

L'apport des corpus spécialisés en terminographie multilingue : le cas des groupes nominaux de type Nom-Adjectif dans la langue médicale

François Maniez

Volume 56, Number 2, June 2011

Les corpus et la recherche en terminologie et en traductologie
Corpora and Research in Terminology and Translation Studies

URI: <https://id.erudit.org/iderudit/1006183ar>

DOI: <https://doi.org/10.7202/1006183ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Maniez, F. (2011). L'apport des corpus spécialisés en terminographie multilingue : le cas des groupes nominaux de type Nom-Adjectif dans la langue médicale. *Meta*, 56(2), 391–406. <https://doi.org/10.7202/1006183ar>

Article abstract

Whether their aim is to describe specialized languages or language in its general use, lexicographers have been using computer corpora for a number of years. In this article, we attempt to show how such corpora can be used for the description of complex noun phrases in which a noun is modified by either an adjective or another noun, as most complex terms correspond to that type of structure in many specialized languages (although we focus specifically on the type of language used in medical research articles). The first part of the article examines to what extent frequency lists that are drawn up using specialized corpora make it possible to consider new term-candidates for inclusion in existing terminological databases. The question of how representative such corpora can be is also examined as we study the representation of terms that are already listed in terminological databases in a corpus that includes more than 5,000 medical research articles written in French. The second part of the article deals with the issue of the choice of an English translation equivalent for such terms, as the head noun may often be modified by either a noun or an adjective. The usefulness of resorting to computer corpora appears to be warranted in both cases.

L'apport des corpus spécialisés en terminographie multilingue : le cas des groupes nominaux de type Nom-Adjectif dans la langue médicale

FRANÇOIS MANIEZ

Université Lumière – Lyon 2, Lyon, France

francois.maniez@univ-lyon2.fr

RÉSUMÉ

La lexicographie, qu'il s'agisse de la description des langues de spécialité ou de celle de la langue générale, s'appuie sur l'usage des corpus informatisés depuis de nombreuses années. Nous choisissons ici de montrer son utilité dans le cadre de la description des groupes nominaux de type Nom-Adjectif, qui constituent l'immense majorité des termes de nombreuses spécialités, le domaine choisi étant celui de la langue des articles de recherche de médecine. La première partie de cet article examine dans quelle mesure les listes de fréquence établies grâce à la consultation des corpus spécialisés permettent de suggérer l'inclusion de nouveaux candidats termes à des bases de données terminologiques existantes, tout en rendant compte de la représentation des termes de ces bases de données dans le corpus constitué pour l'occasion, qui se compose de plus de 5000 articles de recherche médicale de langue française. La deuxième partie traite du problème du choix des équivalents de traduction utilisés en anglais pour les termes qui font intervenir le phénomène de la prémodification par un nom ou un adjectif. L'utilité du recours aux corpus semble prouvée dans les deux cas de figure.

ABSTRACT

Whether their aim is to describe specialized languages or language in its general use, lexicographers have been using computer corpora for a number of years. In this article, we attempt to show how such corpora can be used for the description of complex noun phrases in which a noun is modified by either an adjective or another noun, as most complex terms correspond to that type of structure in many specialized languages (although we focus specifically on the type of language used in medical research articles). The first part of the article examines to what extent frequency lists that are drawn up using specialized corpora make it possible to consider new term-candidates for inclusion in existing terminological databases. The question of how representative such corpora can be is also examined as we study the representation of terms that are already listed in terminological databases in a corpus that includes more than 5,000 medical research articles written in French. The second part of the article deals with the issue of the choice of an English translation equivalent for such terms, as the head noun may often be modified by either a noun or an adjective. The usefulness of resorting to computer corpora appears to be warranted in both cases.

MOTS-CLÉS/KEYWORDS

adjectif relationnel, base de données terminologique, corpus, groupe nominal, langue médicale

relational adjective, terminological database, corpus, noun phrase, medical language

1. Introduction

Près d'un demi-siècle s'est à présent écoulé depuis l'apparition des premiers corpus informatisés totalisant plus d'un million de mots. Dans leur exploitation du *Brown Corpus*, Kučera et Francis (1967) avaient ouvert la voie en montrant la manière dont les statistiques tirées de l'analyse lexicométrique pouvaient apporter un nouvel éclairage sur l'usage de la langue. Le travail effectué en Grande-Bretagne sous l'impulsion de John Sinclair avait dans les deux décennies suivantes mené à la création du *British National Corpus*, qui est longtemps resté la référence des linguistes de corpus qui se sont intéressés à l'usage britannique. C'est l'exploitation de ce corpus qui a notamment conduit Sinclair (1991) à jeter les bases de sa théorie de la lexico-grammaire et lui a permis de décrire de manière novatrice les phénomènes collocatifs de l'anglais moderne, travail théorique qui allait trouver son application pratique dans la réalisation d'ouvrages lexicographiques tels que le dictionnaire COBUILD (Sinclair 1995), précédé par le célèbre dictionnaire des collocations de l'anglais de Benson, Benson, *et al.* (1986). Même si certains linguistes de renom (Chomsky 2004) continuent d'exprimer leurs réticences vis-à-vis de cette nouvelle approche méthodologique qu'est la linguistique de corpus, d'autres (Fillmore 1992) expriment des vues plus nuancées et, tout en dénonçant l'importance exagérée accordée aux données statistiques rendues disponibles par ces nouveaux outils, reconnaissent leur utilité dans le cadre d'une nouvelle manière d'aborder l'étude des faits de langue.

L'utilisation des corpus s'est rapidement imposée dans le domaine de la lexicographie unilingue et bilingue, ainsi que dans celui de la terminographie. L'apport des corpus dans le processus d'extraction terminologique, clairement perçu dès le début des années 1990, a notamment amené les spécialistes du traitement automatique de la langue à se pencher sur les propriétés lexicales et syntaxiques des adjectifs relationnels, éléments-clés de la formation des termes complexes (Daille 2001), la réflexion sur le rôle de l'adjectif dans la formation des termes ayant été poursuivie par Normand et Bourigaud (2001), L'Homme (2004) puis Bowker et Hawkins (2006). Les capacités de stockage informatique allant croissant, l'abondance des données fournies par les grands corpus textuels amène les terminologues à se poser un certain nombre de questions, dont les suivantes : la fréquence d'emploi en corpus est-elle un critère fiable d'inclusion terminographique ? Dans le cas des groupes nominaux candidats termes extraits de façon automatique, existe-t-il des critères fiables d'exclusion des données non pertinentes reposant sur une catégorisation sémantique des noms modifiés par l'adjectif ? Enfin, les corpus spécialisés de grande taille (plusieurs millions de mots) permettent-ils de repérer des candidats termes non encore inclus dans les bases de données terminographiques de référence ? C'est à ces questions que nous tenterons de répondre dans le cadre d'une étude des groupes nominaux de patron Nom-Adjectif dans la langue médicale.

2. Les corpus de langue de spécialité et les bases de données terminologiques

Les principaux outils utilisés pour la première partie de cette étude sont le *Grand Dictionnaire terminologique*¹ (GDT) et le corpus de français médical du CRTT (CRTT-MED). Ce corpus est constitué d'articles de recherche dans le domaine médical tirés des revues disponibles sur la base de données *Science Direct*, et les droits de

reproduction et d'utilisation dans le cadre d'un projet de recherche ont été accordés par les éditions Elsevier pour la quasi-totalité des publications. Le corpus a été étiqueté en partie du discours par l'analyseur Cordial, et compte actuellement vingt-trois millions de mots². Sa conception et son utilisation sont notamment détaillées dans Maniez (2009).

Nous avons dans un premier temps comparé les couvertures respectives du GDT et du corpus CRTT-MED pour les expressions de patron syntaxique Nom-Adjectif formées à partir de l'adjectif *cardiaque* totalisant au moins dix occurrences dans le corpus.

TABLEAU 1

Inclusion dans le GDT des expressions de type <Nom-Adjectif> contenant l'adjectif *cardiaque* dans le corpus CRTT-MED

Fréq.	Expression	GDT			
1839	insuffisance cardiaque	X	33	auscultation cardiaque	
1254	fréquence cardiaque	X	33	patient cardiaque	
866	chirurgie cardiaque	X	32	resynchronisation cardiaque	
341	débit cardiaque	X	29	tamponnade cardiaque	X
308	rythme cardiaque	X	27	lésions cardiaques	
284	échographie cardiaque	X	24	arythmie cardiaque	X
262	arrêt cardiaque	X	23	effets cardiaques	
219	atteinte cardiaque		23	maladie cardiaque	
165	massage cardiaque	X	22	traumatisme cardiaque	
163	fonction cardiaque		21	évaluation cardiaque	
139	complication cardiaque		21	réadaptation cardiaque	
131	index cardiaque	X	20	conduction cardiaque	X
120	défaillance cardiaque		20	souffle cardiaque	X
119	cavités cardiaques		20	travail cardiaque	
119	pathologie cardiaque		20	troubles cardiaques	
105	événement cardiaque		19	plaie cardiaque	
98	décompensation cardiaque		18	muscle cardiaque	X
90	origine cardiaque		18	troponines cardiaques	X
89	risque cardiaque		17	remplissage cardiaque	
72	valve cardiaque	X	16	mortalité cardiaque	
71	enzymes cardiaques	X	16	myocytes cardiaques	
70	dysfonction cardiaque		15	pompe cardiaque	X
58	transplantation cardiaque	X	14	contractilité cardiaque	X
54	activité cardiaque		14	état cardiaque	
53	stimulation cardiaque	X	14	imagerie cardiaque	X
47	cathétérisme cardiaque	X	14	retentissement cardiaque	
47	défibrillateur cardiaque	X	14	silhouette cardiaque	
41	manifestations cardiaques		13	anomalies cardiaques	
39	malformation cardiaque		13	battements cardiaques	
37	cause cardiaque		13	séquelles cardiaques	
37	précharge cardiaque		13	tolérance cardiaque	
36	toxicité cardiaque		12	mastocytes cardiaques	
35	antécédent cardiaque		12	pause cardiaque	
35	greffe cardiaque	X	12	performance cardiaque	
34	greffon cardiaque		11	niveau cardiaque	
			10	accidents cardiaques	

Le tableau 1 fait apparaître que le critère de fréquence est fortement prédictif de l'inclusion terminographique dans le cas des items de très haute fréquence. Les exceptions à cette règle de la fréquence apparaissent toutefois rapidement, la première règle d'exclusion semblant être l'appartenance au sens large à la classe sémantique des états morbides (accident, anomalie, antécédent, atteinte, complication, dysfonction, événement, lésion, maladie, mortalité, risque, séquelle, toxicité, trouble). Ce critère n'est cependant jamais totalement fiable (*déficience cardiaque* est consigné dans le GDT), et peut éventuellement être rendu opaque par certains phénomènes de synonymie. Ainsi, l'anglais *failure* a pour équivalents *défaillance* et *insuffisance*, mais seul *insuffisance cardiaque* est présent en tant qu'expression bilingue dans le GDT, *défaillance cardiaque* apparaissant dans des termes de complexité supérieure (*défaillance cardiaque d'amont*, ~ *d'aval*, ~ *antérograde*, ~ *rétrograde*)³.

Les noms exprimant des liens de causalité (cause, effet, origine, retentissement) peuvent par définition être suivis de n'importe quel type d'adjectif relationnel et ne sont pas non plus présents dans les termes de modèle Nom-Adjectif du GDT. Certains noms d'usage très fréquent (état, niveau) sont également absents de ce type de terme, et les déverbaux (évaluation, performance, remplissage, tolérance) semblent généralement peu propices à la formation de candidats termes. Enfin, l'usage de l'adjectif relationnel pour signifier l'appartenance de certains types de cellules à un organe (*mastocytes/myocytes cardiaques*) ne donne pas non plus lieu à une terminologisation⁴.

À l'inverse, un certain nombre des termes présents dans le GDT sont absents du corpus CRTT-MED, comme le révélera l'examen du tableau 3.

TABLEAU 2

Fréquence des termes du GDT de type <Nom-Adjectif> contenant l'adjectif *cardiaque* dans le corpus CRTT-MED et sur la Toile (recherche effectuée sur Google le 30-09-2010)

Terme <Nom-Adjectif>	CRTT-MED	Toile	Rapport Toile/Corpus
insuffisance cardiaque	1 840	278 000	151,1
fréquence cardiaque	1 219	158 000	129,6
chirurgie cardiaque	876	226 000	258,0
débit cardiaque	345	51 100	148,1
rythme cardiaque	315	373 000	1 184,1
échographie cardiaque	288	24 100	83,7
arrêt cardiaque	262	211 000	805,3
massage cardiaque	166	95 200	573,5
index cardiaque	133	10 700	80,5
valve cardiaque	72	7 510	104,3
enzyme cardiaque	71	142	2,0
transplantation cardiaque	64	27 900	435,9
stimulation cardiaque	53	15 000	283,0
cathétérisme cardiaque	47	23 100	491,5
défibrillateur cardiaque	47	12 200	259,6
greffe cardiaque	36	10 300	286,1
arythmie cardiaque	35	22 300	637,1
tamponnade cardiaque	29	3 060	105,5
conduction cardiaque	21	13 200	628,6
souffle cardiaque	21	10 800	514,3
chirurgien cardiaque	18	7 240	402,2
muscle cardiaque	18	45 800	2 544,4

pompe cardiaque	15	6 780	452,0
contractilité cardiaque	14	1 910	136,4
imagerie cardiaque	14	46 000	3 285,7
prothèse cardiaque	14	3 720	265,7
contraction cardiaque	11	5 120	465,5
réanimation cardiaque	9	4 300	477,8
aire cardiaque	8	1 090	136,3
foie cardiaque	8	1 900	237,5
crise cardiaque	7	339 000	48 428,6
volume cardiaque	7	4 870	695,7
action cardiaque	6	4 630	771,7
bruit cardiaque	6	1 460	243,3
stimulateur cardiaque	6	33 900	5 650,0
tropisme cardiaque	6	409	68,2
cycle cardiaque	5	12 200	2 440,0
révolution cardiaque	5	2 880	576,0
synchronisation cardiaque	5	2 060	412,0
asthme cardiaque	4	1 930	482,5
circulation cardiaque	3	802	267,3
angiographie cardiaque	2	349	174,5
innervation cardiaque	2	844	422,0
plexus cardiaque	2	2 920	1 460,0
réponse cardiaque	2	1 070	535,0
bloc cardiaque	1	5 590	5 590,0
défibrillation cardiaque	1	1 340	1 340,0
éréthisme cardiaque	1	1 030	1 030,0
fibrillation cardiaque	1	2 040	2 040,0
instabilité cardiaque	1	149	149,0

Le tableau 2 indique un taux de couverture relativement élevé du corpus CRTT-MED, puisque l'on y trouve 50 des 79 termes (62%) de patron syntaxique Nom-Adjectif contenant l'adjectif *cardiaque* qui sont consignés dans le GDT⁵. La quatrième colonne du tableau 2 indique également une certaine stabilité du rapport des occurrences Toile/Corpus, qui se situe majoritairement dans la fourchette 100-800 pour les expressions de fréquence supérieure à 10 dans le corpus CRTT-MED (la valeur très faible du rapport pour l'expression *enzyme cardiaque* tient au fait que celle-ci est beaucoup plus souvent employée au pluriel). Les valeurs les plus élevées de ce rapport sont observées dans le cas d'expressions génériques comme *crise cardiaque*, qui ne font pas réellement partie du vocabulaire spécialisé (aucune définition n'est donnée dans le GDT pour ce terme et sa traduction anglaise, *heart attack*), ou de termes comme *rythme cardiaque*, *muscle cardiaque* ou *imagerie cardiaque*, qui sont abondamment utilisés sur les sites de vulgarisation médicale.

TABLEAU 3

Fréquence sur la Toile des termes du GDT de type <Nom-Adjectif> contenant l'adjectif *cardiaque* et absents du corpus CRTT-MED (recherche effectuée sur Google le 30-09-2010)

Terme <Nom-Adjectif>	Toile
déficience cardiaque	18 700
automatisme cardiaque	18 600
télesurveillance cardiaque	13 300
valvule cardiaque	6 170
IRM cardiaque	4 850
tube cardiaque	4 130

blocage cardiaque	1 860
dérive cardiaque	1 330
cirrhose cardiaque	947
syncope cardiaque	916
puissance cardiaque	851
névrose cardiaque	654
cathéter cardiaque	468
estomac cardiaque	370
réflexe cardiaque	319
compensation cardiaque	164
poumon cardiaque	1 130
boucle cardiaque	109
conductibilité cardiaque	59
électrogramme cardiaque	40
rate cardiaque	28
dextrorotation cardiaque	5
routine cardiaque	4
dilatateur cardiaque	3
télémoniteur cardiaque	2
télémonitorage cardiaque	1
synchroniseur cardiaque	1
crochet cardiaque	1
ciseaux cardiaques	1

L'absence du corpus des 29 termes du tableau 3 s'explique par divers facteurs. Certains de ces termes (*crochet cardiaque*, *ciseaux cardiaques*, *synchroniseur cardiaque*, *télémonitorage cardiaque*) étaient des hapax sur la Toile lors de notre recherche (notons que si *télémoniteur cardiaque*, *télémonitorage cardiaque* et *télésurveillance cardiaque* sont absents du corpus, leurs équivalents privés du préfixe *télé-* y figurent et comptent plusieurs milliers d'occurrences sur la Toile). Par ailleurs, il est probable que certains termes d'usage courant (par exemple, *valvule cardiaque*) subiront systématiquement une ellipse adjectivale en contexte et que d'autres seront employés dans leur forme pleine là où seule une forme siglée a été retenue par notre mode de sélection (par exemple, *IRM cardiaque* est absent du corpus, où l'on trouve la forme pleine *imagerie par résonance magnétique cardiaque*). Enfin, certaines pathologies sont rares et il n'est pas surprenant qu'elles ne soient pas mentionnées dans le corpus. C'est en particulier le cas de celles pour lesquelles l'adjectif *cardiaque* est précédé d'un nom d'organe (*poumon cardiaque*, *rate cardiaque*)⁶, l'usage de l'adjectif constituant une hypallage par désignation de l'origine de l'état pathologique de l'organe concerné (*foie cardiaque*, qui compte 1 900 occurrences sur la Toile, est bien présent dans le corpus). Il est intéressant de noter que cette relation causale peut être inversée : ainsi, si la cirrhose cardiaque (stade ultime du foie cardiaque) désigne une variété de cirrhose secondaire à une défaillance cardiaque (illustrant la même relation de causalité que dans les exemples précédents), la névrose cardiaque désigne en revanche des troubles cardiovasculaires consécutifs à une dystonie neurovégétative.

L'examen de ces données semble donc montrer qu'un corpus spécialisé d'une vingtaine de millions de mots peut fournir près des deux tiers des termes de longueur 2 de patron syntaxique <Nom-Adjectif> répertoriés dans le GDT, voire davantage si l'on tient compte du phénomène de l'ellipse adjectivale (ce phénomène d'ellipse expliquant par ailleurs le fait que certains des termes de haute fréquence présents dans le tableau 1 ne figurent pas sous leur forme pleine dans le GDT, les termes *précharge* et *postcharge* y étant présents sans être qualifiés par l'adjectif *cardiaque*)⁷.

Mais si le repérage de ces « faux négatifs » terminologiques (le silence informationnel) est nécessairement long et requiert l'aval des experts, l'élimination des « faux positifs » (le bruit) par l'utilisation des quelques critères mentionnés dans notre commentaire du tableau 1 permet indéniablement une utilisation plus efficace des grands corpus textuels dont l'usage se généralise en terminographie.

3. Utilisation des corpus de langue de spécialité en traduction : le cas de la prémodification en anglais

La structure du nom adjectival pose souvent au traducteur et au terminographe dont le français est la langue cible le problème du choix entre deux structures possibles : celle du complément de nom et celle qui utilisera l'adjectif formé par dérivation à partir du nom français correspondant, si cet adjectif existe. Ce choix est déterminé par l'usage : *cell repository* est traduit par *banque de cellules*, *cell wall* par *paroi cellulaire*, la variation étant rarissime dans le cas de ces deux combinaisons lexicales. Quand un choix existe, il dépend cependant du locuteur et du contexte de communication. Un médecin écrira (et dira dans certains contextes) *cancer mammaire* ou *infarctus myocardique* là où le non-spécialiste parlera de *cancer du sein* et d'*infarctus du myocarde*⁸. Il s'agit là de deux exemples pour lesquels l'anglais ne présente pas réellement de variation, puisque la première expression utilise quasi exclusivement une prémodification nominale (*breast cancer*) et l'autre un adjectif (*myocardial infarction*). Toutefois, on peut supposer que l'adjectif relationnel anglais joue lui aussi ce rôle de démarcation entre le discours du spécialiste et celui du non-spécialiste. La variation entre prémodification par un adjectif (*splenic rupture*) et post-modification prépositionnelle (*rupture of the spleen*) en anglais ne sera toutefois pas étudiée dans le cadre de cet article, faute de place.

La formation de nouveaux adjectifs relationnels français est ainsi souvent provoquée par la nécessité de traduire des expressions venant de l'anglais dans lesquelles on observe un phénomène syntaxique typique des langues germaniques, la prémodification nominale par un nom adjectival. Même si la traduction par le complément de nom est souvent utilisée pour ce type de structure, l'utilisation de l'adjectif offre plusieurs avantages : d'une part, elle confère à l'expression utilisée une consonance plus technique par son aspect figé ; d'autre part, elle évite au traducteur ou au terminologue un choix doublement délicat : celui de la préposition et du type de détermination à utiliser dans le groupe prépositionnel qui traduit le prémodificateur. Ainsi, comme on l'a souligné plus haut, l'adjectif *cardiaque* signifie « consécutif à une défaillance cardiaque » dans *cirrhose cardiaque* alors qu'il signifie « ayant des conséquences au niveau cardiaque » dans *névrose cardiaque*. L'utilisation de l'adjectif relationnel permet donc dans de tels cas une traduction plus fidèle à l'original dans la mesure où l'opacité de la relation entre prémodificateur et nom modifié est préservée.

Afin de comparer l'usage des prémodificateurs de forme nominale et adjectivale en langue générale et en langue de spécialité, nous avons fait porter nos requêtes sur une section et une sous-section du *Corpus of Contemporary American English* (accessible à l'adresse <http://www.americancorpus.org/>, et désigné ci-dessous par le sigle CCAE). Le corpus, qui totalise 410 millions de mots, contient une partie composée exclusivement d'articles de recherche de genre universitaire (nommée *ACADEMIC*), elle-même divisée en plusieurs sous-sections, dont une partie médicale (nommée

ACAD: Medicine) qui totalise un peu plus de 6 millions de mots (les cinq parties principales du corpus, intitulées *Spoken, Fiction, Magazine, Newspaper* et *Academic* contiennent chacune entre 81 et 87 millions de mots). Sa conception et son utilisation sont décrites par son créateur dans Davies (2009). Nous avons fait porter notre comparaison sur la section *Newspaper* (qui compte un peu plus de 83 millions de mots) et la sous-section *ACAD: Medicine*. Nos interrogations du corpus ont porté sur les fréquences comparées dans les sous-ensembles précités des mots heart et cardiac, et sur la fréquence d'emploi de ces deux mots en tant que prémodificateurs d'un nom⁹.

TABLEAU 4

Fréquence comparée de heart et cardiac dans la section *Newspaper* et la sous-section *ACAD: Medicine* du CCAE

word	freq1	freq2	pm1	pm2	pm2/pm1
heart	12 132	885	145,27	147,46	1
cardiac	347	626	4,15	104,30	25
heart NN	4 135	659	49,51	109,80	2
cardiac NN	327	559	3,92	93,14	23

N. B. : *freq1* et *freq2* sont les fréquences absolues respectives; *pm1* et *pm2* sont les fréquences relatives (par million de mots).

L'examen des trois dernières colonnes du tableau 4 révèle que l'utilisation de heart est aussi fréquente dans la section *Newspaper* que dans la sous-section *ACAD: Medicine*, alors que cardiac est 25 fois plus employé que heart dans cette sous-section. Les emplois de heart en tant que prémodificateur y sont deux fois plus nombreux, et ceux de cardiac 23 fois plus nombreux. Mais si l'on compare l'utilisation des deux types de prémodification à l'intérieur du même corpus, on voit que leur fréquence relative est comparable (109,80 et 93,14) alors que la prémodification par heart est 12 fois plus fréquente que par cardiac dans la section *Newspaper*. On peut donc en déduire que l'utilisation de l'adjectif cardiac en tant que prémodificateur est caractéristique de la langue de spécialité médicale, le nom heart étant plus souvent utilisé dans cette fonction dans la langue générale.

3.1. Utilisation de la prémodification par l'adjectif dans les deux corpus

Le tableau 5 donne la liste des 16 expressions de type cardiac NN de fréquence supérieure à 2 dans la section *Newspaper* du CCAE, classées par ordre décroissant de l'indice de spécificité relative ($pm2/pm1$) des deux parties comparées.

TABLEAU 5

Utilisation de la prémodification par l'adjectif cardiac dans la section *Newspaper* du CCAE

Expression	freq1	freq2	pm1	pm2	pm2/pm1
cardiac rehab	5	0	0,06	0,00	5,99
cardiac unit	4	0	0,05	0,00	4,79
cardiac cath	3	0	0,04	0,00	3,59
cardiac arrests	7	1	0,08	0,17	0,50
cardiac care	10	2	0,12	0,33	0,36
cardiac surgeon	4	1	0,05	0,17	0,29
cardiac arrest	152	43	1,82	7,16	0,25

cardiac patients	4	2	0,05	0,33	0,14
cardiac problems	5	3	0,06	0,50	0,12
cardiac rhythm	4	4	0,05	0,67	0,07
cardiac monitor	3	3	0,04	0,50	0,07
cardiac surgery	11	27	0,13	4,50	0,03
cardiac catheterization	6	16	0,07	2,67	0,03
cardiac deaths	4	13	0,05	2,17	0,02
cardiac disease	3	26	0,04	4,33	0,01
cardiac death	3	29	0,04	4,83	0,01

L'examen de la colonne pm2/pm1 est instructif, dans la mesure où seules les trois premières expressions sont d'un emploi relatif plus fréquent dans la section *Newspaper* que dans la sous-section *ACAD: Medicine*. Il s'agit de deux cas d'apocope qui relèvent de la langue familière – cardiac rehab(ilitation) et cardiac cath(erization) – et d'une expression localisante désignant un service hospitalier (cardiac unit) et par conséquent peu employée dans les articles de recherche médicale, où l'on trouverait plus vraisemblablement une forme non elliptique pour la désigner (cardiac care unit). Dans la totalité des autres cas, la fréquence relative des expressions concernées est au moins deux fois supérieure, l'indice pm2/pm1 chutant rapidement en dessous de 10%. Ces données confirment l'utilisation relativement rare de l'adjectif cardiac dans la langue générale. Notons d'ailleurs que d'après les données observées dans la section *Newspaper* du CCAE, la substitution de heart à cardiac est possible pour l'ensemble des expressions du tableau, même si l'emploi de l'adjectif est beaucoup plus fréquent pour la majorité d'entre elles.

Le tableau 6 donne la liste des 19 expressions de type cardiac NN de fréquence supérieure à 4 dans la sous-section *ACAD: Medicine* du CCAE, classées par ordre décroissant de l'indice de spécificité relative (pm1/pm2) des deux parties comparées.

TABLEAU 6

Utilisation de la prémodification par l'adjectif cardiac dans la sous-section *ACAD: Medicine* du CCAE

Expression	freq1	freq2	pm1	pm2	pm1/pm2
cardiac output	44	0	7,33	0,00	733,12
cardiac function	25	0	4,17	0,00	416,54
cardiac arrhythmias	12	1	2,00	0,01	166,98
cardiac dysfunction	10	0	1,67	0,00	166,62
cardiac events	9	0	1,50	0,00	149,96
cardiac death	29	3	4,83	0,04	134,51
cardiac tamponade	8	0	1,33	0,00	133,29
cardiac failure	8	0	1,33	0,00	133,29
cardiac disease	26	3	4,33	0,04	120,60
cardiac complications	7	0	1,17	0,00	116,63
cardiac causes	7	0	1,17	0,00	116,63
cardiac myocytes	7	0	1,17	0,00	116,63
cardiac life	8	1	1,33	0,01	111,32
cardiac valves	5	0	0,83	0,00	83,31
cardiac syndrome	5	0	0,83	0,00	83,31
cardiac monitoring	5	0	0,83	0,00	83,31
cardiac involvement	5	0	0,83	0,00	83,31
cardiac cycle	5	0	0,83	0,00	83,31
cardiac enzymes	5	0	0,83	0,00	83,31

Ces résultats confirment, si besoin était, la fréquence de la prémodification par l'adjectif en langue spécialisée. La quasi-totalité des expressions sont des hapax dans la section *Newspaper*, et l'indice de spécificité est beaucoup plus élevé que dans le tableau 5.

3.2. Utilisation de la prémodification par le nom dans les deux corpus

Le tableau 7 donne la liste des 16 expressions de type heart NN de fréquence supérieure à 10 dans la section *Newspaper* du CCAE, classées par ordre décroissant de l'indice de spécificité relative (pm1/pm2) des deux parties comparées.

TABLEAU 7

Utilisation de la prémodification par le nom heart dans la section *Newspaper* du CCAE

Expression	freq1	freq2	pm1	pm2	pm1/pm2
heart transplant	73	0	0,87	0,00	87,41
heart ailment	46	0	0,55	0,00	55,08
heart ailments	37	0	0,44	0,00	44,30
heart trouble	33	0	0,40	0,00	39,51
heart surgeon	32	0	0,38	0,00	38,32
heart patients	23	0	0,28	0,00	27,54
heart problem	18	0	0,22	0,00	21,55
heart defect	16	0	0,19	0,00	19,16
heart palpitations	16	0	0,19	0,00	19,16
heart pounding	15	0	0,18	0,00	17,96
heart defects	14	0	0,17	0,00	16,76
heart conditions	14	0	0,17	0,00	16,76
heart scan	13	0	0,16	0,00	15,57
heart monitor	13	0	0,16	0,00	15,57
heart operation	12	0	0,14	0,00	14,37
heart medication	11	0	0,13	0,00	13,17

Le fait le plus marquant est l'absence totale des formes concernées de la sous-section *ACAD: Medicine*, qui laisse à penser que l'utilisation de la prémodification par heart donne lieu à une division relativement étanche de la catégorie des noms prémodifiés entre langue générale et langue spécialisée.

Le tableau 8 donne la liste des 16 expressions de type heart NN de fréquence supérieure à 5 dans la sous-section *ACAD: Medicine* du CCAE.

L'enseignement principal à tirer des données de ce tableau est le fait que la prémodification par heart est présente dans les deux types de langue observés, les hapax correspondant à ces formes étant relativement rares dans la section *Newspaper*.

TABLEAU 8

Utilisation de la prémodification par le nom heart dans la sous-section *ACAD: Medicine* du CCAE

	Expression	freq1	freq2	pm1	pm2	pm1/pm2
1	heart rate	164	84	27,33	1,01	27,17
2	heart disease	156	663	25,99	7,94	3,27
3	heart failure	147	301	24,49	3,60	6,80
4	heart association	14	103	2,33	1,23	1,89

5	heart foundation	11	5	1,83	0,06	30,61
6	heart study	10	7	1,67	0,08	19,88
7	heart infusion	9	0	1,50	0,00	149,96
8	heart block	8	0	1,33	0,00	133,29
9	heart trauma	8	0	1,33	0,00	133,29
10	heart surgery	8	130	1,33	1,56	0,86
11	heart attack	8	1 109	1,33	13,28	0,10
12	heart sounds	7	1	1,17	0,01	97,41
13	heart tissue	7	4	1,17	0,05	24,35
14	heart strain	6	0	1,00	0,00	99,97
15	heart thrombi	6	0	1,00	0,00	99,97
16	heart valves	6	27	1,00	0,32	3,09

La fonction de comparaison du module d'interrogation du CCAE permet de se faire une idée plus précise des items lexicaux susceptibles de subir l'un ou l'autre type de prémodification en fonction du type de langue choisi, et permet d'établir les listes ci-dessous.

Section *Newspaper*:

- Noms exclusivement prémodifiés par heart (fréquence supérieure à 10): attack(s), failure, rate, transplant(s), condition, ailment(s), valves, health, problem, defect(s), palpitations, pounding, scan, monitor, murmur, operation, medication;
- Noms exclusivement prémodifiés par cardiac (fréquence supérieure à 2): arrest(s), care, rehab, unit, death(s), cath;
- Noms modifiés au moins deux fois par le nom et l'adjectif: disease, monitor, muscle, patients, problems, rhythm, surgeon, surgery.

Sous-section *ACAD: Medicine*:

- Noms exclusivement prémodifiés par cardiac (fréquence supérieure à 4): arrest, arrhythmias, catheterization, causes, complications, death(s), dysfunction, enzymes, events, life, monitoring, myocytes, output, tamponade;
- Noms exclusivement prémodifiés par heart (fréquence supérieure à 2): attack(s), block, diseases, health, hospital, infusion, involvement, level, rate, sound(s), strain, study, thromb[us/i], tissue, trauma, valve;
- Noms modifiés au moins une fois par le nom et l'adjectif: disease, failure, murmur, surgery, valves.

La comparaison des listes fait apparaître une variation dans la prémodification et permet de déduire l'absence de combinaisons telles que heart monitor, heart muscle, heart patients, heart problems, heart rhythm et heart surgeon de la sous-section *ACAD: Medicine*, combinaisons qui étaient toutes présentes dans la section *Newspaper*. Ces expressions contiennent toutes des noms de fréquence élevée et dont le degré de technicité peut être qualifié de modéré, et leurs équivalents de traduction français utilisent l'adjectif *cardiaque* (heart patients peut également être traduit par l'adjectif substantivé). On peut donc émettre l'hypothèse selon laquelle, lors du processus de traduction du français vers l'anglais en langue de spécialité, les adjectifs spécialisés sont plus fréquemment traduits par des adjectifs que par des noms lorsqu'ils entrent en combinaison avec des noms n'appartenant pas au vocabulaire spécialisé, le rôle de l'adjectif étant peut-être de conférer à l'ensemble de l'expression un plus haut degré de spécialisation.

3.3. Comparaison des deux types de prémodification par le nom dans les deux corpus pour d'autres couples Nom-Adjectif

Pour le couple lung/pulmonary, on ne trouve les deux types de prémodification dans la section *Newspaper* que pour les noms ailment, disease(s), infection, specialist, tests. Les noms cancer(s), capacity, condition, damage, failure, function, power, problems, surgery, tissue, transplant(s), tumors sont modifiés au moins trois fois par lung et jamais par pulmonary¹⁰.

Dans la sous-section *ACAD: Medicine*, on trouve les deux types de prémodification pour un nombre de noms plus élevé. Outre disease(s) et infection, adenocarcinoma, disorders, inflammation, injury, involvement, function, metastasis et nodules sont précédés de l'un ou l'autre prémodificateur. Les noms biopsy, cancer(s), carcinoma, complication, development, fields, infiltration, lesions, scan(s), sections, tissue, tumors, volume(s) sont modifiés au moins trois fois par lung et jamais par pulmonary, les noms cancer(s), tissue et tumors étant les seuls items communs aux deux sections. C'est surtout dans la catégorie des noms modifiés au moins trois fois par pulmonary et jamais par lung que l'on observe la plus grande différence quantitative entre les deux sections étudiées: on y retrouve cinq des six items de la liste de la section *Newspaper* (artery, edema, embolism, fibrosis, hypertension) et 27 autres noms du vocabulaire spécialisé ayant trait pour la plupart au domaine de la pathologie¹¹.

L'étude de la paire kidney/renal révèle des résultats similaires, même si renal est de fréquence moindre que pulmonary dans la section *Newspaper*. Les noms ailments, cancer, damage, dialysis, donor(s), function, infection, patients, problem(s), stone(s), tissue, transplant(s) et trouble y sont exclusivement modifiés par kidney, alors qu'aucun nom n'est modifié uniquement par renal plus de deux fois. Les deux seuls items qui puissent être indifféremment modifiés par le nom et l'adjectif sont disease et failure. Dans la sous-section *ACAD: Medicine*, les seuls noms qui soient exclusivement modifiés par kidney au moins trois fois sont recipients, donors et cells, mais un grand nombre de noms du vocabulaire spécialisé (34) sont modifiés exclusivement par l'adjectif renal¹². On retiendra quelques observations de la comparaison de ces données:

- Un petit nombre de substantifs sont modifiés exclusivement par le nom en langue générale (cancer, function, patient, tissue, transplant)¹³, les mots disease et failure étant les seuls à recevoir systématiquement les deux types de modification;
- Un nombre relativement faible de substantifs peuvent recevoir les deux types de prémodification en langue de spécialité;
- Un nombre conséquent de substantifs sont modifiés exclusivement par l'adjectif en langue de spécialité.

Il reste toutefois à déterminer l'apport que ces données peuvent constituer pour la mise à jour des ressources lexicographiques bilingues existantes, notamment pour la traduction et la rédaction en langue de spécialité.

3.4. Apport potentiel des corpus aux ressources lexicographiques bilingues

Le traducteur/rédacteur pouvant être amené à intervenir dans le domaine de la vulgarisation aussi bien que dans celui d'un domaine de spécialité, une alternative à la solution consistant en la consultation de deux dictionnaires distincts peut être envi-

sagée grâce au support électronique. On peut ainsi imaginer un mode de consultation qui renseignerait l'utilisateur quant à la possibilité des deux types de modification en fonction du type de langue envisagé (générale ou spécialisée), à l'aide d'une structure similaire à celle du tableau 9.

TABLEAU 9

Exemple de structure d'une base de données lexicale renseignant l'utilisateur sur les possibilités de prémodification d'un nom en anglais

			Type de langue			
			Langue générale		Langue de spécialité	
Nom (FR)	Mod. (FR)	Nom (EN)	Adjectif (EN)	Nom (EN)	Adjectif (EN)	Nom (EN)
insuffisance	rénale	failure	renal	kidney	renal	
calcul	rénal	stone		kidney	renal	kidney
lésion	rénale	injury		kidney	renal	
donneur	de rein	donor		kidney		kidney

La structure d'une telle base pourrait éventuellement être affinée par l'apport de détails similaires concernant la nature de la modification en français (adjectif ou complément nominal) en fonction du type de langue envisagé.

D'une manière plus générale, les corpus spécialisés de taille conséquente permettent d'améliorer la couverture des dictionnaires existants. Dans le cas de la prémodification par kidney, les expressions formées à partir des noms suivants sont absentes du GDT : biopsy, cells, damage, dialysis, donor, dysfunction, failure, filtration, graft, maldevelopment, malformation, preservation, recipients, section, sonogram, transplant, transplantation, tumor. La majorité de ces expressions ne sont pas des termes, et leur inclusion dans le GDT ne s'impose donc pas. Toutefois, kidney dialysis, kidney failure, kidney graft, kidney transplant et kidney transplantation sont bien des candidats termes potentiels (kidney failure, kidney graft et kidney transplant sont répertoriés dans la banque de données *TERMIUM Plus*)¹⁴.

Afin d'évaluer l'apport potentiel de l'examen de la sous-section *ACAD: Medicine* pour les expressions utilisant l'adjectif renal, nous avons établi la liste des termes du corpus présents dans au moins l'une de ces deux bases de données terminologiques de référence. Les résultats sont présentés dans le tableau 10¹⁵.

TABLEAU 10

Couverture comparée du GDT et de Termium pour les termes du corpus de type renal N présents dans au moins l'une des deux sources

Nom modifié par renal	GDT	TERMIUM
angiography	X	
artery	X	X
biopsy	X	X
calculus	X	
colic		X
cyst		X
disease	X	X
failure	X	X
function	X	X
glomerulus		X

insufficiency	X	X
lithiasis		X
medulla		X
papilla	X	
pelvis	X	X
perfusion		X
pyramid	X	X
scan	X	X
scarring		X
tubule	X	X
ultrasound		X

Sur les vingt et un termes présents dans le corpus, dix sont communs aux deux sources, huit spécifiques à Terminium et trois au GDT. Précisons cependant que Terminium comporte 36 termes de longueur 2 commençant par renal, et que le GDT en compte 38, ce qui rend leur couverture comparable. Nous avons relevé dans la sous-section *ACAD: Medicine* 12 noms pouvant servir à former des candidats termes potentiels en association avec l'adjectif renal qui sont absents des deux bases de données terminologiques: allograft, angioplasty, azotaemia, capacity, clearance, excretion, hypoperfusion, infarct, insulinase, ischaemia, outflow, parenchyma. Ce choix est bien entendu arbitraire, l'inclusion dans une banque de données terminologique multilingue dépendant de facteurs intrinsèques (place du concept dans l'ontologie du domaine) et extrinsèques (difficulté de traduction potentielle) qui nécessitent une évaluation collective par des linguistes et des experts du domaine concerné. Nous considérons néanmoins que le nombre de candidats termes sélectionnés montre que l'utilisation des corpus spécialisés en terminographie est une base d'enrichissement de telles banques de données.

4. Conclusion

L'examen manuel des données extraites automatiquement des corpus spécialisés grâce à des requêtes utilisant des expressions régulières permet de repérer un nombre important de candidats termes de patron syntaxique Nom-Adjectif en langue de spécialité. La fréquence d'emploi en corpus, qui peut dans certains cas constituer un critère fiable d'inclusion terminographique, ne peut cependant pas être utilisée de manière systématique, certaines collocations ne constituant pas des candidats termes valides. Il est probable que chaque langue de spécialité contient un petit nombre de substantifs pouvant constituer une liste d'exclusion, certains semblent peu propices à la formation de candidats termes. Enfin, les corpus spécialisés de grande taille comme CRTT-MED permettent clairement de repérer des candidats termes non encore inclus dans les bases de données terminographiques de référence, le critère quantitatif ne devant pas occulter le fait que de nombreux termes sont des hapax, y compris dans des corpus de très grande taille. Quant aux corpus bilingues comparables, ils constituent un outil indispensable pour l'appariement multilingue des candidats termes et permettent de rendre compte de l'usage en langue de spécialité en évitant de surestimer le phénomène de l'emprunt à l'anglais souvent constaté dans les corpus traduits. L'apport de la linguistique de corpus s'applique ainsi pleinement au domaine de la terminologie multilingue, auxiliaire indispensable du traducteur spécialisé.

NOTES

1. Le GDT (<<http://www.granddictionnaire.com/>>, consultée le 10 mai 2011) est une ressource désormais incontournable, et dont la couverture dans le domaine médical est impressionnante. La fiabilité des traductions anglaises est systématiquement confirmée par l'usage observé sur la Toile et dans les corpus spécialisés pour les termes de longueur 2, ce qui nous a amené à l'utiliser comme source terminologique bilingue unique dans le cadre de cette étude. Les traductions des termes de longueur supérieure semblent d'une fiabilité moindre, ce phénomène étant l'une des raisons pour lesquelles nous avons limité cette étude aux termes de longueur 2. À titre d'exemple, l'équivalent français donné pour heart sound est correct (*bruit du cœur*), mais le GDT mentionne pour les dix termes anglais de patron syntaxique heart sound NN des traductions contenant la suite *son du/de cœur* dont l'authenticité peut être mise en doute puisqu'elles sont totalement absentes de la Toile.
2. Les diverses spécialités médicales sont à peu près également représentées, comme en témoigne la liste des revues utilisées: *Annales de cardiologie et d'angiologie*, *Annales de chirurgie*, *Annales de chirurgie plastique esthétique*, *Annales françaises d'anesthésie et de réanimation*, *Annales médico-psychologiques*, *Annales de réadaptation et de médecine physiques*, *Annales d'urologie*, *Médecine et maladies infectieuses*, *Néphrologie & Thérapeutique*, *Revue française d'allergologie et d'immunologie clinique*, *Revue de médecine interne*, *Revue du rhumatisme*, *Transfusion clinique et biologique*.
3. Bien que *défaillance* ait été choisi comme équivalent de failure pour ces termes, on dénombre sur la Toile un plus grand nombre d'occurrences des termes *insuffisance cardiaque antérograde* et *insuffisance cardiaque rétrograde* que des équivalents utilisant *défaillance*. En particulier, *insuffisance cardiaque antérograde* est présent dans 6 sources distinctes non dictionnairiques, alors que *défaillance cardiaque antérograde* n'est présent que dans des dictionnaires en ligne, à l'exclusion de tout contexte d'emploi en discours.
4. Le terme *tolérance bactérienne* est inclus dans le GDT, mais la relation entre le nom et l'adjectif (*/tolérance aux bactéries/*) est différente de celle qui existe dans l'expression *tolérance cardiaque*. Quant à *évaluation*, il est absent des termes médicaux bilatéraux du GDT, mais son quasi-synonyme médical *bilan* entre dans la composition de nombreux termes en association avec un adjectif relationnel (*bilan hépatique*, ~ *isotopique*, ~ *lipidique*, ~ *lymphocytaire*, ~ *musculaire*).
5. Seules les expressions vedettes figurent dans ce tableau, à l'exclusion des synonymes (*rhumatisme cardiaque* pour *cardite rhumatismale*), des expressions considérées comme à éviter par le GDT (*gating cardiaque* pour *synchronisation cardiaque*) et des termes relevant d'autres domaines que la médecine (*estomac cardiaque*, qui relève de la zoologie).
6. L'anglais utilise également ce type d'hyppallage (cardiac lung, cardiac liver) mais peut également avoir recours à une prémodification nominale (heart failure spleen). Ce schéma terminologique dans lequel l'adjectif relationnel désignant l'étiologie qualifie une partie du corps atteinte par la maladie est récurrent dans le langage médical (par exemple, le */ped diabétique/*, syndrome pouvant affecter toutes les parties du pied chez le diabétique grave, est appelé diabetic foot en anglais, diabetische Fuß en allemand, piede diabetico en italien et pie diabético en espagnol).
7. Cette ellipse est vraisemblablement le reflet de l'usage; elle est néanmoins surprenante dans le cas de *postcharge*, puisque la présence de l'adjectif permettrait de distinguer la */postcharge myocardique/* de la */postcharge ventriculaire/*.
8. L'emploi de l'adjectif relationnel par opposition au complément de nom n'est toutefois pas systématique dans le discours spécialisé: par exemple, *cancer du poumon* et *bruits du cœur* sont plus fréquemment utilisés que *cancer pulmonaire* et *bruits cardiaques*.
9. Certains des emplois des structures prémodificatives (heart NN et cardiac NN) correspondent à des contextes dans lesquels le terme complexe prémodifie lui-même un autre nom, mais il s'agit de cas relativement rares, notamment dans la partie médicale du corpus. On ne compte que trois expressions de ce type totalisant au mois dix occurrences dans la partie *Newspaper* (heart bypass surgery, heart attack patients, heart attack risk) et une seule dans la sous-section *ACAD:Medicine* (heart rate variability), les expressions de patron syntaxique cardiac NN NN étant de fréquence inférieure à 10 dans les deux cas de figure.
10. Pour un corpus de cette taille, ce seuil semble un bon indicateur, les résultats obtenus sur la Toile confirmant une certaine spécificité à partir des valeurs 3 et 0. À titre d'exemple, la section *Newspaper* contient 2 occurrences de lung complications et aucune de pulmonary complications, qui est pourtant dix fois plus employé sur la Toile. Les noms artery, edema, embolism, fibrosis, hypertension, lab sont modifiés au moins trois fois par pulmonary et jamais par lung.
11. Il s'agit des noms suivants: abnormalities, angiography, artery[ies], aspergillosis, candidiasis, complications, component, dysfunction, edema, embolism, embol[us/], eosinophilia, fibrosis,

- hemorrhage, hemosiderosis, hypertension, infection, koch, lymphangiomatosis, manifestations, medicine, mycosis, oedema, rehabilitation, symptoms, tb, toilet, tree, tuberculosis, ventilation.
12. Il s'agit des noms suivants: allograft, angiography, angioplasty, arter[y/ies], azotaemia, blood, calculi, carcinoma, cell, circulation, clearance, colic, cyst(s), dysplasia, excretion, failure, hypoperfusion, impairment, injury, insufficiency, involvement, medulla, microvasculature, parenchyma, pelvis, perfusion, pyramids, recovery, replacement, scarring, sodium, toxicity, ultrasound, vasodilator.
 13. On relève quelques occurrences de renal transplant en langue de spécialité, où il semble également qu'il y ait variation libre entre renal transplantation et kidney transplantation.
 14. <<http://www.btb.termiumplus.gc.ca>>, consultée le 10 mai 2011.
 15. La comparaison de ces résultats avec ceux que l'on obtiendrait pour les expressions *N du rein* ou *N rénal(e)* n'a pas été effectuée, car nous ne disposons pas pour le français d'un corpus permettant de comparer la langue générale et la langue de spécialité de manière aussi systématique que le CCAE.

RÉFÉRENCES

- BENSON, Morton, BENSON, Evelyn et ILSON, Robert (1986): *The BBI Dictionary of English Word Combinations*. Amsterdam: John Benjamins.
- BOWKER, Lynn et HAWKINS, Shane (2006): Variation in the Organization of Medical Terms, Exploring some Motivations for Term Choice. *Terminology*. 12(1):79-110.
- CHOMSKY, Noam (2004): (Interviewed by Jozsef ANDOR). The master and his performance: An interview with Noam Chomsky. *Intercultural Pragmatics*. 1(1):93-111.
- DAILLE, Béatrice (2001): Qualitative terminology extraction. In: Didier BOURIGAULT, Christian JACQUEMIN et Marie-Claude L'HOMME, dir. *Recent Advances in Computational Terminology. Natural Language Processing*. Amsterdam: John Benjamins, 149-166.
- DAVIES, Mark (2009): The 385+ million word *Corpus of Contemporary American English* (1990-2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*. 14(2):159-190.
- FILLMORE, Charles (1992): Corpus linguistics or computer-aided armchair linguistics. In: Jan SVARTVIK, dir. *Directions in Corpus Linguistics*. Berlin: Mouton, 35-60.
- KUČERA, Henry et FRANCIS, W. Nelson (1967): *Computational Analysis of Present-day American English*. Providence: Brown University Press.
- L'HOMME, Marie-Claude (2004): Adjectifs dérivés sémantiques (ADS) dans la structuration des terminologies. *Actes du colloque Terminologie, ontologie et représentation des connaissances*, Université Jean-Moulin Lyon-3, 22-23 janvier 2004.
- MANIEZ, François (2009): L'adjectif dénominal en langue de spécialité: étude du domaine de la médecine. In: John HUMBLEY, dir. *La terminologie: orientations actuelles. Revue française de linguistique appliquée*. 14(2):117-130.
- NORMAND, Sylvie et BOURIGAULT, Didier (2001): Analysing adjectives used in a histopathology corpus with NLP tools. *Terminology*. 7(2):155-166.
- SINCLAIR, John (1991): *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- SINCLAIR, John (1995): *Collins Cobuild English Dictionary*. London: Harper Collins.