus. Le Tableau 62 montre la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 62. Répartition des PRS par corpus

Relation	Médecine	Informatique	Médecine et Informatique
Causale	Résultat de X Y, Y cause X (cause de Y X (3x), cause de X Y), Y consécutif à X, X donne Y, X entraîne Y (3x) (Y entraîne X, Y n'entraîne pas X), X lié à Y (2x) (Y lié à X (2x)), responsable de X Y (2x) (responsable de Y X), générateur de X Y, secondaire à X Y (2x), Y précipite X, X favorise Y, origine de X Y, X entre en ligne de compte pour Y, Y influe sur X, Y aggrave X, Y modifie X, en fonction de X Y, Y dépend de X	X vient du fait Y, Y apporte X, Y provoque X, Y déterminé par X	X dû à Y (1x, 1x)
Hyperonymique	X il s'agit de Y, X en particulier Y (Y en particulier X), type de X Y, autre Y X, X comme Y	*	<u>Y est X (8x, 1x)</u> <u>(X est Y (1x, 0x),</u> <u>Y c'est X (1x,</u> <u>0x), X c'est Y</u> <u>(1x, 0x))</u>
Méronymique	*	Y contient X, Y comporte X (5x), Y n'est pas exempt de X	*
Patron métalinguistique	X ou Y (Y ou X (2x))	X c'est-à-dire Y (Y c'est-à-dire X) (2x)	*
Finalité	On utilise Y dans X (3x)	*	*

Comme le montre le Tableau 62, la majorité des PRS sont propres au domaine de l'informatique ou de la médecine. En effet, malgré le grand nombre de patrons observés pour ce TS, soit 35 patrons, seulement 2 (6 %) d'entre eux sont partagés par les deux corpus. Par ailleurs, 94 % des patrons (33/35) sont propres à l'informatique ou à la médecine.

4.8.3 *Remarques générales sur le TS [STATE]*

Pour conclure, en ce qui concerne le TS [STATE], la relation causale est la plus importante dans le corpus de médecine. En effet, elle est exprimée dans 58 % des contextes étudiés. En informatique, c'est la méronymie qui est la relation la plus importante. En effet, elle apparaît dans 47 % des contextes étudiés. Cependant, dans les deux corpus, la relation de finalité est l'une des moins importantes puisqu'aucune n'a été relevée dans le corpus d'informatique, et qu'elle n'apparaît que dans 5 % des contextes extraits du corpus de médecine.

Le patron le plus fréquent pour l'ensemble du TS [STATE] est le patron hyperonymique **X est Y**, qui regroupe les formes : **X est Y**, **Y est X (9x)**, **Y c'est X**, **X c'est Y**. Ce patron est présent dans 67 % (12 occurrences sur 18) pour l'ensemble des patrons métalinguistiques, et dans 17 % (12 occurrences sur 72) des contextes étudiés pour l'ensemble des relations de ce TS. Ce patron n'est présent que dans le corpus de médecine.

La majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, seulement 6 % des patrons sont partagés par les deux corpus alors que 94 % des patrons sont exclusifs de l'informatique ou de la médecine.

4.9 **TS : [ABSTRACTION]**

La présente section ainsi que la section 4.10 sont consacrées à des TS pour lesquels nous n'avons pas étudié deux termes par corpus. Rappelons que les

difficultés liées à la sélection des termes pour chacun des TS sont abordées à la section 3.5.1. Bien que nous présentions des résultats, nous ne pouvons généraliser à partir des données obtenues pour ces deux TS.

Nous avons retenu au total 235 contextes pour le TS [ABSTRACTION]. En médecine, nous n'avons pu sélectionner qu'un seul terme, soit *électrocardiogramme* (incluant *ECG*), et en informatique, les termes *langage* et *logiciel*. Le Tableau 63 présente, par domaine, les résultats obtenus pour le TS [ABSTRACTION] : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés par corpus.

Le Tableau 64 présente la répartition dans les deux corpus des 5 relations sémantiques par ordre décroissant.

Tableau 63. Répartition des PRS pour le TS [ABSTRACTION]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
FINALITÉ	86	15	<u>X permet Y (8x), X ne permet pas Y, Y repose sur X (4x), X utile pour Y, clef de Y X</u>	Médecine
		71	<u>X permet Y (43x), X pour Y (4x), sans l'aide de X Y, à l'aide de X Y (4x), au moyen de X Y (2x), but de X Y, objectif de X Y, X sert à Y, X sert de Y, on confie Y à X, on utilise X pour Y (3x), si vous souhaitez Y il faut X, X destiné à Y (3x), X conçu pour Y (2x), X assure Y, X peut Y, X contribue à Y</u>	Informatique
PATRON MÉTALINGUISTIQUE	21	21	Terme X s'oppose à Y, on appelle X Y (2x), on appelle Y X, appellation de X Y, X appelé Y (3x), X signifie Y, on nomme X Y (2x), X nommé Y, on parle de X plutôt que de Y, X du mot Y, terme X fait référence à Y, terme X sens de Y, X désigne Y, X c'est-à-dire Y (3x), Y ou X	Informatique
HYPERONYMIQUE	93	93	Y est X (15x), X est Y (16x), Y c'est X (4x), X c'est Y (3x), Y d'autres X (2x), Y autre X (2x), Y il s'agit de X (3x), X il s'agit de Y (3x), X se classe comme Y, Y constitue X, exemple de X Y (3x), X comme par exemple Y (2x), X parmi Y, parmi X on compte Y, X comme Y (22x), X du genre Y, catégorie de X Y (3x), X tel Y (6x), type de X Y (2x), le plus connu de X Y (2x)	Informatique

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
MÉRONYMIQUE	32	32	X contient Y (2x), Y contient X (4x), Y extrait de X, X inclut Y, ensemble de Y forme X, X se trouve dans Y, X se trouve sur Y, sous-ensemble de X Y (3x), X constitue Y, X comporte Y (4x), composante de Y X, X dispose de Y, Y dispose de X (2x), Y comprend X, X comprend Y (2x), Y dispose d'un ensemble de X, X constitue un ensemble de Y, X consiste en Y, Y possède X, X possède Y, Y regroupe X	Informatique
CAUSALE	3	2	Y modifie X, effet de Y sur X	Médecine
		1	Responsable de Y X	Informatique

Tableau 64. Répartition des relations sémantiques par ordre décroissant pour le TS [ABSTRACTION]

Relation	Nombre/Total	Pourcentage
Relation hyperonymique	93/235	40 %
Relation de finalité	86/235	37 %
Relation méronymique	32/235	14 %
Patron métalinguistique	21/235	9 %
Relation causale	3/235	1 %

Toutes les relations sont représentées pour le TS [ABSTRACTION]. La relation hyperonymique est la plus importante. Cependant, elle est suivie de très près par la relation de finalité. La différence est peu marquée entre ces deux relations puisque l'écart les séparant est de 3 %. En effet, la relation hyperonymique, qui n'est observée que dans notre corpus d'informatique, apparaît dans 40 % des contextes étudiés. La relation de finalité, quant à elle, observée dans nos corpus de médecine et d'informatique, est exprimée dans 37 % des contextes étudiés. Nous avons reproduit un exemple de la relation hyperonymique en (150).

150. Ces *logiciels* sont de trois **types** : des logiciels constructeurs, [...] des progiciels [...], et des logiciels développés pour les besoins spécifiques de l'entreprise [...] (genilo).

Nous avons reproduit des exemples (151 et 152) de la relation de finalité tirés respectivement de nos corpus de médecine et d'informatique.

151. Dans tous les autres cas, seul l'*ECG endocavitaire* **permet** le diagnostic en démontrant l'existence d'un trouble de conduction significatif à l'étage tronculaire (potentiel hisien déformé et élargi, ou mieux dédoublé) ou infra-hisien (HV allongé) (troub3).
152. Si les standards d'Internet permettent d'uniformiser les réseaux ainsi que les services offerts dans ces réseaux, les *langages de programmation*, eux, **servent de** matériau de construction à l'autoroute de l'information (ceveil).

Le PRS le plus important est le patron de finalité **X permet Y**, qui regroupe les formes suivantes : **X permet Y (51x)** et **X ne permet pas Y (1x)**. Il est présent dans 60 % des contextes exprimant une relation de finalité (52 patrons sur 86) pour cette relation, et dans 22 % des contextes étudiés (52 patrons sur 235) pour ce TS. Ce patron est présent dans les deux corpus.

Pour le TS [ABSTRACTION], on trouve aussi 9 % de patrons métalinguistiques. Le contexte (153), tiré de notre corpus d'informatique, en présente un exemple.

153. Il a donc fallu mettre au point des liaisons physiques entre les ordinateurs pour que l'information puisse circuler, mais aussi *un langage de communication* pour qu'il puisse y avoir un réel échange, **on** a décidé de **nommer** ce langage : protocole (comint).

Le TS [ABSTRACTION] est également associé à des relations méronymiques dans 14 % des contextes analysés. Les relations méronymiques se répartissent de la manière suivante : partie fonctionnelle – tout, élément – ensemble et lieu – zone. Elles proviennent exclusivement de notre corpus d'informatique. Les contextes 154, 155 et 156 illustrent les relations méronymiques relevées.

Partie fonctionnelle – tout :

154. Les matériels et les *logiciels* livrés aujourd'hui **comportent** en effet un serveur Web permettant d'interroger l'application (reseauip).

Élément – ensemble :

155. Le *logiciel* qui **consiste en** un ensemble de un ou plusieurs programmes développés par l'humain permettant d'exécuter des tâches spécifiques. Par exemple, un logiciel comptable qui transforme des données provenant de transactions comptables comme les ventes en de l'information financière (zmspc1).

Lieu – zone :

156. Si l'installation ne se lance pas toute seule, il vous faut le faire vous-même : - soit en allant dans le poste de travail, sur votre lecteur **contenant** le *logiciel* et cliquer sur Install.exe ou Setup.exe. - soit à partir de la fonction exécuter du menu démarrer et faire, dans le cas où votre logiciel se trouve sur un CD-ROM... (window).

Enfin, les relations causales relevées expriment la modification (157 et 158) et l'augmentation (159). Celles qui expriment la modification proviennent de notre corpus de médecine, alors que celle qui exprime l'augmentation provient de notre corpus d'informatique.

157. La quinidine **modifie** l'*électrocardiogramme*, allongeant les espaces PR, QRS et surtout QT, l'espace ST restant normal (zschmit).

158. Steiness et Olesen (1976) ont étudié, chez des patients en insuffisance cardiaque chronique, les **effets** sur l'*ECG* **d'**un traitement prolongé digitalodiurétique (digoxine + furosémide ou bumétanide) sans apports potassiques. 50 p. cent des sujets ont présenté des arythmies et la perte de potassium était identique chez les sujets avec ou sans trouble du rythme (xgiroud).

159. *Ce langage* de haut niveau [HTML] [...] est le principal **responsable de** la croissance d'Internet, autant pour le nombre d'utilisateurs que pour le nombre d'applications (ceveil).

Le Tableau 65 reprend la répartition des relations sémantiques pour le TS [ABSTRACTION], cette fois-ci par terme clé.

**Tableau 65. Répartition des relations sémantiques par terme clé pour le TS
[ABSTRACTION]**

	RELATION CAUSALE	PATRON MÉTALIN- GUISTIQUE	RELATION DE FINALITÉ	RELATION HYPERO- NYMIQUE	RELATION MÉRONYMIQU E.
Électrocardio- gramme (ECG) (médecine)	2 (2 modification)	0	15	0	0
Langage (informatique)	1	9	24	42	9 (6 partie fonctionnelle – tout, 3 élément – ensemble)
Logiciel (informatique)	0	12	47	51	23 (12 partie fonctionnelle – tout et 6 élément – ensemble, 5 lieu – zone)

Les relations sémantiques sont plus nombreuses pour les termes *langage* et *logiciel* que pour le terme *électrocardiogramme (ECG)*. Les relations sémantiques sont peu variées pour *électrocardiogramme (ECG)* (causale et finalité), alors que les termes *logiciel* et *langage*, sont, quant à eux, associés à quatre et cinq relations respectivement. De plus, les relations méronymiques de ces deux termes d'informatique, qui sont principalement de nature partie fonctionnelle – tout, sont assez diversifiées.

4.9.1 *Répartition des relations sémantiques par corpus*

Seulement 17 contextes sur 235 (7 %) proviennent du corpus de médecine, contre 218 contextes (93 %) pour le corpus d'informatique. Ce déséquilibre est normal puisqu'un seul terme est choisi pour la médecine (contre deux termes pour l'informatique). Les Tableaux 66 et 67 présentent, par corpus, la répartition des relations sémantiques observées par ordre décroissant.

Tableau 66. Répartition des relations sémantiques par ordre décroissant pour le corpus de médecine pour le TS [ABSTRACTION]

Relation	Nombre/Total	Pourcentage
Relation de finalité	15/17	88 %
Relation causale	2/17	12 %
Relation hyperonymique	0/17	0 %
Relation méronymique	0/17	0 %
Patron métalinguistique	0/17	0 %

Tableau 67. Répartition des relations sémantiques par ordre décroissant pour le corpus d'informatique pour le TS [ABSTRACTION]

Relation	Nombre/Total	Pourcentage
Relation hyperonymique	93/218	43 %
Relation de finalité	71/218	33 %
Relation méronymique	32/218	15 %
Patron métalinguistique	21/218	10 %
Relation causale	1/218	0 %

Comme le montrent les Tableaux 66 et 67, la relation de finalité est la principale relation observée dans le corpus de médecine (88 %), alors que c'est la relation hyperonymique qui est la plus importante dans le corpus d'informatique (43 %). La relation de finalité apparaît tout de même dans 33 % des contextes extraits du corpus d'informatique.

On note également que la relation causale est peu importante dans le corpus de médecine (12 %) et qu'elle est quasiment absente du corpus d'informatique pour ce TS.

Par ailleurs, l'hyperonymie ne se présente pas dans les mêmes proportions d'un corpus à l'autre. En effet, le corpus de médecine ne contient aucune relation hyperonymique, alors que le corpus d'informatique en contient 43 %. Enfin, les relations méronymiques et les PM sont absents du corpus de médecine, alors qu'ils représentent respectivement 15 % et 10 % pour le corpus d'informatique. Ces écarts entre les corpus sont probablement attribuables à la taille des corpus ainsi qu'aux caractéristiques distinctes des termes, d'autant plus que pour ce TS, un seul terme de médecine contre deux termes en informatique ont pu être étudiés.

4.9.2 *Répartition des PRS pour le TS* [ABSTRACTION]

Cette sous-section examine la forme des patrons relevés pour le TS [ABSTRACTION] dans chacun des corpus et dans les deux corpus. Le Tableau 68 montre la répartition des patrons selon les corpus, et lesquels sont présents à la fois dans le corpus de médecine et le corpus d'informatique.

Tableau 68. Répartition des PRS par corpus pour le TS [ABSTRACTION]

Relation	Médecine	Informatique	Médecine et Informatique
Causale	Y modifie X, effet de Y sur X	Responsable de Y X	*
Hyperonymique	*	Y est X (16x) (X est Y (15x), Y c'est X (4x), X c'est Y (3x)), Y d'autres X (2x) (Y autre X (2x)), Y il s'agit de X (3x) (X il s'agit de Y (3x)), X se classe comme Y, Y constitue X, exemple de X Y (3x) (X comme par exemple Y (2x)), X parmi Y (parmi X on compte Y), X comme Y (22x), X du genre Y, catégorie de X Y (3x), X tel Y (6x), type de X Y (2x), le plus connu de X Y (2x)	*
Méronymique	*	X contient Y (2x) (Y contient X (4x)), Y extrait de X, X inclut Y, ensemble de Y forme X, X se trouve dans Y (X se trouve sur Y), sous-ensemble de X Y (3x), X constitue Y, X comporte Y (4x), composante de Y X, X dispose de Y (Y dispose de X (2x)), Y comprend X (X comprend Y (2x)), Y dispose d'un ensemble de X, X constitue un ensemble de Y, X consiste en Y, Y possède X (X possède Y), Y regroupe X	*
Patron métalinguistique	*	On appelle X Y (2x) (on appelle Y X, appellation de X Y, X appelé Y (3x)), X signifie Y, on nomme X Y (2x) (X nommé Y), on parle de X plutôt que de Y, X du mot Y, terme X s'oppose à Y (terme X fait référence à Y, terme X sens de Y), X désigne Y, X c'est-à-dire Y (3x), Y ou X	*
Finalité	Y repose sur X (4x), clef de Y X	X pour Y (4x), à l'aide de X Y (4x) (sans l'aide de X Y), au moyen de X Y (2x), but de X Y, objectif de X Y, X sert à Y (X sert de Y), on confie Y à X, si vous souhaitez Y il faut X, X destiné à Y (3x), X conçu pour Y (2x), X assure Y, X peut Y, X contribue à Y, X développé pour Y	<u>X permet Y (8x, 43x), (X ne permet pas Y(1x, 0x)), X utile pour Y (1x, 0x) (on utilise X pour Y (0x, 2x))</u>

Comme le montre le Tableau 68, la majorité des PRS sont propres au domaine de l'informatique ou de la médecine. En effet, seulement 2 patrons (3 %) (exprimant une relation de finalité) sont partagés par les deux corpus. Par ailleurs, 57 (97 %) patrons sont propres à l'informatique ou à la médecine.

4.9.3 *Remarques générales sur le TS [ABSTRACTION]*

En conclusion, étant donné que nous n'avons retenu qu'un seul terme pour le corpus de médecine contre deux termes pour le corpus d'informatique, nous ne pourrions faire des analyses aussi poussées que pour les autres TS. La recherche sur les termes *langage* et *logiciel* dans notre corpus d'informatique nous a permis de recueillir 218 contextes (133 contextes pour le terme *logiciel* et 85 contextes pour le terme *langage*), contre 17 contextes seulement pour le terme *électrocardiogramme* (ECG) provenant de notre corpus de médecine.

Notre analyse nous a permis de noter que la relation hyperonymique (40 %) est la relation la plus fréquemment observée pour ce type sémantique. Elle est suivie de très près par la relation de finalité (37 %) puisque seulement 3 % d'écart sépare les deux relations. Tout comme pour les TS [ENTITY OBJECT] et [STATE], la relation la plus importante diffère d'un corpus à l'autre, puisque, dans le corpus de médecine, c'est la relation de finalité qui est la plus importante alors que dans le corpus d'informatique, c'est la relation hyperonymique qui est la mieux représentée.

Le PRS le plus important est le patron de finalité **X permet Y**, qui regroupe les formes suivantes : **X permet Y (51x)** et **X ne permet pas Y (1x)**. Il est présent dans 60 % des contextes exprimant une relation de finalité (52 patrons sur 86) pour cette relation, et dans 22 % des contextes étudiés (52 patrons sur 235) pour l'ensemble du TS [ABSTRACTION]. Ces patrons sont présents dans nos deux corpus.

La majorité des PRS ne sont observés que dans un corpus (informatique ou médecine). En effet, seulement 2 patrons sur un total de 59 patrons (3 %) sont partagés par les deux corpus alors que les 57 autres patrons (97 %) sont exclusifs de l'informatique ou de la médecine.

4.10 TS : [PSYCHOLOGICAL FEATURE]

Comme dans la section 4.9, la présente section est consacrée à un TS pour lequel nous n'avons pas étudié deux termes par corpus. Bien que nous présentions les résultats, nous ne pouvons généraliser à partir des données obtenues pour ces deux TS.

Nous avons retenu, au total, 56 contextes pour le TS [PSYCHOLOGICAL FEATURE]. Nous avons sélectionné les termes *douleur* et *symptôme* en médecine, mais aucun terme en informatique. Le Tableau 69 présente les résultats obtenus pour ce TS : les relations sémantiques accompagnées des PRS qui leur sont associés ainsi que le nombre de contextes relevés dans le corpus de médecine.

Le Tableau 70 présente la répartition des 5 relations sémantiques par ordre décroissant.

Tableau 69. Répartition des PRS pour le TS [PSYCHOLOGICAL FEATURE]

RELATION	NOMBRE TOTAL	NOMBRE PAR CORPUS	PATRON	DOMAINE
CAUSALE	29	29	Y diminue X, facteur de X Y (2x), Y générateur de X, Y provoque X (4x), Y entraîne X (3x), X fait évoquer Y, cause de X Y, Y cause X (2x), Y ne cause pas X (2x), à l'origine de X Y, Y provoque X, responsable de X Y (2x), Y déclenche X, on met X sur le compte de Y, X fait suspecter Y, X dû à Y, X favorise Y, X s'observe dans Y, X fait craindre Y, X fait penser à Y	Médecine
PATRON MÉTALINGUISTIQUE	1	1	Y c'est-à-dire sans X	Médecine
HYPERONYMIQUE	25	25	<u>Y est X (5x), X est Y (7x), X qu'est Y, type de X Y (8x), divers X Y, d'autres X Y (2x), X d'allure Y</u>	Médecine
MÉRONYMIQUE	1	1	Y comprend X	Médecine

Tableau 70. Répartition des relations sémantiques par ordre décroissant pour le TS [PSYCHOLOGICAL FEATURE]

Relation	Nombre/Total	Pourcentage
Relation causale	29/56	52 %
Relation hyperonymique	25/56	45 %
Patron métalinguistique	1/56	2 %
Relation méronymique	1/56	2 %

La relation causale est la relation sémantique la plus importante pour le TS [PSYCHOLOGICAL FEATURE]. Elle est présente dans 52 % des cas pour ce TS. Elle se subdivise de la manière suivante : création, diminution et conservation. Les contextes 160, 161 et 162 montrent un exemple de chacune des subdivisions.

Création :

160. Les kystes du cholédoque sont une **cause de** *douleur abdominale chronique* (doul).

Diminution :

161. ... (alors que la *douleur est diminuée par* le repos, les sédatifs, les anti-inflammatoires) (pericar)...

Conservation :

162. Généralement, ces changements dans la rétine **ne causent aucun** *symptôme* (ret_diab).

Même si c'est la relation causale qui est la plus importante, le PRS le plus fréquent pour le TS [PSYCHOLOGICAL FEATURE] est un patron exprimant l'hyponymie : **X est Y**, qui regroupe les formes suivantes : **X est Y (7x)**, **Y est X (5x)**, **X qu'est Y**. Il est présent dans 23 % des contextes étudiés (13 occurrences sur 56) pour ce TS, et dans 52 % des contextes étudiés exprimant une relation hyperonymique (13 occurrences sur 25). Le contexte 163 en présente un exemple.

163. Le diagnostic avec les autres causes de rougeur oculaire est facile si l'on considère le *symptôme* capital **qu'est** la mydriase (opht124).

Outre la relation causale, le TS [PSYCHOLOGICAL FEATURE] est également associé à un nombre important de relations hyperonymiques (45 %). Voici un exemple tiré de notre corpus de médecine. Nous avons reproduit un exemple en 164.

164. Les uvéites antérieures. Signes fonctionnels. *Douleur sourde, profonde*, à **type de** lourdeur péri-orbitaire (opht118).

Pour le TS [PSYCHOLOGICAL FEATURE], on compte aussi un patron métalinguistique. L'exemple 165 montre le contexte comprenant ce PM.

165. L'ischémie myocardique silencieuse (IMS) On entend sous ce terme, toutes les insuffisances coronariennes qui comportent des épisodes ischémiques asymptomatiques, **c'est-à-dire** sans *douleur angineuse simultanée* (imt).

Enfin, pour le TS [PSYCHOLOGICAL FEATURE], on compte une relation méronymique partie fonctionnelle – tout que présente l'exemple 166. Nous sommes cependant consciente que son classement parmi les relations de nature partie fonctionnelle – tout est un cas limite puisque le méronyme *symptômes neurologiques* et l'holonyme tableau clinique sont des éléments abstraits.

166. Le tableau clinique **comprend** : fièvre, douleur rachidienne et *symptômes neurologiques* (rach).

Le Tableau 71 reprend la répartition des relations sémantiques pour le TS [PSYCHOLOGICAL FEATURE], cette fois-ci par terme clé.

Tableau 71. Répartition des relations sémantiques par terme clé pour le TS [PSYCHOLOGICAL FEATURE]

	RELATION CAUSALE	PATRON MÉTALINGUISTIQUE	RELATION HYPERONYMIQUE	RELATION MÉRONYMIQUE
Symptôme (médecine)	7 (2 conservation, 5 création)	0	14	1 (partie fonctionnelle – tout)
Douleur (médecine)	22 (21 création, 1 diminution)	1	11	0

Comme le montre le Tableau 71, le terme *douleur* est nettement plus fréquemment associé à un PRS que *symptôme*. Toutes les relations sémantiques représentées pour ce TS sont représentées pour ce terme à l'exception de la relation méronymique. La relation causale exprimant une création est particulièrement fréquente pour ce terme puisqu'on la retrouve dans 21 contextes sur 22 (95 %).

Quant au terme *symptôme*, il est associé à toutes les relations sémantiques de ce TS à l'exception des patrons métalinguistiques. La relation la plus importante pour ce terme est la relation hyperonymique, qu'on retrouve dans 14 contextes sur 25 (56 %). Encore une fois, la proportion de relations causales exprimant une création est assez élevée, puisqu'on les retrouve dans 5 contextes sur 7.

4.10.1 *Remarques générales sur le TS [PSYCHOLOGICAL FEATURE]*

En conclusion, en ce qui concerne le TS [PSYCHOLOGICAL FEATURE], la relation causale est la plus importante. En effet, elle est exprimée dans 52 % des contextes étudiés. La deuxième relation la plus importante est l'hyponymie, qui est exprimée dans 45 % des contextes étudiés. En contrepartie, la relation de finalité est complètement absente pour ce TS.

Le PRS le plus important pour le TS [PSYCHOLOGICAL FEATURE] est le patron hyperonymique **X est Y**, qui regroupe les formes suivantes : **X est Y (7x)**, **Y est X (5x)**, **X qu'est Y**. Ce patron est présent dans 23 % des cas (13 occurrences sur 56) pour l'ensemble de ce TS, et dans 52 % des cas pour la relation hyperonymique (13 occurrences sur 25).

4.11 Synthèse des résultats

Cette section présente les principales constatations découlant de nos analyses sur chacun des types sémantiques. Nous y présentons les relations les plus importantes observées pour chacun des types sémantiques. Nous rappelons également les relations sémantiques qui ne sont jamais associées à des TS. Enfin, nous reproduisons les PRS les plus fréquents pour chaque relation sémantique.

Le Tableau 72 présente les relations les plus importantes observées pour chacun des types sémantiques.

Tableau 72. Relation la plus importante selon le TS

TS	RELATION LA PLUS IMPORTANTE
[ENTITY OBJECT]	Relation hyperonymique dans 34 % des cas
[EVENT]	Relation causale dans 83 % des cas
[ACTIVITY]	Relation finalité dans 71 % des cas
[PSYCHOLOGICAL FEATURE (FIELD)]	Relation de finalité dans 81 % des cas
[PSYCHOLOGICAL FEATURE]	Relation causale dans 52 % des cas
[ENTITY ORGANISM]	Relation de finalité dans 50 % des cas
[PHENOMENON]	Relation causale dans 74 % des cas
[STATE]	Relation causale dans 53 % des cas
[ABSTRACTION]	Relation hyperonymique dans 40 % des cas

Le Tableau 72 montre que, pour 4 des 9 TS, la relation causale est la plus importante, en particulier pour le TS [EVENT] où l'on constate qu'elle est présente dans 83 % des cas. Par ailleurs, pour 3 autres TS, la relation de finalité est celle qui apparaît le plus souvent, notamment pour le TS [PSYCHOLOGICAL FEATURE (FIELD)] où on la rencontre dans 81 % des cas. En terminologie, ces deux relations sont souvent définies comme étant les relations centrales, et nous supposons qu'elles seraient déterminantes dans notre étude. Toutefois, il est étonnant que l'hyponymie ou la méronymie ne soient presque jamais la relation la plus fréquemment observée pour un type sémantique. L'hyponymie est la relation sémantique la mieux représentée pour les TS [ENTITY OBJECT] et [ABSTRACTION] uniquement.

Pour certains types sémantiques cependant, la relation sémantique la plus importante n'est pas forcément la même dans les deux corpus comme nous l'avons vu plus haut. C'est le cas du TS [ENTITY OBJECT], pour lequel la relation la plus importante en médecine est la relation de finalité, alors qu'en informatique, c'est la relation hyperonymique qui se dégage (la relation de finalité arrive en troisième position dans ce corpus). De même, en ce qui concerne le TS [STATE], la relation causale est la relation la plus importante dans le corpus de médecine, mais en informatique, c'est la relation méronymique qui est la mieux représentée (la relation causale arrive en deuxième position dans ce corpus). Enfin, pour le TS [ABSTRACTION], la relation la plus fréquemment observée dans le corpus de médecine est la relation de finalité, alors que le corpus de médecine accorde une plus grande place à l'hyponymie (la relation de finalité arrive en deuxième position dans ce corpus).

Les termes appartenant aux TS [EVENT] et [PSYCHOLOGICAL FEATURE (FIELD)] sont très fortement associés aux relations causale et de finalité respectivement. En effet, dans le cas du TS [EVENT], la relation causale se démarque vraiment des autres, tant dans le corpus de médecine que dans le corpus d'informatique puisque 73 % d'écart séparent la relation la plus importante, la

relation causale en l'occurrence, de la relation arrivant en deuxième position et ce, tant dans le corpus de médecine que dans le corpus d'informatique. En ce qui a trait au TS [PSYCHOLOGICAL FEATURE (FIELD)], la relation de finalité est exclusive au corpus de médecine, et fortement représentée dans le corpus d'informatique puisque 58 % d'écart séparent la relation la plus importante, la relation de finalité en l'occurrence, de la deuxième relation.

Quant aux termes appartenant aux TS [ENTITY OBJECT] et [ABSTRACTION], la relation la plus importante à laquelle ils sont liés se démarque peu des autres relations. En effet, pour le TS [ENTITY OBJECT], seulement 7 % d'écart sépare la relation la plus importante de la troisième. De même pour le TS [ABSTRACTION], seulement 3 % d'écart sépare la relation la plus importante soit l'hyponymie, de la deuxième relation, soit la relation de finalité.

Enfin, rappelons que, comme nous l'avons vu à la section 4.1, les types sémantiques peu fréquemment associés à des relations sémantiques sont [PSYCHOLOGICAL FEATURE (FIELD)] et [ENTITY ORGANISM].

Le Tableau 73 fait un rappel des relations sémantiques qui ne sont jamais associées à certains TS dans les deux corpus étudiés.

Tableau 73. Relations n'étant jamais associées à certains TS

TS	RELATIONS
[ENTITY OBJECT]	Relation causale
[EVENT]	Relation de finalité
[PSYCHOLOGICAL FEATURE (FIELD)]	Relation causale et patron métalinguistique
[PSYCHOLOGICAL FEATURE]	Relation de finalité
[PHENOMENON]	Patron métalinguistique, relation méronymique

Le Tableau 73 montre que la relation causale n'est jamais associée au TS [ENTITY OBJECT] ni au TS [PSYCHOLOGICAL FEATURE (FIELD)]. La relation de finalité, quant à elle, n'est jamais associée aux TS [EVENT] et [PSYCHOLOGICAL FEATURE]. Aucune relation méronymique n'est associée au TS [PHENOMENON]. De même, aucun patron métalinguistique n'est associé aux TS [PSYCHOLOGICAL FEATURE (FIELD)] ou [PHENOMENON].

À l'exception du TS [ENTITY OBJECT], qui compte 146 contextes, les TS auxquels une ou plusieurs relations ne sont jamais associées comptent un nombre de contextes nettement moins élevé (soit 38) que la moyenne globale, qui est de 79 contextes pour l'ensemble des neuf types sémantiques. Lorsqu'on inclut le TS [ENTITY OBJECT], le nombre de contextes moyen par TS pour les cinq types sémantiques indiqués dans le Tableau 73 est de 60.

Quant aux autres TS ([ACTIVITY], [ENTITY ORGANISM], [STATE] et [ABSTRACTION]), ils sont tous associés aux cinq relations étudiées (causale, de finalité, hyperonymique, méronymique et patrons métalinguistiques). Pour deux de ces TS, soit [ACTIVITY] et [ABSTRACTION], le nombre de contextes enregistrés (87 et 235 respectivement) est supérieur à la moyenne, bien que pour le TS [ENTITY ORGANISM], le nombre de contextes enregistrés ne soit que de 16. Enfin, pour l'ensemble de ces quatre TS, le nombre de contextes moyen est égal à 103 contextes. Ce nombre est presque deux fois plus élevé que la moyenne obtenue pour les TS auxquels moins de cinq relations sont associées.

Dans le 0, nous reproduisons les PRS les plus importants observés pour chaque relation sémantique. Pour chacune des relations, nous indiquons le nombre total de contextes retenus et leur répartition par corpus, ainsi que le patron le plus important dans chaque corpus.

Tableau 74. PRS le plus important pour une relation sémantique

TS	PATRON LE PLUS IMPORTANT (pour une relation sémantique donnée)
	Relation hyperonymique
[ENTITY OBJECT]	X est Y dans 28 % des contextes – hyperonymique
[PSYCHOLOGICAL FEATURE]	X est Y dans 52 % des contextes – hyperonymique
[STATE]	X est Y dans 67 % des contextes – hyperonymique
	Relation causale
[EVENT]	X cause Y dans 16 % des contextes – causale
[PHENOMENON]	X entraîne Y dans 18 % des contextes – causale
	Relation de finalité
[ACTIVITY]	On utilise X pour Y dans 23 % des contextes – finalité et X permet Y dans 24 % des contextes – finalité
[PSYCHOLOGICAL FEATURE (FIELD)]	Objectif de X Y dans 31 % des contextes – finalité
[ABSTRACTION]	X permet Y dans 60 % des contextes – finalité
[ENTITY ORGANISM]	Rôle de X Y dans 25 % des contextes – finalité (égalité avec la relation méronymique)
	Relation méronymique
[ENTITY ORGANISM]	X fait partie de Y dans 25 % des contextes – méronymique (égalité avec la relation méronymique)

Comme le montre le 0, le patron **X est Y** est le patron qui est le plus productif lorsqu'il est associé au TS [STATE]. Il exprime en effet une relation hyperonymique dans 67 % des contextes enregistrés pour cette relation. Ce patron est cependant moins productif lorsqu'il est associé au TS [PSYCHOLOGICAL FEATURE] (52 %), et encore moins productif pour le TS [ENTITY OBJECT] (28 %). Quant au patron **X permet Y**, il est plus productif lorsqu'il est associé au TS [ABSTRACTION]. Il représente en effet 60 % des contextes enregistrés pour la relation de finalité, alors que pour le TS [ACTIVITY], il n'apparaît que dans 24 % des contextes.

5. CONCLUSION

La présente thèse visait à élaborer un modèle d'association des patrons de relations sémantiques (PRS) aux types sémantiques (TS) auxquels appartiennent les termes. Notre objectif principal consistait à vérifier si ce modèle pouvait transcender les différences entre domaines de spécialité et entre corpus.

D'abord, dans l'état de la question, nous avons rappelé l'intérêt que peut présenter l'identification des PRS dans les corpus spécialisés, intérêt qui avait déjà été souligné par nombre de chercheurs. Nous avons passé en revue divers travaux, dont les travaux de Davidson *et al.* (1998) et de Meyer *et al.* (2001), qui ont cherché à faire l'inventaire des patrons lexicaux, des patrons paralinguistiques et des patrons grammaticaux qui expriment des relations conceptuelles et ce, en anglais et en français. Nous avons aussi présenté la méthode d'extraction de relations sémantiques à partir de marqueurs lexico-syntaxiques de Séguéla (1999). L'auteur distingue, pour sa recherche, les marqueurs génériques des marqueurs spécifiques. Pour ce travail, nous supposons que les patrons lexicaux sont ceux qui sont susceptibles d'être génériques et qu'ils devraient donc être présents dans nos deux corpus. Nous avons aussi vu que la majorité des recherches sur les patrons de relations sémantiques ont porté sur des corpus de langue anglaise. Cependant, quelques travaux, dont ceux de Condamines *et al.* (1999), Meyer *et al.* (1999), Rebeyrolle (2000) et Séguéla (1999) portent sur le français. Tout comme ces auteurs, nous avons, nous aussi, choisi des corpus de langue française. Nous avons également souligné que les travaux de Condamines (2003) ont montré que la définition des genres textuels qui entrent dans la composition du corpus en fonction des structures à décrire est déterminante. De plus, pour cette auteure, la notion de « marqueur » est fortement liée au genre textuel. Enfin, nous avons vu que les textes didactiques et de vulgarisation sont des types de corpus qui sont tous deux susceptibles de comporter un grand nombre de PRS, comme le

suggèrent les études de Loffler-Laurian (1983, 1994), de Pearson (1998), et de Rebeyrolle (2000).

Notre apport à ces travaux consistait à montrer que l'association de patrons aux TS permettrait des requêtes plus précises dans les corpus spécialisés. Nous cherchions à répondre aux besoins de la terminologie en facilitant davantage le repérage de contextes explicatifs dans les corpus.

Pour atteindre cet objectif, nous avons utilisé une typologie caractérisant les différents types sémantiques et avons tenté de trouver des termes appartenant à ces TS. Nous avons, par la suite, relevé les PRS qui leur étaient associés en contexte. Les contextes étaient extraits de deux corpus spécialisés, soit un corpus d'informatique et un corpus de médecine. Ces corpus sont essentiellement composés de textes didactiques et de vulgarisation.

Les contextes sélectionnés pour la recherche devaient contenir le terme de la requête (auquel nous accédions grâce à l'un des trente-trois termes clés choisis pour la recherche), un PRS lexical (exprimant une relation de finalité, une relation causale, la méronymie ou l'hyponymie; nous avons également retenu les patrons métalinguistiques) ainsi que l'information sémantique qui se rapporte au terme. Les patrons de relations sémantiques ont ensuite été neutralisés et classés en fonction de la relation exprimée. Afin de déterminer à quels types sémantiques respectifs appartenaient les termes clés choisis pour la recherche, nous nous sommes inspirée des grands génériques de WordNet. Enfin, nous avons classé dans une base de données les informations recueillies pour la recherche (contexte, terme, type sémantique, patron de relation sémantique neutralisé, relation sémantique, corpus et source).

La principale contribution à laquelle nous nous sommes efforcée de parvenir est présentée dans le chapitre 4, consacré aux résultats. Nous avons montré que, dans la plupart des cas, on peut mettre en rapport une relation sémantique et un TS, et ce, quel que soit le domaine. Les cas de mise en relation évidente sont associés à différents types sémantiques, dont les principaux sont

repris ici. D'abord, le TS [EVENT], pour lequel la relation la plus importante pour les deux corpus est la relation causale, présente dans 83 % des cas, alors qu'aucune relation de finalité n'a été répertoriée ni dans le corpus de médecine ni dans le corpus d'informatique. Pour le TS [PSYCHOLOGICAL FEATURE (FIELD)], la relation la plus importante pour les deux corpus est la relation de finalité, présente dans 81 % cas, et aucune relation causale ni aucun patron métalinguistique n'ont été répertoriés ni dans le corpus de médecine ni dans le corpus d'informatique. Pour les TS [ACTIVITY] et [ENTITY ORGANISM], la relation la plus importante pour les deux corpus est la relation de finalité, présente dans 71 % et 50 % des contextes étudiés.

Quant au TS [PHENOMENON], la relation la plus importante pour les deux corpus est la relation causale, présente dans 73 %. Enfin, pour le TS [ENTITY OBJECT], bien que la relation la plus importante diffère d'un corpus à l'autre, la relation causale n'est présente ni dans le corpus de médecine ni dans le corpus d'informatique. Il semble donc possible de dégager des généralisations pour les relations sémantiques entre les domaines.

Nous avons toutefois observé quelques écarts entre les corpus quant à la répartition des relations sémantiques. En effet, nous avons vu que pour le TS [STATE], il y a déséquilibre entre les deux corpus. La relation causale est la principale relation dans le corpus de médecine (58 %). Dans le corpus d'informatique, c'est la méronymie qui est la plus importante (47 %), mais la relation causale apparaît tout de même dans 33 % des contextes analysés. On constate également un déséquilibre entre les deux corpus pour la relation hyperonymique puisque dans le corpus de médecine, 32 % des contextes expriment cette relation, alors qu'elle est complètement absente du corpus d'informatique. Enfin, la relation de finalité n'est jamais associée au TS [STATE] dans le corpus d'informatique, alors qu'on en compte trois (5 %) dans le corpus de médecine.

Un autre cas d'écart important entre nos deux corpus concerne le TS [ABSTRACTION]. En effet, la relation de finalité est la principale relation observée dans le corpus de médecine (88 %), alors que c'est l'hyponymie qui est la mieux représentée dans le corpus d'informatique (43 %). La relation de finalité apparaît tout de même dans 33 % des contextes extraits du corpus d'informatique. On note également que la relation causale est peu importante dans le corpus de médecine (12 %) et qu'elle est absente du corpus d'informatique pour ce TS. Enfin, les relations méronymiques et les patrons métalinguistiques sont également absents du corpus de médecine, alors qu'ils représentent 15 % et 10 % respectivement des contextes dans le corpus d'informatique.

Pour le TS [ENTITY OBJECT], on se serait attendu à ce que la relation de finalité soit la relation la plus importante pour ce TS et ce, dans les deux corpus. Il est vrai que cette relation est la plus fréquente en médecine (90 %), mais en informatique, bien qu'il n'y ait que 11 % d'écart entre cette relation et la relation hyperonymique, qui est la principale relation pour ce type sémantique, la relation de finalité est seulement la troisième en importance, derrière les relations hyperonymique et méronymique.

Il est également important de rappeler que deux TS, soit [ENTITY ORGANISM] et [PSYCHOLOGICAL FEATURE (FIELD)], sont surtout représentés par des termes dont la proportion de contextes retenus par rapport au nombre de contextes étudiés est peu élevée. Quant au TS [PSYCHOLOGICAL FEATURE], il est représenté par des termes dont la proportion de contextes retenus par rapport au nombre de contextes étudiés est la plus élevée. Ce TS est suivi par les TS [STATE] et [ABSTRACTION].

Cependant, pour ce qui est des patrons de relations sémantiques eux-mêmes, la mise en relation des TS est beaucoup moins marquée. En effet, la proportion la plus élevée de PRS communs aux deux corpus est de 18 %, résultat obtenu pour les TS [PHENOMENON] et [ACTIVITY]. Pour le TS [PHENOMENON], ces PRS sont : **avec X Y, X entraîne Y (Y entraîne X, Y entraîne conséquence sur**

X), **X résulte de Y (résultat de Y X)**, **conséquence de X Y**, **X conduit à Y** et **cause de X Y (cause de Y X)**, patrons exprimant tous la causalité. Quant au TS [ACTIVITY], les patrons communs aux deux corpus sont les suivants : **X permet Y (Y permet X, Y ne permet pas X)**, **rôle de X Y**, **objectif de X Y** (patrons exprimant la finalité); **exemple de X Y, Y est X (X est Y)** (patrons exprimant l'hyponymie) et **étape de X Y** (patron exprimant la méronymie). Ces résultats rejoignent les constats faits par Condamines (2002, 2003) selon lesquels les patrons sont davantage liés aux genres textuels et non aux TS comme nous le croyions au moment d'entreprendre cette recherche. Il serait néanmoins possible de mettre à profit les quelques PRS communs aux deux corpus afin de préciser des requêtes dans le cadre d'une recherche terminologique.

Les résultats obtenus sur les patrons de relations sémantiques remettent également en question la notion de « patron générique ». Nous avons formulé l'hypothèse voulant que les patrons lexicaux soient plus susceptibles de se trouver dans des corpus de nature différente. Il semble que cela ne soit pas le cas.

Enfin, nous avons relevé des patrons de relations sémantiques moins bien documentés, comme les PRS exprimant une relation de finalité : **on confie Y à X**, **X prend en charge Y**, **X orienté vers Y**, **base de X Y**, **Y au service de X**. Pour la relation causale, nous avons relevé le PRS suivant : **X entre en ligne de compte pour Y**. Pour la relation hyperonymique, nous avons relevé les PRS suivants : **le plus connu de X Y** et **X vu comme Y**. Pour la relation méronymique, nous avons relevé les PRS suivants : **dimension de X Y**, **Y aspect du processus de X**, **Y extrait de X**.

Contrairement à une idée très répandue en terminologie et en ingénierie des connaissances, les résultats de notre étude montrent que les relations méronymique et hyperonymique ne sont pas les plus fréquentes dans les corpus. En effet, la relation la plus représentée est la relation de finalité (32 % des contextes). Cette relation est suivie par les relations hyperonymique (29 %), causale (20 %) méronymique (13 %) et enfin, les patrons métalinguistiques (6 %).

Ce travail permet d'enrichir les connaissances dans les domaines de la terminologie et de l'analyse des textes spécialisés.

Cette étude a également permis de prouver que la détermination des patrons de relations sémantiques (PRS) en tant qu'outil mis en relation avec les types sémantiques (TS) des termes peut permettre de prévoir la nature de la définition.

Au terme de cette étude, nous proposons quelques pistes qui pourraient constituer un prolongement de notre travail.

Parmi les nouvelles voies à explorer, nous pourrions étendre notre étude aux PRS paralinguistiques et aux PRS grammaticaux. Notre choix a été de nous concentrer sur les PRS lexicaux.

Il serait également intéressant d'étudier d'autres relations. Notre choix a été d'étudier les relations causale, méronymique, hyperonymique, de finalité et les patrons métalinguistiques. Nous avons étudié les relations méronymique et causale plus en profondeur grâce aux typologies de Winston *et al.* (1987) et de Barrière (2002), respectivement. Toutefois, les types sémantiques choisis peuvent être associés à des relations sémantiques différentes que notre étude ne permet pas de mettre au jour.

Ensuite, il serait intéressant de tester notre modèle sur d'autres corpus pour voir si nos résultats concordent. En effet, en faisant une requête à partir des PRS plutôt que des termes, on pourrait ainsi voir si effectivement les termes repérés suivent la tendance de notre modèle.

Nous avons choisi des corpus portant sur des domaines différents, mais il serait intéressant de voir si la proportion des patrons qui sont communs à différents corpus est plus élevée lorsque ces corpus sont de nature semblable. Un corpus technique pourrait peut-être présenter davantage de similitude avec notre corpus d'informatique, par exemple.

Enfin, on pourrait envisager d'effectuer cette recherche sur des corpus de textes rédigés dans d'autres langues.

INDEX DES NOTIONS

- Contexte pertinent, 112, 113
- Corpus, 14, 15, 33, 34, 98
- Corpus de médecine, 98, 99, 100, 102
- Corpus d'informatique, 98, 99, 100, 102
- Corpus général, 13
- Corpus spécialisé, 13, 25
- Fréquence des contextes pertinents, 19, 29, 105
- Genre textuel, 23, 24
- Grand générique, 127, 128, 129, 130
- Langue de spécialité, 21, 25, 31, 74, 86, 128
- Neutralisation des patrons, 121
- Niveau de spécialisation, 25, 26, 27, 28, 29, 30, 100
- Patron de relation sémantique, 58, 59
- Forme du PRS, 114, 115
- Patron métalinguistique, 57
- PRS grammatical, 77
- PRS lexical, 76
- PRS paralinguistique, 77
- Registre, 25, 26, 27, 28, 29
- Regroupement des patrons, 125
- Relation causale, 44, 56
- Relation conceptuelle, 37
- Relation de finalité, 54, 55, 56
- Relation hyperonymique, 38, 39, 40, 55
- Relation méronymique, 40, 41, 42, 43, 44, 56
- Relation sémantique, 34, 35, 37, 55, 133
- Représentativité d'un corpus, 18
- Reprise anaphorique, 113
- Terme, 103
- Extension de sens du terme, 113
- Terme clé, 104, 105, 107
- Terme complexe, 108, 109, 110, 111, 112
- Terme simple, 108, 109, 110, 111, 112
- Type de texte, 22, 23
- Type sémantique, 130, 131, 133

BIBLIOGRAPHIE

- AHMAD, K. (1995). « Pragmatic of specialist terms: The acquisition and representation of terminology », in P. Steffens (ed.), *Machine Translation and the Lexicon: Proceedings of the 3rd International EAMT Workshop*, 26-28 avril 1993 in Heidelberg, Berlin/New York, Springer Verlag, 51-76.
- AHMAD, K. et H. FULFORD (1992). « Knowledge processing 4. Semantic relations and their use in elaborating terminology », in *Computing Sciences Technical Report CS-92-07*, Guildford, Surrey, University of Surrey.
- AHMAD, K. et M. ROGERS (2001). « Corpus linguistics and terminology extraction », in S.E. Wright et G. Budin (eds.), *Handbook of Terminology Management*, Amsterdam, Philadelphia, J. Benjamins, vol. 2, 725-760.
- ALLEN, J. P. B. et H. G. WIDDOWSON (1974). *English in Focus: Physical Science*, Oxford, Oxford University Press.
- APRESJAN, J. D. (1992). « Systematic lexicography », in H. Tommola et K. Varantola (eds.), *Proceedings of EURALEX '92, 5th International Congress on lexicography, I-II*, coll. *Studia translologica*, 4-9 août 1992, Tampere, Finland, Tampereen yliopisto, Ser. A, vol. 2, 3-16.
- ATKINS, S., J. CLEAR, et N. OSTLER (1992). « Corpus design criteria », in *Literary and Linguistic Computing*, 7 (1), 1-16.
- AUGER, A. (1998). *Repérage des énoncés d'intérêt définitoire dans les bases de données textuelles*, document maison rédigé à partir de la thèse de doctorat de l'auteur, Suisse, Université de Neuchâtel, 1997.

- AUSSENAC-GILLES, N et A. CONDAMINES (2001). « Entre textes et ontologies formelles : les bases de connaissances terminologiques », in *Ingénierie et capitalisation des connaissances*, Paris, Hermès, 153-176.
- BAKHTINE, M. (1984). *Esthétique de la création verbale*, Paris, Gallimard, Tel.
- BARNBROOK, G. et J. SINCLAIR (1995). « Parsing cobuild entries », in J. Sinclair, M. Hoelter et C. Peters (eds.), *The Languages of Definition: The Formalization of Dictionary Definitions for Natural Language Processing*, Luxembourg, Office for Official Publication of the European Communities, 13-58.
- BARRIÈRE, C. (2001). « Investigating the causal relation in informative texts », in *Terminology*, 7 (2), 135-154.
- BARRIÈRE, C. (2002). « Hierarchical refinement and representation of the causal relation », in *Terminology*, 8 (1), 91-111.
- BERNARD, J. R. L. (ed.) (1990). *The Macquarie Encyclopaedic Thesaurus: The Book of Words*, Melbourne, Macquarie Library Pty.
- BIBER, D. (1988). *Variation Across Speech and Writing*, Cambridge [England]/New York, Cambridge University Press.
- BIBER, D. (1989). « A typology of English texts », *Linguistics*, 27, 3-43.
- BIBER, D. (1993). « Representativeness in corpus design », in *Literary and Linguistic Computing*, 8 (4), 243-257.
- BODSON, C. (1997). Méthode de description des verbes spécialisés en vue d'applications en traduction automatique, accompagnée d'une étude portant sur trente-cinq verbes du domaine de l'informatique, mémoire de maîtrise, Montréal, Université de Montréal.

- BORILLO, A. (1996). « Exploration automatisée de textes de spécialité : repérage et identification de la relation lexicale d'hyponymie », in *Linx*, 34-35, 113-124.
- BOURIGAULT, D. et M. SLODZIAN (1999). « Pour une terminologie textuelle », in C. Enguehard et A. Condamines (eds.), *Terminologie et intelligence artificielle (TIA-99) : actes des 3^e rencontres*, 10 et 11 mai 1999, Nantes, *Terminologies nouvelles*, 19, 29-32.
- BOWDEN, P. R., P. HALSTEAD, et T. G. ROSE (1996). « Extracting conceptual knowledge from text using explicit relation markers », in *Proceedings of the 9th European Knowledge Engineering Workshop (EKAW-96): Lecture notes in Artificial Intelligence*, 1076, 14-17 mai 1996, University of Nottingham, Springer Verlag, 147-162.
- BOWKER, L. (1996). « Towards a corpus-based approach to terminology », in *Terminology*, 3 (1), 27-52.
- BOWKER, L. et J. PEARSON (2002). *Working with specialized language. A practical guide to using corpora*, London/New York, Routledge.
- BRONCKART, J. P. (1996). *Activités langagières, textes et discours*, Lausanne, Delachaux et Niestlé.
- CONDAMINES, A. (1993). « Un exemple d'utilisation de connaissances de sémantique lexicale : acquisition semi-automatique d'un vocabulaire de spécialité », in *Cahiers de lexicologie*, 62 (1), 25-65.
- CONDAMINES, A. (2000b). « Chez dans un corpus de sciences naturelles : un marqueur de relation méronymique? », in *Cahiers de lexicologie*, 77 (2), 165-187.
- CONDAMINES, A. (2002). « Corpus analysis and conceptual relation patterns », in *Terminology*, 8 (1), 141-162.

- CONDAMINES, A. (2003). *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*, in Carnets de grammaire, Rapports internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), Rapport 13, UMR 5610, Toulouse, CNRS et Université de Toulouse-le Mirail.
- CONDAMINES, A. et P. AMSILI (1993). « Terminology between language and knowledge: an example of terminological knowledge base », in *Proceedings of TKE'93: Terminology and Knowledge Engineering*, 25-27 août 1993, Francfort, Indeks Verlags, 316-323.
- CONDAMINES, A., M. P. JACQUES, A. M. SOUBEILLE, N. AUSSENAC-GILLES, J. FELIU et J. REBEYROLLE (1999). *SGGD – Analyse des corpus en vue de la constitution d'un référentiel terminologique*, ERSS (Équipe de recherche en syntaxe et sémantique), UMR 5610, Toulouse, CNRS et Université de Toulouse-le Mirail.
- CONDAMINES, A. et J. REBEYROLLE (1998). « CTKB: a corpus-based approach to a terminological knowledge base », in *Proceedings of Computerm '98: First Workshop on Computational Terminology*, 15 août 1998, Montreal, Université de Montréal, COLING-ACL '98, 29-35.
- CONDAMINES, A. et J. REBEYROLLE (1997). « Point de vue en langue spécialisée », in *Meta*, 42 (1), 174-184.
- CONDAMINES, A. et J. REBEYROLLE (2001). « Searching for and identifying conceptual relationships via a corpus-based approach to a terminological knowledge base (CTKB): method and results », in D. Bourigault, C. Jacquemin et M. C. L' Homme (eds.), *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 127-148.
- CRUSE, D. A. (1986). *Lexical Semantics*, Cambridge/New York, Cambridge University Press.

- DAVIDSON, L. (1997). *Knowledge Extraction Technology For Terminology*, Master's Thesis, Ottawa, University of Ottawa.
- DAVIDSON, L., J. KAVANAGH, K. MACKINTOSH, I. MEYER, et D. SKUCE (1998). « Semi-automatic extraction of knowledge-rich contexts from corpora », in D. Bourigault, C. Jacquemin et M. C. L'Homme (eds.), in *Proceedings of Computerm'98: First Workshop on Computational Terminology*, 15 août 1998, Montreal, Université de Montréal, *COLING-ACL'98*, 50-56.
- DE BESSÉ, B. (1990). « La définition terminologique », in J. Chaurand et F. Mazière (eds.), *La Définition*, Paris, Larousse, 252-261.
- DELAMARE, J. (1999). *Dictionnaire abrégé des termes de médecine*, 3^e édition, Paris, Maloine.
- DELAS, D. et D. DELAS-DEMON (1991). *Dictionnaire des idées par les mots (analogique)*, coll. « Les usuels du Robert », Paris, Le Robert.
- DESCLÉS, J. P. (1996). « Appartenance/inclusion, localisation, ingrédience et possession », in *Faits de langues*, 7, 91-100.
- EDWARDS, P. (ed.) (1967). *The Encyclopedia of Philosophy*, vol. 1, New York/London, Macmillan.
- ENGWALL, G. (1994). « Not chance but choice: criteria in corpus creation », in Atkins, BTS & Zampoli, A. (eds.), *Computational Approaches to the Lexicon*, Oxford, Clarendon Press, 49-82.
- EVENS, M. (ed.) (1988). *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*, Cambridge/New York, Cambridge University Press.

- FANG, C. Y. (1991). « Building a corpus of the English of computer science », in J. Aarts, P. de Haan, and N. Oostdijk (eds.), *English Language Corpora*, Amsterdam, Rodopi, 73-78.
- FLOWERDEW, J. (1991). « Pragmatic modifications on the ‘representative’ speech act of defining », in *Journal of Pragmatics*, 15 (3), 253-264.
- FLOWERDEW, J. (1992a). « Salience in the performance of one speech act, the case of definitions », in *Discourse Processes*, 15, 165-181.
- FLOWERDEW, J. (1992b). « Definitions in science lectures », in *Applied Linguistics*, 13 (2), 202-221.
- FLOWERDEW, J. (1993). « Concordancing as a tool in course design », in *System* 21 (2), 231-244.
- FRADIN, B. et J. M. MARANDIN (1979). « Autour de la définition : de la lexicographie à la sémantique », in *Langue française*, 43, 60-83.
- FRANCIS, W. N. (1982). « Problems of assembling and computerizing large corpora », in S. Johansson (ed.), *Computer Corpora in English Language Research*, Bergen, Norwegian Computing Centre for the Humanities, 7-24.
- FRANCIS, W. N. (1992). « Language corpora B.C. », in J. Svartvik (ed.), *Directions in Corpus Linguistics, Proceedings of the Nobel Symposium 82*, 4-8 août 1991, Berlin/New York, Mouton de Gruyter, 17-32.
- GARCIA, D. (1996). « COATIS, un outil d’aide à l’acquisition des connaissances causales exprimées dans les textes », in P. Bouffard et A. Kharrat (eds.), *Actes du premier colloque étudiant de linguistique informatique de Montréal, CLIM’96*, 8-11 juin 1996, Montréal, 97-103.

- GARCIA, D. (1997). « Structuration du lexique de la causalité et réalisation d'un outil d'aide au repérage de l'action dans les textes », in *Terminologie et intelligence artificielle (TIA-97): actes des deuxièmes rencontres*, Équipe de Recherche en Syntaxe et Sémantique, 3 et 4 avril 1997, Toulouse, Université Toulouse-le Mirail, 7-26.
- GARNIER *et al.* (2002). *Dictionnaire des termes de médecine*, 27^e édition revue et mise à jour, Paris Maloine.
- GINGUAY, M. (1999). *Dictionnaire d'informatique anglais/français*, 12^e édition, Paris, Dunod.
- GLADSTONE, W. J. (1996). *Dictionnaire anglais-français des sciences médicales et paramédicales : English-French Dictionary of Medical and Paramedical Sciences*, 4^e édition, Saint-Hyacinthe, Québec, Edisem.
- GOUVERNEMENT DU CANADA. *Termium*, en ligne. Adresse URL: <http://www.termiumplus.translationbureau.gc.ca/site/>.
- GUHA, R. V. et D. B. LENAT (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*, Reading, Mass., Addison-Wesley Pub.
- HABERT, B., A. NAZARENKO et A. SALEM (1997). *Les linguistiques de corpus*, coll. U. Série «Linguistique», Armand Colin.
- HARRIS, Z. (1982). « Discourse and sublanguage », in R. Kittredge et J. Lehrberger (eds.), *Sublanguage: Studies of Sublanguage in Restricted Semantic Domains*, Berlin/New York, Walter de Gruyter, 232-236.
- HATOUT, N. (1996). « Pour la construction d'une base de connaissances lexicologiques à partir du Trésor de la langue française : les marqueurs superficiels dans les définitions spécialisées », in *Cahiers de lexicologie*, 68 (1), 137-173.

- HEARST, M. A. (1992). « Automatic acquisition of hyponyms from large text corpora », in *Proceedings of the Fifteenth International Conference on Computational Linguistics*, Coling-92, 23-28 juillet 1992, Nantes, France, 539-545.
- IRIS, M. A., B. E. LITOWITZ, et M. EVENS (1988). « Problems of the part-whole relation », in M. W. Evens (ed.), *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*, Cambridge, Cambridge University Press, 261-288.
- JACKSON, H. (1988). *Words and their Meaning*, London/New York, Longman.
- JACOBI, D. (1994). « Lexique et reformulation intradiscursive dans les documents de vulgarisation scientifique », in D. Candel (ed.), *Études de sémantique lexicale, français scientifique et technique et dictionnaire de langue*, coll. Études de sémantique lexicale, CNRS, Institut national de la langue française, Paris, Didier Érudition, 77-91.
- JAMES, G., R. DAVISON, H. A. CHEUNG et S. DEERWESTER (1994), *English in Computer Science: A Corpus-Based Lexical Approach*, Hong Kong, Longman Asia.
- JOHANSSON, S. (1991). « Times change, and so do corpora », in Aijmer et Altenburg (eds.), *English Corpus Linguistics: Studies in Honour of J. Svartvik*, London/New York, Longman, 305-314.
- KOCOUREK, R. (1991). *La langue française de la technique et de la science. Vers une linguistique de la langue savante*, 2^e édition augmentée, refondue, mise à jour avec une nouvelle bibliographie, Wiesbaden, O. Brandstetter.
- LARIVIÈRE, L. (1996). « Comment formuler une définition terminologique », in *Meta*, 41 (3), 405-418.

- LAROUSSE (1996). *Dictionnaire de l'informatique*, coll. Références Larousse, Paris, Larousse-Bordas.
- LEECH, G. (1991). « The state of the art in corpus linguistics », in K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics: Studies in Honour of J. Starvik*, London, Longman, 8-29.
- LEVIN, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*, Chicago, The University of Chicago Press.
- L'HOMME, M. C. (1995). « Processing word combinations in existing term banks », *Terminology* 2 (1), 137-158.
- L'HOMME, M. C. (1996). « A computerized model for processing lexical combinations in technical language », in Järborg *et al.* (eds.), *Proceedings of Euralex '96*, M., 12-17 août 1996, Göteborg, Sweden, Göteborg University, 797-806.
- L'HOMME, M. C. (1997). « Organisation des classes conceptuelles pour l'accès informatisé aux combinaisons lexicales spécialisées verbe + nom (terme) », in *Terminologie et intelligence artificielle (TIA-97) : actes des deuxièmes rencontres*, Équipe de Recherche en Syntaxe et Sémantique, 3 et 4 avril 1997, Toulouse, Université Toulouse-le Mirail, 161-174.
- L'HOMME, M. C. (2004). *La terminologie : principes et techniques*, Montréal, Les Presses de l'Université de Montréal.
- LOFFLER-LAURIAN, A. M. (1983). « Typologie des discours scientifiques : deux approches », in *Études de linguistique appliquée*, 51, 8-20.
- LOFFLER-LAURIAN, A. M. (1994). « Les définitions dans la vulgarisation scientifique (PRESSE, MUSÉES) », in D. Candel (ed.), *Français scientifique et technique et dictionnaire de langue*, coll. Études de sémantique lexicale, CNRS, Institut national de la langue française, Paris, Didier Érudition, 93-112.

- LYONS, J. (1968). *Introduction to Theoretical Linguistics*, Cambridge, Cambridge University Press.
- LYONS, J. (1977). *Semantics: Volume 1*, Cambridge, Cambridge University Press.
- MARSHMAN, E. (2002). The Cause Relation in Biopharmaceutical Corpora: English and French Patterns for Knowledge Extraction, Master's Thesis, Ottawa, University of Ottawa.
- MARSHMAN, E., T. MORGAN, et I. MEYER (2002). « French patterns for expressing concept relations », in *Terminology*, 8 (1), 1-29.
- McENERY, T. et A. WILSON (1996). *Corpus linguistics*, Edinburgh, Edinburgh University Press.
- MEYER, I. (1994). «Linguistic strategies and computer aids for knowledge engineering approach », in *L'actualité terminologique/Terminology Update*, 27 (4), 6-10.
- MEYER, I. (2001). « Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework », in D. Bourigault, C. Jacquemin et M. C. L'Homme (eds.), *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 279-302.
- MEYER, I. et K. MACKINTOSH (1996a). « Refining the terminographer's concept-analysis methods: How can phraseology help? », in *Terminology*, 3 (1), 1-26.
- MEYER, I. et K. MACKINTOSH (1996b). « The corpus from a terminographer's viewpoint », in *International Journal of Corpus Linguistics*, 1 (2), 257-285.

- MEYER, I., K. MACKINTOSH, C. BARRIÈRE, et T. MORGAN (1999). « Conceptual sampling for terminological corpus analysis », in P. Sandrini (ed.), *Proceedings of TKE '99: Fifth International Congress on Terminology and Knowledge Engineering*, 23-27 août 1999, Innsbruck, Austria, TermNet, 256-267.
- MEYER, I. *et al.* (Page consultée le 18 avril 2001). *Les contextes définitoires dans les corpus spécialisés : vers la détection automatique*, en ligne. Adresse URL: <http://aix1.uottawa.ca/~imeyer/burtrad/default.htm>.
- MILLER, G. A. (1990). « Nouns in WordNet: a lexical inheritance system », in *International Journal of Lexicography*, 3 (4), 245-264.
- MILLER, G. A. (1998) : « Nouns in WordNet » in C. Fellbaum (ed.): *WordNet, An Electronic Lexical Database*, Cambridge, Massachusetts, MIT Press, 23-46.
- MORGAN, T. (2000). *A Comparative Study of Hypernymic Patterns for Knowledge Extraction*, Master's Thesis, Ottawa, University of Ottawa.
- MORIN, E. (1999a). *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques*, thèse de doctorat, Nantes, Université de Nantes.
- MORIN, E. (1999b). « Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique », in *Traitement automatique des langues (T.A.L.)*, 40 (1), 143-166.
- MORIN, Y., G. AMARENCO et J. BARBAUT (2000), *Larousse médical*, Paris Larousse-Bordas.
- MULLER, C. (1979). *Langue française et linguistique quantitative : recueils d'articles*, Genève, Slatkine.

- MULLER, C. (1992a). *Principes et méthodes de statistique lexicale*, coll. Unichamp, 33, Paris, Champion.
- MULLER, C. (1992b). *Initiation aux méthodes de la statistique linguistique*, coll. Unichamp, 32, Paris, Champion.
- NATIONAL LIBRARY OF MEDICINE (Page consultée le 22 avril 2003). *Unified Medical Language System (UMLS) Semantic Network*, en ligne. Adresse URL : <http://www.nlm.nih.gov/pubs/factsheets/umlssemn.html>.
- NUOPPONEN, A. (1994). « Causal relations in terminological knowledge representation », in *Terminology, Science & Research (Journal of the International Institute for Terminology Research, IITF)*, 5 (1), 36-44.
- OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE (2002). *Guide méthodologique du Grand dictionnaire terminologique* (document interne), Québec, Gouvernement du Québec.
- OFFICE QUÉBÉCOIS DE LA LANGUE FRANÇAISE (Page consultée le 9 juin 2003). *Grand dictionnaire terminologique (GDT)*, en ligne. Adresse URL: http://www.oqlf.gouv.qc.ca/ressources/gdt_bdl2.html.
- OTMAN, G. (1996). *Les représentations sémantiques en terminologie*, Paris Masson.
- PALMER, F. R. (1981). *Semantics*, Cambridge, Cambridge University Press.
- PEARSON, J. (1996). « The expression of definitions in specialised texts: a corpus-based analysis », in M. Gellerstam, J. Järborg, S.G. Malmgren, K. Norén, L. Rogström & C. Pappmehl (eds.), *Proceedings of Euralex '96: Seventh International Congress on Lexicography*, Part II, 12-17 août 1996, Göteborg, Sweden, Göteborg University, 759-769.
- PEARSON, J. (1998). *Terms in Context*, Amsterdam/Philadelphia, John Benjamins.

- PÉCHOIN, D. (dir.) (1991). *Thésaurus Larousse*, Paris, Larousse.
- PÉRY-WOODLEY, M. P. (1995). « Quels corpus pour quels traitements automatiques? », in *Traitement automatique des langues (TAL)*, 36 (1-2), 213-232.
- PICOCHÉ, J. (1992). *Précis de lexicologie française : l'étude et l'enseignement du vocabulaire*, 2^e édition revue et augmentée, Paris, Nathan.
- PRINCETON UNIVERSITY, COGNITIVE SCIENCE LABORATORY. (Page consultée le 22 avril 2003). *WordNet: a Lexical Database for the English Language*, en ligne. Adresse URL : <http://www.cogsci.princeton.edu/~wn/>.
- QUÉRIN, S. (1998). *Dictionnaire des difficultés du français médical*, Saint-Hyacinthe, Edisem.
- QUEVAUVILLIERS. J. et A. FINGERHUT (1999). *Dictionnaire médical*, Paris Masson.
- REBEYROLLE, J. (2000). *Forme et fonction de la définition en discours*, thèse de doctorat, Toulouse, Université Toulouse-le Mirail.
- REBEYROLLE, J. et L. TANGUY (2000). « Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires », *Les Cahiers de Grammaire, Sémantique et Corpus*, 25, 153-174.
- RÉSEAU PANLATIN DE TERMINOLOGIE (2004). *Dictionnaire panlatin des termes de base de l'informatique*, version 2.05, en ligne. Adresse URL : <http://www.realiter.net/microinf/CDICO30.HTM#Concept:50>.
- REY-DEBOVE, J. (1978). *Le métalangage : étude linguistique du discours sur le langage*, coll. L'ordre des mots, Paris, Le Robert [réédition A. Colin, 1997].

- RICHARDSON, S. D., W. B. DOLAN et L. VANDERWENDE (1998). « Mindnet: acquiring and structuring semantic information from text », in *Proceedings of COLING – ACL' 98*, vol. 2, 10-14 août 1998, Montréal, Université de Montréal, 1098-1102.
- RIEGEL, M. (1987). « Définition directe et indirecte dans le langage ordinaire : les énoncés définitoires copulatifs », in *Langue française*, 73, 29-53.
- ROE, P. J. (1977). *The Notion of Difficulty in Scientific Text*, Ph.D. Thesis, Birmingham, University of Birmingham.
- SAGER, J. C. (1990). *A Practical Course in Terminology Processing*, Amsterdam/Philadelphia, John Benjamins.
- SAGER, J., D. DUNGWORTH, et P. F. Mc DONALD (1980). *English Special Languages: Principles and Practice in Science and Technology*, Wiesbaden, Brandstetter Verlag.
- SAGER, N. et C. FRIEDMAN (1987). *Medical Language Processing: Computer Management of Narrative Data*, Reading, MA, Addison-Wesley.
- SÉGUÉLA, P. (1999). « Adaptation semi-automatique d'une base de marqueurs de relations sémantiques sur des corpus spécialisés », in *Terminologie et intelligence artificielle (TIA-99) : actes du colloque de Nantes*, 10 et 11 mai 1999, Nantes, *Terminologies nouvelles*, 19, 52-60.
- SÉGUÉLA, P. (2001). *Construction de modèles de connaissances par analyse linguistique de relations lexicales dans les documents techniques*, thèse de doctorat, Toulouse, Université Paul Sabatier.
- SINCLAIR, J. M. (1991). *Corpus. Concordance. Collocation*, Oxford, Oxford University Press.

- SINCLAIR, J. (1994). « Corpus typology: a framework for classification », Expert Advisory Group on Language Engineering Standards (EAGLES) document 1-18, now published as J. Sinclair 1995. “Corpus typology: a framework for classification”, in G. Melchers et B. Warren, *Studies in Anglistics*, Stockholm, Almquist et Wiksell International, 17-34.
- SWALES, J.(1981). « Definitions in science and law – evidence for subject-specific course components? », in *Fachsprache*, 81 (3), 106-112.
- TALMY, L. (1985). « Lexicalization patterns: semantic structure in lexical forms », in *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, vol. 3, T. Shopen (ed.), Cambridge, Cambridge University Press.
- TALMY, L. (1988). « Force dynamics in language and cognition », in *Cognitive Science*, 12, 49-100.
- TALMY, L. (2000). *Toward a Cognitive Semantics*, vol. 1-2., Cambridge, MA, MIT Press.
- Université du Québec à Montréal, Département de linguistique et de didactique des langues, Centre d’analyse de texte par ordinateur de la Faculté des sciences humaines. (Page consultée le 22 avril 2003). *SATO-4, Système d’analyse de texte par ordinateur*, en ligne. Adresse URL : <http://www.ling.uqam.ca/ato/>.
- University of Amsterdam, Institute for Logic, Language and Computation (ILLC). (Page consultée le 22 avril 2003). *Welcome to EuroWordNet: Building a Multilingual database with wordnets for several European languages*, en ligne. Adresse URL : <http://www.ilic.uva.nl/EuroWordNet/>.
- VAN CAMPENHOUDT, M., Centre de recherche Termisti (Page consultée le 19 juin 2004). *Les relations hiérarchiques*, en ligne. Adresse URL : <http://www.termisti.refer.org/theoweb4.htm>

VOSSEN, P. (ed.) (1998). *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Dordrecht/Boston/London, Kluwer Academic Publishers.

WINSTON, M. E., R. CHAFFIN, et D. HERRMANN (1987). « A Taxonomy of Part-Whole Relations », in *Cognitive Science*, 11 (4), 417-444.

WORDNET: a Lexical Database for the English Language, en ligne. Adresse URL : <http://www.cogsci.princeton.edu/~wn/>

ANNEXE A : Description des corpus

Le tableau suivant précise, pour chaque fichier, la langue de spécialité, le contenu, le niveau de spécialisation et le nombre de mot :

Nom de fichier	Langue de spécialité	Contenu	Niveau de spécialisation	Nombre de mots
Genilo	Informatique	Le génie logiciel – la fabrication de systèmes informatisés	Vulgarisation	27 840
Impri	Informatique	Utilisation d'une imprimante. Complément au manuel d'utilisation	Général	30 049
ordi96	Informatique	Les bases de données – Portrait de six logiciels qui ont su évoluer et s'adapter aux nouveaux besoin	Spécialisé	4765
pce96	Informatique	Les portables couleur	Public avisé, connaisseur du domaine	14 385
Pcprem	Informatique	Introduction à l'usage et au fonctionnement de l'ordinateur	Général	28 519
Begupa	Informatique	Principales fonctions des logiciels de traitement de textes	Formation	5899
Claviez1	Informatique	Notions de base de l'informatique (unités de mesure, unités de performance)	Général	7676
Delavi1	Informatique	Informatique de gestion. Théorie et activités pratiques. Notions fondamentales de l'informatique	Formation	19 170
Acces1	Informatique	Programmation	Spécialisé	6938
Pierre1	Informatique	Introduction aux ordinateurs : organisation, exploitation et programmation	Formation	17 974
Plaisen1	Informatique	Initiation à la micro-informatique	Formation	21 405
Mspcdos1	Informatique	Le micro-ordinateur: introduction à MS/PC-DOS, LOTUS 1-2-3 et dBASE III PLUS	Formation	41 171
Net	Informatique	Les virus informatiques et l'installation d'un antivirus	Vulgarisation	488
Infor96b	Informatique	Les serveurs web	Spécialisé	4021
Zope	Informatique	Zope : plate-forme dynamique de publication d'objets sur le Web	Technique	2954
Synchro	Informatique	Informatique mobile et la synchronisation	Technique	4635
Corba	Informatique	Intranet+Extranet+Internet, CORBA	Vulgarisation	1693
Linux	Informatique	Composants logiciels et Linux	Vulgarisation	3783
Domino	Informatique	Lotus Domino Rnext Bêta 2. Environnement de travail collaboratif	Vulgarisation	3418
Internet	Informatique	Définition d'Internet et histoire de	Formation	2758

Nom de fichier	Langue de spécialité	Contenu	Niveau de spécialisation	Nombre de mots
		son évolution		
Trajet	Informatique	Taux de transfert de données	Formation	2066
Macosx	Informatique	Présentation de Mac OS X	Vulgarisation	8799
Asp	Informatique	Location d'applications et le modèle ASP	Vulgarisation	4418
Reseauip	Informatique	Réseaux TCP/IP	Spécialisé	6139
Persoxp	Informatique	Windows XP	Formation	7968
Ecran	Informatique	L' écran, histoire de son évolution	Vulgarisation	3262
Linux2	Informatique	Installation de Linux	Vulgarisation	1916
Linux3p1	Informatique	Guide d'utilisation de Linux	Technique	47 003
Linux3p2	Informatique	Guide d'utilisation de Linux	Technique	50 642
Linux3p3	Informatique	Guide d'utilisation de Linux	Technique	39 771
Linux3p4	Informatique	Guide d'utilisation de Linux	Technique	35 084
Windows	Informatique	Windows	Spécialisé	13 539
Cosnar	Informatique	Globalisation des ressources informatiques et des données informatiques, systèmes	Spécialisé	1576
Glossa	Informatique	Glossaire informatique français-anglais	Général (rédigé par des experts du INRIA, s'adresse aux usagers de l'informatique)	5865
Comint	Informatique	Initiation à Internet	Général	991
Sysint	Informatique	Vulgarisation des systèmes d'exploitation	Formation	3735
Bddint	Informatique	Introduction aux bases de données	Général	1132
Conres	Informatique	Le concept de réseau	Général	4071
Foncmtot	Informatique	Initiation aux moteurs de recherche et à Internet	Vulgarisation	5842
Inordi	Informatique	Introduction à la notion d'ordinateur	Général	8964
Rollan	Informatique	Introduction à Internet	Formation	16 891
Ceveil	Informatique	Problèmes liés à l'internationalisation d'Internet	Spécialisé (rédigé par un comité gouvernemental pour les contribuables)	9602
Giri	Informatique	Initiation à la recherche sur Internet	Formation	15189
Linint	Informatique	Introduction à Linux	Formation	5596
Optrep	Informatique	Initiation au dépannage informatique	Formation	10 515
Secint	Informatique	Initiation à la sécurité informatique	Formation	548
Adevim	Informatique	Notions fondamentales de l'informatique	Général (rédigé par des fonctionnaires pour les néophytes)	3253
Antho	Informatique	Composantes et modification d'un PC	Technique (rédigé par des experts pour les initiés)	5621

Nom de fichier	Langue de spécialité	Contenu	Niveau de spécialisation	Nombre de mots
Biosrv	Informatique	Le BIOS	Général	2382
Disdur	Informatique	Installation d'un deuxième disque dur	Formation	1423
Fonord	Informatique	Notions fondamentales sur l'informatique et les ordinateurs (composantes)	Général	2500
Formp3	Informatique	Le format mp3	Vulgarisation	1547
Ovlock	Informatique	Les principes de base de l'overlocking	Technique	1972
Societ	Informatique	Mondialisation de l'Internet	Vulgarisation	6245
Toutsa	Informatique	Les virus informatiques	Vulgarisation	1391
Window	Informatique	Notions de base de Windows	Général	1 889
Anosec	Informatique	Anonymat sur Internet et sécurité en ligne	Général (rédigé par des experts pour les initiés)	3712
Chemla	Informatique	Internet pour les débutants	Général	3386
Etenir	Informatique	Entretien d'un disque dur	Général	1439
Instxp	Informatique	Installation de Windows XP	Général	1245
Mesenc	Informatique	Rapport sur le retard informatique de la France	Général	3057
Missio	Informatique	Rapport sur le développement informatique et de l'Internet en France	Spécialisé	35 129
Resloc	Informatique	Installation d'un réseau local	Technique	1345
Touze	Informatique	Dactylographie informatique. Utilisation du clavier	Formation	1381
Webcam	Informatique	Les webcams	Vulgarisation	743
Aspect	Informatique	Programmation Orientée Aspect	Vulgarisation	4486
Contrat	Informatique	Programmation par contrat et qualité logicielle	Vulgarisation	3492

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
Reta	Médecine	Retard ou avance de croissance, retard ou avance pubertaire – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2982
Zadven	Médecine	Pharmacologie cardio-vasculaire et respiratoire	S'adresse à des professionnels de la santé.	15 451
Opac	Médecine	Opacité médiastino-thoracique – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2934
Isch	Médecine	Accidents vasculaires cérébraux ischémiques	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4516
Hemor	Médecine	Accidents vasculaires cérébraux hémorragiques	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3804
Zdjiane	Médecine	Thérapeutique cardiovasculaire	Professionnels de la santé	21 416
Trans	Médecine	Échographie transfontanellaire – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1252
Thyr	Médecine	Imagerie – thyroïde	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2534
Patho1	Médecine	Pathologie de l'orbite	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5428
Saliv	Médecine	Imagerie – glandes salivaires	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5619
Zcarraz	Médecine	Médicaments du cœur, des vaisseaux et de la circulation du sang	S'adresse à des professionnels de la santé	10 620
Malf	Médecine	Malformations et obstructions des voies urinaires – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4860
Lestum	Médecine	Tumeurs cérébrales	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3255
Image8	Médecine	Imagerie radiologique – thrombophlébites cérébrales	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes	1205

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
			(France).	
Image1	Médecine	Imagerie cérébrale – sujet âgé	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2178
Scia	Médecine	Imagerie – sciatique opérée et récidivante	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1999
patho3	Médecine	Pathologie infectieuse encéphalique	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5642
Crani	Médecine	Traumatismes crânio-encéphaliques	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5054
Intrach	Médecine	Radiologie interventionnelle du rachis	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1970
Image9	Médecine	Traumatismes crânio-encéphaliques	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	6133
Cong	Médecine	Anomalies congénitales – thorax – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3028
Doul	Médecine	Douleurs abdominales – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2733
Hanc	Médecine	Échographie – hanche – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5886
Squel	Médecine	Traumatismes du squelette (sauf crâne et rachis) – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3182
Abdo	Médecine	Diagnostic – masse abdominale ou pelvienne – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	8047
Image7	Médecine	Imagerie des maladies de la substance blanche	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3103
Steno	Médecine	Sténoses du canal rachidien	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5444

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
Toux	Médecine	Toux chronique – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1929
Xrenal1	Médecine	Pharmacologie cardiovasculaire, pulmonaire et rénale	S'adresse à des professionnels de la santé	39 088
Image4	Médecine	Imagerie – Épilepsie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5578
Xgiroud	Médecine	Pharmacologie clinique	S'adresse à des professionnels de la santé	58 370
Lary	Médecine	Imagerie – cancers du larynx	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3005
Ovas	Médecine	Diagnostic radiologique – obstruction des voies aériennes – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3512
Bpa	Médecine	Radiologie de bronchopneumopathies aiguës (BPA) – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4955
Malgen	Médecine	Radiologie – maladies généralisées – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3434
Cardio	Médecine	Imagerie et cardiopathies congénitales – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2448
Osteo	Médecine	Diagnostic des ostéoporoses – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1129
Band	Médecine	Bandes claires métaphysaires – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1308
Vomi	Médecine	Diagnostic des vomissements – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3735
Lechat	Médecine	Pharmacologie médicale	S'adresse à des professionnels de la santé	14 942
Xmedpat	Médecine	Médicaments en pathologie cardio-vasculaire	S'adresse à des professionnels de la santé	60 119
Tumeur	Médecine	Tumeurs du rachis et de la moelle	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes	3440

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
Rach	Médecine	Infections du rachis	(France). Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4207
Traum2	Médecine	Traumatismes rachidiens et médullaires	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5645
Ostei	Médecine	Ostéite et Ostéoarthrite (diagnostic) – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3187
Tume	Médecine	Tumeurs osseuses – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2935
Faci	Médecine	Massif facial (traumatismes)	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2410
Urin	Médecine	Infection urinaire – pédiatrie	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3989
Snc	Médecine	Malformations du système nerveux central	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	10 630
Hydro	Médecine	Hydrocéphalies	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4611
Cran	Médecine	Pédiatrie – Traumatismes crâniens	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3116
Xabreg	Médecine	Cardiologie	S'adresse à des professionnels de la santé	3831
Zschmit	Médecine	Cardiologie	S'adresse à des professionnels de la santé	12 550
Psyc	Médecine	Retard psychomoteur; épilepsie; malformations cérébrales	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5269
Ortho	Médecine	Orthopédie (problèmes)	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3299
Ztoutou	Médecine	Pharmacie-médicaments	S'adresse à des professionnels de la santé	4531
Orei	Médecine	ORL (pathologie de l'oreille et des sinus)	S'adresse à des professionnels de la santé	6047

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
Thromvei	Médecine	Thromboses veineuses profondes	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2782
Troub1	Médecine	... les troubles du rythme constituent un chapitre important de la cardiologie. Ils posent d'importa	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3193
Troub2	Médecine	Dans ce chapitre seront successivement abordés : – Les troubles de la commande sinusale ; – Les ar	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5092
Troub3	Médecine	Bradycardies et troubles de la conduction auriculo-ventriculaire	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3020
Troub4	Médecine	Stimulation cardiaque définitive	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1340
Imt	Médecine	Ischémie myocardique	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	7534
Insuaort	Médecine	Insuffisance aortique	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4146
Insumitr	Médecine	Insuffisance mitrale	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2509
Image6	Médecine	Imagerie des glandes salivaires	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5643
Image5	Médecine	Imagerie des cancers du larynx	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3034
Image3	Médecine	Imagerie de la thyroïde	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2556
Image2	Médecine	Imagerie de la sciatique opérée et récidivante	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2021
Leshyd	Médecine	Les hydrocéphalies	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes	4628

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
			(France).	
Lestra	Médecine	Les traumatismes du massif facial	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2431
patho2	Médecine	Pathologie de l'oreille et des sinus	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	6068
Radiol	Médecine	Radiologie interventionnelle du rachis	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1970
Traum1	Médecine	Traumatismes crânio-encéphaliques	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	5054
Circul	Médecine	Principes élémentaires de la circulation cardio-pulmonaire	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2965
Cflotta	Médecine	Optométrie – La maladie des yeux – Corps flottants et éclairs (flashes de lumière)	Association des Optométristes du Québec – Informations au public, informations générales	758
Cataract	Médecine	Optométrie – La maladie des yeux – Les cataractes	Association des Optométristes du Québec – Informations au public, informations générales	961
Cardicon	Médecine	Cardiopathies congénitales	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2965
Brazeau	Médecine	Problèmes pathophysiologiques en lentilles cornéennes : survol clinique	Association des Optométristes du Québec – Informations aux optométristes (dont la Revue l'Optométriste), p. 9	2882
Anevriss	Médecine	Anévrismes de l'aorte abdominale sous-rénale	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2861
Idm	Médecine	Infarctus du myocarde	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4950
Grefcorn	Médecine	Ophtalmologie – La greffe de cornée	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	788
Glauco	Médecine	Optométrie – La maladie	Association des Optométristes du	894

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
		des yeux – Le glaucome	Québec– Informations au public, informations générales	
Focales2	Médecine	Ophtalmologie – Glaucome – Sclérectomie profonde	Information médicale – Les pages professionnelles	1916
Focales1	Médecine	Les conjonctivites allergiques	Information médicale – Les pages professionnelles	1428
Embolie	Médecine	L'embolie pulmonaire	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2224
Dissecti	Médecine	Dissection aiguë de l'aorte	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3163
Degmac	Médecine	Optométrie – La maladie des yeux – La dégénérescence maculaire	Association des Optométristes du Québec– Informations au public, informations générales	1039
Decore	Médecine	Optométrie – La maladie des yeux – Le décollement de la rétine	Association des Optométristes du Québec– Informations au public, informations générales	752
Cornee	Médecine	Ophtalmologie – La greffe de cornée	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rouen (France).	1881
oph-mg	Médecine	Ophtalmologie – Médecine générale et ophtalmologie	Information médicale – Les pages professionnelles (plusieurs auteurs)	6696
Myocar	Médecine	Les myocardopathies (cardiomyopathies)	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3785
Langism	Médecine	Les lentilles P.A.G. (lentilles perméables au gaz)	Association des Optométristes du Québec – Informations aux optométristes (dont la Revue l'Optométriste)	2099
Insutric	Médecine	Insuffisance tricuspidienn	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	1000
Recherc	Médecine	Les lasers ophtalmiques	Vulgarisation scientifique pour informer le public (publicité)	544
Pericar	Médecine	Affections du péricarde	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2502
Orgelet	Médecine	Optométrie – La maladie des yeux – L'orgelet (ou hordeolum externe), l'hordeolum interne et le chala	Association des Optométristes du Québec– Informations au public, informations générales	434
oph126	Médecine	Ophtalmologie – La rétinopathie diabétique	Corpus médical ophtalmologie. La faculté de médecine de Grenoble	978

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
		(introduction, étude clinique, diagnostic, traitement, co	propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	
opht125	Médecine	Ophtalmologie – Les oblitérations vasculaires rétiniennes (généralités, occlusion de l'artère centra	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	998
opht124	Médecine	Ophtalmologie – Glaucome par fermeture de l'angle (introduction, clinique, traitement)	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	1045
opht123	Médecine	Ophtalmologie – Décollment de rétine (introduction, clinique, traitement)	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	1559
opht122	Médecine	Ophtalmologie – Cataracte (introduction, étude clinique de la cataracte sénile, traitement, les autr	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	1382
opht121	Médecine	Ophtalmologie – Les paralysies oculaires (introduction, diagnostic différentiel, paralysies périphér	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	1000
opht120	Médecine	Ophtalmologie – L'exophtalmie (introduction, exophtalmies endocriniennes, exophtalmies unilatérales)	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type :	1490

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
opht119	Médecine	Ophtalmologie – Pathologie des voies optiques (introduction, atteintes rétiniennes, atteintes du nerf)	Matériel d'enseignement. Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	1670
opht118	Médecine	Ophtalmologie – (Définition – Généralités, Sémiologie clinique, évolution, diagnostic étiologique, d)	Corpus médical ophtalmologie. La faculté de médecine de Grenoble propose l'ensemble de l'enseignement du 2e cycle médical (+ de 250 questions de cours, et 1500 QCM d'évaluation) (site éditeur Université Joseph Fournier) – Type : Matériel d'enseignement.	882
Therape	Médecine	Introduction à la thérapeutique cardiologique	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	3256
Td504	Médecine	Ophtalmologie – Néovascularisation	Didactique. Cours diffusé par deux professeurs.	361
Td503	Médecine	Ophtalmologie – Anomalies microvasculaires (intra-rétiniennes (AMIR), anomalies artérielles et veineuses)	Didactique. Cours diffusé par deux professeurs.	179
Td502	Médecine	Ophtalmologie – Oedème et exsudats secs du pôle postérieur (exsudats cotonneux (nodules dysoriques))	Didactique. Cours diffusé par deux professeurs.	257
Td501	Médecine	Ophtalmologie – Microanévrismes (hémorragies intra-rétiniennes)	Didactique. Cours diffusé par deux professeurs.	251
Rheume	Médecine	Optométrie. Lentilles de contacts – Problèmes de lentilles cornéennes et médicaments thérapeutiques	Association des Optométristes du Québec – Informations aux optométristes (dont la Revue l'Optométriste)	2997
Retr_mit	Médecine	Rétrécissement mitral	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	4872
Retr_aor	Médecine	Le rétrécissement aortique orificiel de l'adulte	Cours diffusé sur Internet à l'intention des étudiants de la Faculté de médecine de l'Université de Rennes (France).	2395
Retr_pig	Médecine	Optométrie – La maladie des yeux – La rétinopathie	Association des Optométristes du Québec – Informations au public,	609

Nom de fichier	Domaine	Sous-domaine	Niveau de spécialisation	Nombre de mots
Ret_diab	Médecine	pigmentaire Optométrie – La maladie des yeux – La rétinopathie diabétique	informations générales Association des Optométristes du Québec– Informations au public, informations générales	1000
Yeuxsecs	Médecine	Optométrie – La maladie des yeux – Les yeux secs (anatomie et physiologie, les problèmes des larmes,	Association des Optométristes du Québec– Informations au public, informations générales	689

ANNEXE B : Acceptions des termes retenus pour le corpus de médecine et pour le corpus d'informatique

Ces tableaux présentent la liste des termes clés retenus, classés par domaine. Nous avons également reproduit une définition, car c'est dans leurs sens techniques que les termes devraient être employés dans les contextes retenus.

CORPUS DE MÉDECINE

Terme	Définition	Source
Angiographie	Technique d'imagerie médicale qui consiste à injecter des produits de contraste dans les vaisseaux sanguins, afin de les opacifier et de les visualiser dans un but diagnostique ou thérapeutique.	GDT
Cathéter	Instrument chirurgical tubulaire servant à évacuer le liquide d'une cavité ou à y injecter un produit, ou à dilater un orifice ou un conduit naturel.	Quevauvilliers et Fingerhut (1999)
Chirurgie	Partie de la thérapeutique médicale qui comporte une intervention manuelle et généralement instrumentale afin de corriger les difformités, soigner les blessures ou diagnostiquer et traiter certaines maladies.	GDT
Décès	Cessation définitive de tout signe de vie se produisant à tout moment après la naissance vivante (arrêt post-natal des fonctions vitales sans possibilité de réanimation).	GDT
Développement	Propriété fondamentale que possèdent les organes ou les êtres vivants de croître (en poids, en volume) et de se modifier jusqu'à ce qu'ils atteignent leur forme parfaite.	Quevauvilliers et Fingerhut (1999)
Douleur	Sensation anormale et désagréable provoquée par l'excitation excessive d'une partie vivante de l'organisme, perçue et intégrée au niveau le plus élevé du cortex cérébral, qui lui confère son caractère de « vécu personnel » et sa tonalité émotionnelle et affective.	GDT
Électrocardiogramme (ECG)	Représentation graphique des signaux électriques émis par le cœur en fonction du temps.	GDT
État	État de santé d'un malade.	Quérin (1998)
Maladie	Altération de l'état de santé, attribuée à des causes internes ou externes, se traduisant par des symptômes et des signes, et se manifestant par une perturbation des fonctions ou par des lésions.	GDT

Terme	Définition	Source
Médecin	Personne qui a suivi avec succès le programme d'études de médecine reconnu dans le pays où elle se trouve et qui remplit les conditions requises pour exercer légalement la médecine.	GDT
Médecine	Science et art du diagnostic, du traitement et de la prévention des maladies humaines.	GDT
Patient	Personne qui a recours aux services médicaux ou paramédicaux, qu'elle soit malade ou non. Le terme <i>patient</i> n'est plus réservé aux malades qui souffrent, comme le voudrait son étymologie latine. Il peut désigner une personne soumise à un examen médical, suivant un traitement ou subissant une intervention chirurgicale de même que les femmes enceintes.	GDT
Scanner	Appareil d'imagerie médicale utilisé en tomodensitométrie et qui comprend notamment un système tomographique, un ordinateur et une console de visualisation.	GDT
Symptôme	Signe général local, physique ou fonctionnel provoqué par la maladie, perçu ou manifesté par le malade, dont l'étude sert à poser le diagnostic.	GDT
Systole	Contraction du muscle cardiaque. La s. simultanée des 2 oreillettes précède celle, également simultanée, des 2 ventricules.	Delamare (1999)
Traitement	Ensemble des moyens hygiéniques, pharmacologiques, chirurgicaux et psychiques mis en œuvre pour combattre la maladie.	GDT
Vieillesse	Processus naturel de transformation graduelle que subit tout organisme vivant du fait de son avancée en âge.	GDT

CORPUS D'INFORMATIQUE

Terme	Définition	Source
Bogue (bug)	Défaut du logiciel ou du matériel se manifestant par des anomalies de fonctionnement.	GDT
Codage	Transformation d'informations au moyen d'un code, ce qui permet leur représentation sous une forme différente et offre la possibilité de revenir à leur forme d'origine.	GDT
Développement	Création, réalisation et mise au point d'un programme, d'un logiciel, d'une application ou d'un système.	GDT
Erreur	Faute d'écriture dans un programme, qui conduit soit à l'impossibilité de traduire ce programme (erreur de syntaxe) soit à l'impossibilité de l'exécuter (erreur sur les données), soit à des résultats aberrants (erreur de programmation).	Larousse (1996)
État	Situation dans laquelle se trouve, à un moment particulier, un dispositif matériel ou logiciel. L'état d'un élément est généralement transmis au programme sous la forme d'une valeur dont la signification détermine la suite du traitement en cours. Certains programmes d'application utilisent des lignes d'état pour afficher des messages à l'intention de l'utilisateur.	GDT
Évolution	Augmentation des performances ou élargissement des fonctions d'un ensemble logiciel ou matériel.	Réseau panlatin de terminologie
Génie	Discipline qui traite de la conception, du développement et de la fabrication de systèmes informatiques (génie informatique). Ensemble des connaissances, des procédés et des acquis scientifiques et techniques mis en application pour la conception, le développement, la vérification et la documentation de logiciels, dans le but d'en optimiser la production, le support et la qualité (génie logiciel).	GDT
Imprimante	Périphérique qui permet de produire des sorties papier de l'information (caractères ou images) provenant de l'ordinateur.	GDT
Informatique	Discipline qui traite de tous les aspects, tant théoriques que pratiques, reliés à la conception, à la programmation, au fonctionnement et à l'utilisation des ordinateurs.	GDT
Langage	Ensemble organisé de symboles, de mots-clés, de caractères et de règles (instructions et syntaxe) utilisé pour adresser des commandes à l'ordinateur et assurer la communication avec la machine.	GDT
Logiciel	Ensemble des programmes destinés à effectuer un traitement particulier sur un ordinateur.	GDT

Terme	Définition	Source
Ordinateur	Machine programmable de traitement de l'information, commandée par des programmes enregistrés en mémoire, qui accepte des entrées structurées et les traite selon des règles établies afin de produire un résultat en sortie.	GDT
Panne	Incident de fonctionnement d'un matériel.	Larousse (1996)
Programmation	Ensemble des activités techniques reliées à l'élaboration d'un programme informatique.	GDT
Programmeur	Spécialiste qui traduit les opérations que l'ordinateur doit effectuer en une instruction que ce dernier peut comprendre.	GDT
Utilisateur	Toute personne faisant appel aux ressources de l'informatique pour mener à bien toutes ou une partie de ses activités. Ce terme est utilisé pour désigner les utilisateurs de ressources informatiques, par opposition aux producteurs de ressources informatiques (consultants, services informatiques, distributeurs de matériels et logiciels, etc.).	GDT

ANNEXE C : Définitions des grands génériques de WordNet

Ci-dessous, nous présentons les définitions de chacun des TS retenus :

1.A [ENTITY ORGANISM] : « That which is perceived or known or inferred to have its own physical existence (living or nonliving) » (WordNet 1.7.1) –Entité vivante qui a sa propre existence physique [ENTITY ORGANISM].

1.B [ENTITY OBJECT] : « That which is perceived or known or inferred to have its own physical existence (living or nonliving) » (WordNet 1.7.1) –Entité non vivante qui a sa propre existence physique.

2. [ABSTRACTION] : « A general concept formed by extracting common features from specific examples » (WordNet 1.7.1) – Concept général créé à partir de l'extraction de caractéristiques générales tirées d'exemples spécifiques.

3. [PSYCHOLOGICAL FEATURE, PSYCHOLOGICAL FEATURE (FIELD)] : « A feature of the mental life of a living organism » (WordNet 1.7.1) – Caractéristique du psychique d'un organisme vivant. Nous avons ajouté l'étiquette *field* aux domaines du savoir afin de les distinguer des autres termes associés au TS [PSYCHOLOGICAL FEATURE].

4. [PHENOMENON³⁶] : « Any state or process known through the senses rather than by intuition or reasoning » (WordNet 1.7.1) – État ou processus perçu par les sens plutôt que par l'esprit.

5. [ACTIVITY] : « Something that people do or cause to happen » (WordNet 1.7.1) – Actes ou travaux de l'être humain.

6. [EVENT] : « Something that happens at a given place and time » (WordNet 1.7.1) – Fait survenant en un lieu et un temps donné.

8. [STATE] : « The way something is with respect to its main attributes; “the current state of knowledge”; his state of health”; in a weak financial state » (WordNet 1.7.1) –État d'une chose ou d'une personne selon sa nature propre.

³⁶ Il semble y avoir disparité entre le tableau de Miller (1998 : 30), présenté à la section 3.5.1, et les chaînes d'hyponymes que l'on trouve effectivement dans WordNet. Il nous paraissait plus logique de considérer les chaînes d'hyponymes de WordNet comme correctes puisque NATURAL PHENOMENON (PHÉNOMÈNE NATUREL) devrait en effet être un hyponyme de PHENOMENON.