

2m 11.3450.4

Université de Montréal

Adjectifs dérivés de noms : analyse en corpus médical et élaboration d'un
modèle d'encodage terminologique

par

Isabelle Carrière

Département de linguistique et traduction

Faculté des arts et des sciences

Mémoire présenté à la Faculté des études supérieures

en vue de l'obtention du grade de maître

en traduction

option recherche

novembre, 2006

© Isabelle Carrière, 2006



P
25
U54
2007
v.003

Direction des bibliothèques

AVIS

L'auteur a autorisé l'Université de Montréal à reproduire et diffuser, en totalité ou en partie, par quelque moyen que ce soit et sur quelque support que ce soit, et exclusivement à des fins non lucratives d'enseignement et de recherche, des copies de ce mémoire ou de cette thèse.

L'auteur et les coauteurs le cas échéant conservent la propriété du droit d'auteur et des droits moraux qui protègent ce document. Ni la thèse ou le mémoire, ni des extraits substantiels de ce document, ne doivent être imprimés ou autrement reproduits sans l'autorisation de l'auteur.

Afin de se conformer à la Loi canadienne sur la protection des renseignements personnels, quelques formulaires secondaires, coordonnées ou signatures intégrées au texte ont pu être enlevés de ce document. Bien que cela ait pu affecter la pagination, il n'y a aucun contenu manquant.

NOTICE

The author of this thesis or dissertation has granted a nonexclusive license allowing Université de Montréal to reproduce and publish the document, in part or in whole, and in any format, solely for noncommercial educational and research purposes.

The author and co-authors if applicable retain copyright ownership and moral rights in this document. Neither the whole thesis or dissertation, nor substantial extracts from it, may be printed or otherwise reproduced without the author's permission.

In compliance with the Canadian Privacy Act some supporting forms, contact information or signatures may have been removed from the document. While this may affect the document page count, it does not represent any loss of content from the document.

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé :

Adjectifs dérivés de noms : analyse en corpus médical et élaboration d'un modèle
d'encodage terminologique

présenté par :

Isabelle Carrière

a été évalué par un jury composé des personnes suivantes :

Sylvie Vandaele, présidente-rapporteuse
Marie-Claude L'Homme, directrice de recherche
Patrick Drouin, co-directeur
Gilles Bélanger, membre du jury

Résumé

Notre travail vise, dans un premier temps, à isoler les adjectifs dérivés de noms extraits d'un corpus médical pour en désambiguïser le sens au moyen d'un modèle d'encodage destiné aux traducteurs et aux rédacteurs médicaux.

Depuis quelques années, d'autres travaux ont porté sur les adjectifs dans les langues de spécialité (Bae (2006), L'Homme (2004), Maniez (2001) et Zweigenbaum, Hadouche et Grabar (2003)) et leur traitement automatique (Daille (2001), Grabar et Zweigenbaum (2003) et Normand et Bourigault (2001)). Parmi ces travaux, peu ont porté sur l'analyse du sens des adjectifs, particulièrement des adjectifs dérivés de noms. Toutefois, nous avons observé que, contrairement à ce que laissent supposer les définitions généralement données dans les dictionnaires médicaux, le sens de l'adjectif dérivé de nom ne se ramène pas uniquement au sens exprimé par la base nominale de celui-ci. Par exemple, lorsque l'adjectif *abdominal* se combine au nom *abcès*, il signifie « qui se situe dans l'abdomen », et lorsqu'il se combine au nom *chirurgie*, il signifie « qui est effectué sur l'abdomen ». Dans les dictionnaires qui comptent des adjectifs parmi leurs entrées, la définition d'*abdominal* est généralement « qui se rapporte à l'abdomen ».

Pour pouvoir procéder à l'analyse des adjectifs, nous avons mis en forme un corpus de 220 000 mots portant sur l'infectiologie, un sous-domaine de la médecine. Dans un deuxième temps, nous avons procédé au traitement automatique du corpus. Une fois la liste des adjectifs extraits nettoyée, ces derniers ont été triés en ordre décroissant selon leur spécificité à notre corpus d'analyse. Ensuite, nous avons analysé les adjectifs en les classant d'après une typologie lexico-sémantique et nous nous sommes concentrée sur l'analyse du sens des adjectifs dérivés de noms. Au moment de l'analyse, nous avons observé tous les

syntagmes nominaux présents dans notre corpus qui contiennent les adjectifs à l'étude et nous avons été en mesure d'expliquer leur sens au moyen de paraphrases définitoires.

Pour assurer une uniformité, nous avons élaboré un système de paraphrases qui regroupe également les noms de tête et les adjectifs associés à chacune des paraphrases. Le système de paraphrases s'appuie sur deux éléments : 1) la relation entre le nom de tête et le nom de base; et 2) l'utilisation des collocatifs appelés par le type de nom de tête. En adoptant cette approche, nous informons le traducteur et le rédacteur de l'utilisation idiomatique des termes présents dans la base de données terminologiques, nommée *MédiTerm*, où est présenté notre modèle d'encodage.

L'analyse des adjectifs nous a permis d'isoler les adjectifs dérivés de noms et d'observer que les adjectifs spécifiques à notre corpus étaient majoritairement dérivés de noms (88 %). De plus, l'analyse du sens des adjectifs nous a permis de constater que les adjectifs dérivés de noms possèdent en moyenne 5,10 acceptions. Enfin, *MédiTerm* permet à l'utilisateur de procéder à des recherches selon différents critères et d'accéder à des données dont l'encodage facilite la compréhension du sens des adjectifs.

Mots-clés : adjectif dérivé de nom, adjectif relationnel, corpus médical, base de données terminologiques, relation sémantique, analyse sémantique, encodage terminologique, paraphrase définitoire.

Abstract

This research focuses on adjectival derivatives of nouns in medical corpora. Its main goals are to isolate these adjectives once extracted from a corpus, and to analyse their various senses in order to produce a terminological database designed for medical writers and translators.

In recent years, linguists and terminologists have shown a growing interest in the role and the place of adjectives in specialized texts (Bae (2006), L'Homme (2004), Maniez (2001), and Zweigenbaum, Hadouche and Grabar (2003)) and their automated processing (Daille (2001), Grabar and Zweigenbaum (2003), and Normand and Bourigault (2001)). But few studies have been conducted on the semantic analysis of the adjectives, especially those derived from nouns. That being said, we have noticed that these adjectives can carry a richer semantic content than a simple reference to their nominal bases, as indicated in most medical dictionaries. For example, we have found that the adjective *abdominal*, when combined with *abcès* (abcess), means "which affects the abdomen," while when it is combined with *chirurgie* (surgery) means "which is performed on the abdomen." When we look in existing dictionaries that include adjectives, we generally find the definition "related to the abdomen" (my translation).

In conducting our study, we used a corpus of 220,000 words on infectiology, a sub-domain of medicine. From this we automatically extracted a list of candidate adjectives that we reviewed to obtain a list specific to our corpus, sorted according to their specificity. Using the list as a starting point, we classified the adjectives using lexico-semantic criteria and analysed every noun phrase containing adjectival noun derivatives included in the study, in order to define their specific sense using a defining paraphrase.

To ensure that our paraphrases were uniform, we developed a model that groups the nouns combined with an adjective that corresponded to a given meaning. The model was based on: 1) the relation that links the head noun of the noun phrase and the base noun of the adjective; and 2) the use of a verbal collocatives usually referred to by the head nouns. This allows the terminological database user to develop an awareness of the idiomatic use of these terms in medical text.

The study allowed us to isolate the adjectival derivatives of nouns and to observe that the adjectives that were most specific to our corpus were mostly noun derivatives. In addition, by analysing the senses of our adjectives, we discovered that they had an average of 5,10 acceptions each. Finally, we created a database, *MédiTerm*, which allows the user to search for adjectives according to various criteria and to access data encoded in a way that allows a better understanding of their meanings.

Keywords: adjectival derivatives of nouns, relational adjective, medical corpora, terminological database, semantic relation, terminological encoding, defining paraphrase.

Table des matières

Introduction.....	1
Chapitre 1 : Caractéristiques linguistiques des adjectifs et typologie.....	7
1.1 Rôle de l'adjectif dans le discours	7
1.1.1 Caractéristiques syntaxiques	7
1.1.2 Caractéristiques morphologiques	8
1.2 Typologie des adjectifs	9
1.2.1 Typologie lexico-sémantique des adjectifs	10
1.2.1.1 Adjectifs ne faisant pas partie de l'étude	10
1.2.1.2 Adjectifs dérivés de noms	15
1.2.2 Critères de classification des adjectifs	21
Chapitre 2 : État de l'art.....	24
2.1 Place des différentes parties du discours en terminologie	24
2.1.1 Prédominance du nom dans les terminologies	24
2.1.2 Prise en compte du verbe dans les descriptions terminologiques	25
2.1.3 Place des adjectifs en corpus spécialisés.....	27
2.2 Études portant sur les adjectifs spécialisés en médecine	29
2.2.1 Analyse des adjectifs en corpus au moyen d'outils de traitement des langues naturelles	30
2.2.2 Traduction du nom adjectival en anglais médical.....	31
2.2.3 Analyse du sens des termes adjectivaux en coréen.....	34
2.3 Traitement automatique des adjectifs spécialisés	37
2.3.1 Productivité des adjectifs selon le genre et le domaine	38
2.3.2 Relations morphologiques des adjectifs en corpus médicaux.....	40
2.3.3 Identification automatique des adjectifs dérivés de noms.....	44
2.3.4 Distinction entre les termes complexes et les collocations	46
2.4 Contribution de la présente étude.....	48
Chapitre 3 : Méthodologie	50

3.1 Mise en forme du corpus.....	50
3.1.1 Définition	50
3.1.2 Critères de sélection	51
3.1.2.1 Critères de départ	51
3.1.2.2 Critères de sélection des textes	52
3.2 Production d'une liste préliminaire de candidats adjectifs	55
3.2.1 Extraction et spécificité des adjectifs.....	55
3.2.2 Nettoyage de la liste de candidats adjectifs.....	59
3.2.3 Lemmatisation des adjectifs.....	62
3.3 Présentation de la base de données	64
3.3.1 Description des champs de la base de données utilisée lors de l'analyse	65
3.4 Classement et sélection des candidats adjectifs	68
3.4.1 Classification des candidats adjectifs.....	68
a) Adjectifs qualificatifs.....	69
b) Adjectifs déterminatifs.....	70
c) Adjectifs dérivés de verbes	71
d) Adjectifs dérivés de noms	72
3.4.2 Numérotation des entrées.....	76
3.4.3 Sélection des adjectifs analysés	77
3.5 Analyse des adjectifs dérivés de noms.....	77
3.5.1 Étude des contextes au moyen de SYNTEX.....	78
3.5.2 Méthode de distinction des acceptions.....	83
3.5.2.1 Types de noms	83
3.5.2.2 Utilisation de la paraphrase définitoire	86
Chapitre 4 : Données obtenues et modèle d'encodage	92
4.1 Données obtenues	92
4.1.1 Portrait détaillé des adjectifs dans le corpus d'analyse	92
4.1.2 Bases nominales.....	94

4.1.3 Acceptions.....	96
4.1.4 Paraphrases.....	98
4.1.5 Noms de tête.....	101
4.2 Modèle d'encodage	102
4.2.1 <i>MédiTerm</i>	102
4.2.1.1 Champs de la fiche terminologique.....	103
4.2.1.2 Présentation des résultats générés par les recherches	105
Conclusion	110
Bibliographie.....	115
Annexe A : Liste des textes du Corpus	xv
Annexe B : Rapport de lemmatisation manuelle	xxiv
Annexe C : Système de paraphrases	xxvii
Annexe D : Liste des adjectifs analysés.....	xxxvi

Liste des tableaux

Tableau 1.1 Critères de classification des adjectifs	22
Tableau 2.1 Proportion des adjectifs dérivés dans trois corpus spécialisés	28
Tableau 2.2 Échantillon des patrons syntagmatiques du nom adjectival <i>cell</i> (Maniez 2001 : 64)	33
Tableau 2.3 Allomorphes regroupés automatiquement puis complétés manuellement (Grabar et Zweigenbaum 2003 : 111)	39
Tableau 2.4 Suffixes utilisés dans les règles de transformation (Daille 2001 : 157).....	45
Tableau 2.5 Dénombrement des patrons extraits du corpus d'agriculture (AGRO) (Daille 2001 : 159 [trad. libre])	46
Tableau 3.1. Critères de sélection des textes du corpus.....	54
Tableau 3.2 Patrons syntaxiques des paraphrases définitives des adjectifs dérivés de noms	87
Tableau 4.1 Répartition des adjectifs analysés	93
Tableau 4.2 Liste des adjectifs dérivés de noms analysés	94
Tableau 4.3 Liste des bases nominales	95
Tableau 4.4 Bases nominales des adjectifs dérivés de noms	96
Tableau 4.5 Liste des bases nominales produisant deux adjectifs	96
Tableau 4.6 Acceptions des adjectifs analysés	97
Tableau 4.7 Paraphrases définitives produites lors de l'analyse.....	98
Tableau 4.8 Échantillon du système de paraphrases.....	100
Tableau 4.9 Adjectifs qui se combinent au nom <i>infection</i>	101

Liste des figures

Figure 3.1 Rappel des étapes de traitement du corpus en vue de l'extraction des candidats adjectifs	57
Figure 3.2 Liste des candidats adjectifs, triés selon leur spécificité, générée par l'interface Internet de TermoStat.....	58
Figure 3.3 Contextes fournis par TermoStat pour l'adjectif <i>clinique</i>	59
Figure 3.4 Table de la base de données utilisée au moment de nettoyer et de lemmatiser la liste de candidats adjectifs.....	60
Figure 3.5 Échantillon du <i>Rapport de lemmatisation manuelle</i> (le Rapport est présenté en entier à l'annexe B).....	63
Figure 3.6 Formulaire <i>Types d'adjectifs</i>	65
Figure 3.7 Formulaire <i>Contextes</i>	67
Figure 3.8 Exemple d'une fiche <i>Types d'adjectifs</i> pour un adjectif qualificatif.....	70
Figure 3.9 Exemple d'une fiche <i>Contextes</i> pour un adjectif qualificatif	70
Figure 3.10 Exemple d'une fiche <i>Types d'adjectifs</i> pour un adjectif dérivé de verbe.....	72
Figure 3.11 Exemple d'une fiche <i>Contextes</i> pour un adjectif dérivé de verbe	72
Figure 3.12 Exemple d'une fiche <i>Types d'adjectifs</i> pour un adjectif dérivé de nom.....	73
Figure 3.13 Exemple d'une fiche <i>Contextes</i> pour un adjectif dérivé de nom	74
Figure 3.14 Exemple d'une fiche <i>Types d'adjectifs</i> pour un adjectif ayant un lien sémantique ténu avec un nom	75
Figure 3.15 Exemple d'une fiche <i>Contextes</i> pour un adjectif ayant un lien sémantique ténu avec un nom	75
Figure 3.16 Exemple de la numération des acceptions de l'adjectif <i>alimentaire</i>	76
Figure 3.17 Formulaire de recherche dans les sorties Access de SYNTAX	79
Figure 3.18 Liste de résultats de la recherche de l'adjectif <i>viral</i>	81
Figure 3.19 Liste de résultats de <i>viral</i> comme adjectif en expansion de syntagme.	81
Figure 3.20 Contexte qui contient le syntagme <i>infection virale</i>	82
Figure 4.1 Fiche de l'acception <i>viral la</i> dans <i>MédiTerm</i>	103

Figure 4.2 Fenêtre de définition de la base nominale <i>virus</i> dans <i>MédiTerm</i>	105
Figure 4.3 Formulaire de recherche de <i>MédiTerm</i>	106
Figure 4.4 Liste de résultats générée par une requête selon l'adjectif <i>muqueux</i>	107
Figure 4.5 Liste de résultats générée par une requête selon la base <i>air</i>	108
Figure 4.6 Liste de résultats générée lors d'une requête d'après le nom de tête <i>chirurgie</i>	109

À une mère formidable,

Remerciements

Je souhaite d'abord remercier ma directrice de recherche Marie-Claude L'Homme, pour sa grande générosité, son infinie patience et ses conseils inestimables. Elle a fait preuve de beaucoup de confiance et de passion face à mon travail. Au fil des mois, elle m'a appris énormément tant d'un point de vue professionnel que personnel.

Je remercie également mon co-directeur de recherche Patrick Drouin qui a fait preuve d'une grande disponibilité et d'une oreille attentive tout au long du projet. Il m'a également été d'une aide indispensable lors de la préparation de mon corpus et des manipulations techniques.

Je veux également souligner ma reconnaissance envers Didier Bourigault qui a généreusement contribué au traitement automatique de notre corpus.

Je ne pourrais passer sous silence le soutien précieux que m'ont apporté Philippe Caignon, Agnès Tutin et Olivier Kraif. Vous comptez parmi les professeurs qui font une différence dans la vie des étudiants. De plus, Je veux souligner la générosité et le soutien constant d'Hee Sook Bae et d'Elizabeth Marshman, deux collègues et amies qui ont su marquer mes deux dernières années.

Enfin, je remercie ma famille et mes amis qui m'ont soutenue et qui ont su faire preuve d'une patience et d'une compréhension sans borne pendant les dernières années. Il y a un peu de vous dans mon travail et je ne serais pas rendue si loin sans votre présence dans ma vie.

Introduction

À l'ère d'Internet, une pléthore de ressources terminologiques est à la portée des traducteurs. Par contre, lorsque vient le temps pour ceux-ci de comprendre la terminologie utilisée dans un texte, la rareté des adjectifs, ainsi que des verbes, dans les ressources terminologiques rend parfois la recherche terminologique ponctuelle difficile. Notre recherche porte sur une partie du discours qui suscite de plus en plus d'intérêt en terminologie : les adjectifs, plus précisément les adjectifs dérivés de noms (ex. *bactérien* qui est dérivé du nom *bactérie*). Notre recherche s'inscrit dans un projet de recherche plus vaste portant sur la description des adjectifs en corpus spécialisés¹.

Pour traduire un texte spécialisé, il convient de surmonter les obstacles que pose la langue de spécialité. L'apprentissage de cette dernière est crucial à la compréhension du texte de départ et à la rédaction du texte d'arrivée. Parmi les difficultés qu'elle présente, nous comptons, entre autres, les lacunes des ressources terminologiques utilisées par les langagiers (Lethuiller 2003 : 139). Par exemple, prenons la liste de combinaisons nom-adjectif proposée par une base de données terminologiques pour un adjectif. Elle est souvent longue sans nécessairement contenir l'expression recherchée. Les traducteurs sont alors tentés de déduire le sens du syntagme qui pose problème et ils risquent de commettre un faux sens (« faute de traduction résultant d'une mauvaise appréciation du sens d'un mot ou d'un énoncé dans un contexte donné » (Delisle 1993 : 31)). De plus, nous pouvons souligner que les bases de données terminologiques contiennent rarement les adjectifs spécialisés seuls, car ces derniers participent souvent à la construction des termes complexes (Maniez 2002).

Dans le cadre de la présente étude, nous nous attardons uniquement aux adjectifs médicaux. La médecine est un domaine qui fait appel à de nombreux

¹ *Étude du rôle des adjectifs dans la structuration des terminologies* : projet mené par le groupe ÉCLECTIK de l'OLST (Observatoire de Linguistique Sans-Texte) sous la supervision de Marie-Claude L'Homme, qui vise à décrire la fonction des adjectifs (qualificatifs et dérivés sémantiques) dans la structuration des terminologies.

adjectifs. Nous croyons que leur encodage sous forme de base de données électronique permettra de répondre à certains besoins terminologiques dont :

- 1) Le regroupement, dans un dictionnaire ou une ressource terminologique, des adjectifs qui sont essentiels à la rédaction des textes médicaux;
- 2) La désambiguïsation des adjectifs dérivés de noms en expliquant leur sens, d'une part, par rapport au nom dont ils sont dérivés et avec lequel ils partagent des composantes sémantiques et, de l'autre, par rapport aux noms qu'ils modifient.

La nature descriptive des sciences médicales (ex. anatomie, physiologie, infectiologie) exige l'utilisation d'adjectifs spécialisés. Par exemple, l'infectiologie décrit l'origine, les symptômes et le traitement des maladies bactériennes, virales, fongiques ou causées par tout autre agent pathogène. Il est, entre autre, question de décrire la caractérisation de symptômes ou d'états physiologiques (ex. *fièvre hémorragique, abcès abdominal*). Au moment de produire des textes médicaux, les rédacteurs et les traducteurs font appel aux adjectifs pour décrire un état anatomique, physiologique ou pathologique. Dans bien des cas, seules les connaissances en médecine peuvent permettre d'utiliser adéquatement la plupart des adjectifs médicaux.

Un grand nombre d'adjectifs présents dans les textes médicaux sont construits à partir d'un nom (ex. *symptomatique* qui est dérivé de *symptôme*) et ils ont retenu l'attention de chercheurs comme Daille (2001), Grabar et Zweigenbaum (2003), L'Homme (2004), Maniez (2001-2002), Normand et Bourigault (2001) et Zweigenbaum, Hadouche et Grabar (2003). Les adjectifs se caractérisent par leur capacité à établir une relation entre leur base nominale et le nom auquel ils sont combinés. Par exemple, dans le syntagme nominal *patient symptomatique*, une relation (en gras) est observée entre le nom en tête du syntagme (*patient*) et le nom de base de l'adjectif (*symptôme*) : *le patient qui manifeste des symptômes*. Par contre, lorsque l'adjectif fait partie du syntagme *médicament symptomatique*, la relation change : *le médicament qui traite les symptômes*. Cette caractéristique des adjectifs spécialisés a amené certains chercheurs à souligner le besoin de désambiguïser le sens de l'adjectif

(L'Homme 2004a : 159). La présentation des caractéristiques des différents types d'adjectifs, notamment celles des adjectifs dérivés de noms, est faite au chapitre 1.

Chercher un adjectif dans une base de données terminologiques constitue une tâche parfois ardue. Ce dernier fait rarement l'objet d'une entrée, mais il apparaît dans une liste, exhaustive, de termes complexes (L'Homme 2003 : 405). Par exemple, dans le Grand dictionnaire terminologique (GDT), l'adjectif *bactérien* possède une entrée dans la base de données, mais sa définition ne reflète pas toujours son sens précis, car il ne fait que renvoyer à la base nominale, *bactérie*. Toutefois, il apparaît également dans 114 termes complexes (ex. *action bactérienne*, *capsules bactériennes*, *charbon bactérien*, etc.) (OQLF 2006 : s.v.). L'observation d'un tel nombre d'occurrences appuie l'importance que nous donnons aux adjectifs dans les textes médicaux.

Parmi les dictionnaires médicaux, rares sont ceux qui consacrent des articles spécifiques aux adjectifs. Prenons l'exemple du *Dictionnaire médical* (Manuila, Manuila et al. 2001), qui est parmi les plus riches en adjectifs. Un certain nombre d'entre eux possèdent leur propre article (ex. *abdominal*, *pulmonaire*, *gastrique*), mais la plupart ne sont mentionnés qu'à la fin de la définition du nom dont ils sont dérivés (ex. l'adjectif *leucocytaire* est donné sous *leucocyte* (Manuila, Manuila et al. 2001 : 277)). De plus, lorsque l'adjectif possède sa propre entrée, la définition est plutôt générale et elle ne permet pas de distinguer les différentes acceptions qui apparaissent lorsque l'adjectif est en contexte. En effet, dans le *Dictionnaire médical* de Manuila (2001 : s.v.), l'adjectif *abdominal* est défini par « qui se rapporte à l'abdomen ». Parallèlement, dans *TERMIUM Plus*, la base de données terminologiques du Bureau de la traduction, à Ottawa, l'adjectif *abdominal* possède sa propre fiche, mais la définition qui est offerte se limite également à « qui se rapporte à l'abdomen ». Or, une analyse approfondie de l'adjectif *abdominal* révèle des subtilités qui ne sont pas décrites dans les deux répertoires cités. Par exemple, dans le syntagme *foyer abdominal*, l'adjectif signifie « qui se situe dans l'abdomen », dans *infection abdominale*, « qui affecte l'abdomen » et dans *échographie abdominale*, « qui est effectuée sur l'abdomen ».

Nous espérons, grâce à la présente étude, élaborer un modèle d'encodage pour les adjectifs dérivés de noms qui permette de combler certaines lacunes que présentent les dictionnaires et les bases de données terminologiques spécialisées. Une telle ressource terminologique pourrait faciliter l'apprentissage des langues de spécialité à des fins traductionnelles et améliorer les connaissances terminologiques des traducteurs qui désirent se spécialiser en traduction médicale.

Pour mener à bien notre étude, nous isolons les adjectifs dérivés de noms, présents dans un corpus médical, des adjectifs qualificatifs, déterminatifs et dérivés de verbes. L'identification des adjectifs dérivés de noms est une étape essentielle dans notre méthodologie. De plus, cette étape nous a permis d'évaluer, sur la base de leur spécificité, la place occupée par ces adjectifs dans un corpus médical.

Deuxièmement, nous analysons le lien qui unit les adjectifs dérivés aux noms qu'ils modifient. Cette analyse nous permet de bien saisir les nuances sémantiques présentes dans notre corpus pour un même adjectif. Au moyen d'un système de paraphrases, nous dressons une liste des acceptions de chaque adjectif. Ce tableau nous permet d'uniformiser la rédaction des paraphrases définitives.

Troisièmement, à l'aide des résultats de l'analyse, nous élaborons un modèle d'encodage qui permet de désambiguïser les adjectifs. Nous croyons également que le modèle élaboré encouragera l'utilisation des adjectifs appartenant à ce groupe.

Accessoirement à l'objectif général de désambiguïisation des adjectifs dérivés de noms, nous croyons que la présence de ceux-ci dans une ressource terminologique et un encodage permettant de visualiser les adjectifs dans les syntagmes nominaux et dans les termes complexes encourage leur utilisation. En effet, une meilleure connaissance de ce groupe d'adjectifs permet aux traducteurs d'alléger leurs traductions tout en produisant des textes idiomatiques. Par exemple, si un traducteur peut trouver l'adjectif *symptomatique* en effectuant une recherche à partir du nom de

base *symptôme*, il lui est alors possible de remplacer un segment de phrase (en gras) par l'adjectif approprié. En d'autres mots, nous pouvons remplacer *Le patient qui manifeste des symptômes est placé en isolation* par *Le patient symptomatique est placé en isolation*. Nous n'observons aucune perte de sens, car l'adjectif dérivé de nom *symptomatique* signifie « qui manifeste des symptômes » lorsqu'il est combiné à un être animé comme *patient*.

Un modèle d'encodage est proposé à la suite de l'étude des adjectifs extraits d'un corpus médical français de 220 000 mots dont les textes portent sur l'infectiologie, un sous-domaine de la médecine. Au moment de l'analyse, nous avons étudié l'adjectif en contexte pour identifier les adjectifs dérivés de noms. Pour y parvenir, nous avons d'abord procédé à la reconnaissance de la base nominale sur la base d'une ressemblance morphologique, du sens de la base nominale et du sens général de l'adjectif. Dans le cadre de notre projet, nous avons effectué cette portion d'analyse en observant les adjectifs en contexte et nous avons constaté que le changement de contexte permet parfois d'observer plusieurs bases. Par exemple, l'adjectif *digestif* possède deux bases :

- 1) digestion : *infection digestive* (« qui affecte la digestion »)
- 2) appareil digestif : *flore digestive* (« qui se situe dans l'appareil digestif »)

Le présent mémoire est divisé en quatre chapitres.

Le premier chapitre présente les différents concepts sur lesquels repose l'analyse des adjectifs de notre corpus. Il est principalement question des différents types d'adjectifs, ainsi que des caractéristiques que nous avons retenues pour chacun d'entre eux au moment de les classer et, par le fait même, d'isoler les adjectifs dérivés de noms.

Au deuxième chapitre, nous effectuons un survol des travaux qui ont été faits sur les adjectifs spécialisés. Nous nous attardons principalement à l'étude des adjectifs dérivés de noms, à leur identification au moyen de leurs caractéristiques grammaticales et morphologiques, ainsi qu'à leur place dans les corpus médicaux. Dans un deuxième temps, nous survolons les travaux effectués, dans différents domaines, sur le traitement automatique des adjectifs en corpus spécialisés.

Au chapitre trois, nous présentons la méthodologie que nous avons adoptée pour recueillir les données nécessaires à l'élaboration de notre modèle d'encodage. Parmi les principales étapes de notre méthodologie, nous comptons la mise en forme de notre corpus médical, le traitement automatique de celui-ci pour dresser une liste d'adjectifs, la classification de ces derniers pour isoler les adjectifs dérivés de noms et, enfin, l'analyse du sens des adjectifs dérivés de noms dans le but d'encoder les résultats sous forme de base de données terminologiques.

Le chapitre quatre présente d'abord les données chiffrées qui nous permettent d'évaluer la proportion des adjectifs, plus précisément des adjectifs dérivés de noms, dans notre corpus. Ensuite, nous présentons un modèle d'encodage électronique, qui est sous forme de base de données.

Chapitre 1 : Caractéristiques linguistiques des adjectifs et typologie

Avant d'entreprendre la lecture des sections portant sur les adjectifs spécialisés, il est essentiel de présenter les concepts généraux auxquels fait appel notre travail. En effet, cet exercice nous permet de préciser la terminologie ainsi que les critères que nous avons adoptés. Ce chapitre brosse d'abord un portrait de l'adjectif au sens large de façon à en expliquer le rôle dans le discours. Ensuite, nous nous penchons sur les types d'adjectifs afin de présenter les critères qui nous permettent d'identifier les adjectifs qui font l'objet de cette étude, à savoir les adjectifs dérivés de noms.

1.1 Rôle de l'adjectif dans le discours

L'adjectif est la partie du discours qui « groupe des éléments dont le trait commun est d'apparaître de façon facultative dans le syntagme nominal » (Arrivé, Gadet et al. 1986 : 32). En fait, l'adjectif est rarement utilisé seul et il sert soit à exprimer la qualité d'un concept exprimé par un nom, soit à marquer une relation entre le nom qu'il modifie et la base de l'adjectif (voir section 1.2). Dans cette section, nous présentons les principales caractéristiques syntaxiques et morphologiques des adjectifs.

1.1.1 Caractéristiques syntaxiques

L'adjectif peut faire partie d'un syntagme nominal (*adjectif épithète*) ou verbal (*adjectif attribut*).

a) Adjectif épithète

L'épithète est généralement, mais non exclusivement, située après le nom qu'elle modifie et elle accepte parfois les variantes temporelles lorsqu'il s'agit d'un adjectif qualificatif (Arrivé, Gadet et al. 1986 : 36). Par exemple, il est juste de lire *infection soudainement fatale*, alors qu'il est inusité de lire *l'infection soudainement bactérienne*, car l'adjectif *bactérien* ne fait que préciser la nature de l'infection².

b) Adjectif attribut

L'attribut est généralement précédé d'un verbe d'état (ex. *être, sembler, devenir, rester*) et il subit l'effet temporel qui est imposé par le verbe (Arrivé, Gadet et al. 1986; Noailly 1999). Dans le syntagme verbal *l'infection devient grave*, le verbe *devenir* ajoute un élément temporel à la qualification apportée par l'adjectif *grave*.

c) Ambiguïté quant à la fonction

Au moment de classer l'adjectif selon l'une ou l'autre des fonctions grammaticales décrites ci-dessus, il arrive que certains éléments portent à confusion. En effet, un adjectif en apposition (ex. *le randonneur, blessé, a terminé l'excursion*) risque de poser problème (Arrivé, Gadet et al. 1986 : 39), car il peut sous-entendre un verbe d'état (ex. *le randonneur, qui a été blessé, a terminé l'excursion*) qui en fait un attribut, ce qui souligne l'importance temporelle de la blessure. Le même adjectif peut également être épithète sans l'apposition (ex. *le randonneur blessé a terminé l'excursion*). Dans ce dernier cas, l'adjectif précise l'état du randonneur sans nécessiter l'ajout d'un verbe, ce qui en fait une épithète.

1.1.2 Caractéristiques morphologiques

L'adjectif peut se reconnaître au moyen de différentes caractéristiques morphologiques. Parmi les formes adjectivales, nous comptons (Noailly 1999 : 11) :

² Les exemples qui ne sont pas précédés ou suivis d'une citation sont tirés de notre corpus.

- a) Les adjectifs primaires : adjectifs non suffixés (ex. *grand, vieux, bleu*).
- b) Les adjectifs dérivés : adjectifs construits à partir d'un nom ou d'un verbe auquel un suffixe (*-el, -al, -ien, -esque, -ique, -iste, -able et -ible*) a été ajouté (ex. *éléphantesque, abominable, viral*).
- c) Les adjectifs par conversion : noms (ex. *spécialiste, imperméable*), verbes conjugués (ex. *absorbant*) ou adverbes (ex. *bien*) qui changent de partie du discours pour devenir des adjectifs.

De plus, il est possible de différencier les adjectifs à forme simple ou complexe. Les adjectifs à forme simple regroupent les adjectifs primaires (ex. *vide, lent*), ceux qui sont dérivés d'un nom (ex. *virus – viral*), d'un verbe (ex. *aimer – aimable*) ou d'un autre adjectif (ex. *triste – tristounet*). Pour leur part, les adjectifs complexes sont des adjectifs composés de deux adjectifs variables (ex. *blond-roux*) ou d'un adjectif invariable qui joue le rôle d'adverbe suivi d'un adjectif variable (ex. *nouveau-né*) (Riegel, Pellat et al. 2004 : 357).

1.2 Typologie des adjectifs

Comme nous l'avons vu dans les sections précédentes, la partie du discours adjectivale regroupe des membres hétérogènes. D'après Noailly (1999 : 9), « la catégorie adjectivale est composée d'un noyau central autour duquel gravitent des éléments satellites, plus difficiles à intégrer à l'ensemble ». Cette analogie illustre le besoin de faire la lumière sur la typologie des adjectifs qui envisage ceux-ci sous les angles syntaxique (section 1.1.1), morphologique (section 1.1.2) et lexico-sémantique.

Nous avons choisi de privilégier une typologie lexico-sémantique qui s'arrête davantage à « la nature [du] contenu notionnel » des adjectifs (Noailly 1999). Elle s'appuie sur les caractéristiques qui permettent de comprendre le sens des adjectifs. Cette typologie permet de distinguer quatre types d'adjectifs.

1.2.1 Typologie lexico-sémantique des adjectifs

Dans cette section, nous présentons les principaux types d'adjectifs retenus pour effectuer notre classification lexico-sémantique. Nous commençons par présenter les caractéristiques des adjectifs qui ne font pas partie de l'étude et, par la suite, nous procédons à une présentation plus approfondie des adjectifs dérivés de noms.

1.2.1.1 Adjectifs ne faisant pas partie de l'étude

Cette section est une étape importante dans le processus de classification des adjectifs, car elle sert de base à la différenciation des types d'adjectifs. Même si nous n'analysons pas ces premiers types d'adjectifs, nous devons être en mesure de les identifier de façon à les classer avec justesse. Parmi ceux-ci, nous comptons les adjectifs qualificatifs, les adjectifs déterminatifs et les adjectifs dérivés de verbes.

1.2.1.1.1 Adjectifs qualificatifs

Au moment de définir les adjectifs qualificatifs, nous avons constaté qu'ils sont souvent repérés de façon intuitive. Ce phénomène se justifie par leur caractère descriptif et leur faible autonomie référentielle (Riegel, Pellat et al. 2004 : 356). Par contre, ces deux caractéristiques sémantiques compliquent sérieusement l'application de critères de classification. Ceux-ci peuvent servir d'indices pour vérifier l'appartenance des adjectifs à ce type. Parmi les critères de classification des adjectifs qualificatifs, les plus fréquemment évoqués dans la littérature sont : la possibilité d'occuper les fonctions d'attribut et d'épithète, leur capacité de coordination avec un autre adjectif qualificatif et l'acceptation d'une modification par un adverbe de degré ou un complément de l'adjectif.

a) Possibilité d'occuper les fonctions d'attribut et d'épithète

Lorsque l'adjectif est utilisé comme attribut, le patron syntaxique (soit la structure dans laquelle il se présente habituellement dans un texte) est facile à reconnaître, car il se présente généralement dans la forme *nom sujet + verbe d'état + adjectif* (Arrivé, Gadet et al. 1986). Par exemple, nous observons une telle structure dans la phrase *L'œdème est absent ou discret*. Toutefois, la présence d'un patron propre à l'attribut cache parfois un adjectif dérivé de nom. En effet, il arrive parfois qu'un adjectif dérivé de nom, qui rejette habituellement le statut d'attribut, soit placé après un verbe. La phrase sous-entend alors une répétition du nom sujet qui est identique au nom de tête du syntagme nominal auquel appartient l'adjectif. Dans un tel cas, la vraie modélisation de la phrase est la suivante : *nom sujet + verbe d'état + nom (répétition du sujet) + adjectif* (Monceaux 1997). Par exemple, *le malaise est [un malaise] cardiaque*. En présence d'une telle ambiguïté, il convient d'une part d'appliquer les deux autres critères de classification qui suivent puis vérifier l'applicabilité des critères des adjectifs dérivés de noms (voir section 1.2.1.2).

Lorsque l'adjectif est utilisé comme épithète, il fait partie d'un syntagme nominal et, tout comme l'attribut, il désigne une propriété du nom qu'il modifie. L'adjectif ajoute alors un élément descriptif au nom tout en restant facultatif à la syntaxe de la phrase (Noailly 1999 : 9). Il est habituellement situé après le nom de tête (ex. *infection aiguë*) et il est possible de remplacer le syntagme nominal par un segment de phrase dont la structure est la suivante : *tête du syntagme (sujet) + qui + verbe d'état + adjectif* (Arrivé, Gadet et al. 1986 : 36). Par exemple, nous pouvons remplacer le syntagme qui est complément (en gras) comme suit :

L'œdème inflammatoire est un phénomène actif

L'œdème inflammatoire est un phénomène qui est actif

b) Coordination avec un autre adjectif qualificatif

Le critère, tel que présenté par Noailly (1999), est la coordination possible avec un autre adjectif de même nature, ici un adjectif qualificatif. Par exemple, il est

possible d'obtenir la phrase suivante : *L'œdème est absent ou discret*. Il permet, entre autres, de confirmer le comportement qualificatif de certains adjectifs verbaux (voir la section 1.2.1.2.3 pour la différence entre les adjectifs verbaux et dérivés de verbes). Par exemple, *des épanchements qui peuvent être bilatéraux et abondants*. La coordination avec un autre type d'adjectif, comme un adjectif dérivé de nom, (ex. *nom de tête + adjectif dérivé de nom + et + adjectif qualificatif*) est inusitée. Par exemple, l'expression *une infection abdominale et aiguë* est fautive, car l'*infection* qui est *aiguë* est l'*infection abdominale*. En d'autres mots, les deux adjectifs ne jouent pas le même rôle, *aigu* renvoie à la propriété de l'*infection* alors que l'adjectif *abdominal* exprime une relation entre *infection* et *abdomen* (base nominale de *abdominal*).

c) Acceptation d'une modification par un adverbe de degré et un complément de l'adjectif

Premièrement, l'adverbe de degré (*très, peu*) qui accompagne un adjectif qualificatif permet de quantifier l'état ou la qualité exprimé par l'adjectif qu'il modifie, une caractéristique qui n'est généralement pas partagée par les adjectifs dérivés de noms. Par exemple, une *infection* peut être *très fréquente*, mais elle ne peut pas être *très bactérienne* (adjectif dérivé de *bactérie*); elle sera causée par des bactéries ou non.

Deuxièmement, le complément de l'adjectif permet de préciser un aspect de l'état ou de la qualité de l'adjectif qu'il modifie. De plus, l'adjectif qualificatif qui a un complément est généralement attribut. Parmi les différents groupes compléments (en gras dans les exemples), nous comptons les syntagmes prépositionnels (introduits par des prépositions comme *de, à, contre, envers, pour, avec* et *en*) et les propositions « subordonnées analogues aux circonstancielles (*un homme digne qu'on l'appelle aux plus hautes fonctions*) » (Arrivé, Gadet et al. 1986 : 43). Le syntagme prépositionnel est toutefois le groupe complément le plus fréquent dans les langues de spécialité. Le syntagme prépositionnel accompagne souvent des adjectifs qui précisent la présence et la fréquence du concept nommé, comme *possible, impossible* et *absent*. Par exemple, dans la phrase *La fièvre est absente dans 50 % des cas*, la présence du complément *dans 50 % des cas* n'a plus sa raison d'être si nous retirons l'adjectif *absent*. Nous

convenons que la phrase *L'infection est bactérienne dans 50 % des cas* semble juste. Par contre, la phrase réelle est *L'infection est [une infection] bactérienne dans 50 % des cas* (voir le critère *a*) et, à partir de cette phrase, le complément *dans 50 % des cas* devient un complément du terme complexe *infection bactérienne*.

1.2.1.1.2 Adjectifs déterminatifs

Ce groupe d'adjectifs constitue une classe à part. En effet, ils sont jugés obligatoires au même titre que les articles et les autres déterminants (Arrivé, Gadet et al. 1986 : 33), car ils sont essentiels à la syntaxe de la phrase et ils tendent à se comporter comme des déterminants. Il s'agit d'adjectifs dont le sens est général et qui regroupent les éléments suivants (Noailly 1999 : 9) :

- 1) Adjectifs démonstratifs : ex. *celui, lequel*
- 2) Adjectifs possessifs : ex. *leur*
- 3) Adjectifs indéfinis : ex. *autre, même, quelconque*
- 4) Adjectifs interrogatifs : ex. *quel*
- 5) Adjectifs numéraux et ordinaux : ex. *troisième, deux, dernier*

Les adjectifs déterminatifs ne feront pas l'objet d'une analyse plus approfondie, car, par leur statut obligatoire dans la structure de la phrase, ils seront considérés comme des déterminants, tant au moment de procéder à l'analyse qu'au moment de traiter automatiquement le corpus en vue de procéder à l'extraction des adjectifs.

1.2.1.1.3 Adjectifs dérivés de verbes

L'adjectif dérivé de verbe constitue le type d'adjectif se rapprochant le plus des adjectifs dérivés de noms, car ils sont également caractérisés par la relation que leur base (en l'occurrence, un verbe) établit avec le nom qu'ils modifient. De plus, les adjectifs dérivés de verbes se reconnaissent d'après : 1) leur ressemblance

morphologique avec un verbe et 2) le sens de la base qui est présent dans le sens de l'adjectif.

Il est facile de croire que les nombreux adjectifs ayant une forme semblable à celle d'un verbe sont des adjectifs dérivés de verbes, mais il semble, d'après les observations que nous avons faites dans notre corpus, que les véritables adjectifs dérivés de verbes sont assez rares. Il convient de distinguer les adjectifs verbaux des adjectifs dérivés de verbes.

Les adjectifs verbaux sont des verbes conjugués (participes passés et présents) qui se comportent comme des adjectifs qualificatifs. Les participes présents adjectivés se distinguent facilement du verbe conjugué, car l'adjectif s'accorde en genre et en nombre (ex. *bactérie aérotolérante*) alors que le verbe est invariable (ex. *les débris circulant dans le liquide*). Pour ce qui est des participes passés, utilisés avec l'auxiliaire *avoir*, ils sont généralement des verbes conjugués (ex. *le médecin a prodigué des soins*). Le statut des participes passés utilisés avec l'auxiliaire *être* est parfois ambigu. Lorsque le participe exprime un état, il s'agit le plus souvent d'une interprétation adjectivale (ex. *je suis fatigué* qui est utilisé au sens de « je suis las »). Par contre, si le participe exprime la conséquence engendrée par un phénomène externe, s'agit plutôt du verbe conjugué (ex. *je suis fatigué par le voyage* au sens de « résultant du voyage ») (Noailly 1999 : 19-20).

Une fois les participes adjectivaux écartés, le nombre d'adjectifs dérivés de verbes diminue. Parmi ceux-ci, nous trouvons principalement des adjectifs comme *adhérent* qui provient du verbe *adhérer*. Dans ce cas, la ressemblance morphologique est tout aussi évidente que la parenté sémantique. En effet, l'adjectif peut être remplacé par une paraphrase qui explique la relation unissant le nom de tête d'un syntagme et le verbe. Les patrons syntaxiques qui permettent de décrire la relation entre le nom de tête et le verbe de base de l'adjectif sont :

1) Nom de tête + de + verbe de base à l'infinitif

2) Nom de tête + qui + verbe de base conjugué

Par exemple, le syntagme nominal *surface adhérente* donnera la phrase explicative *surface qui adhère* et le syntagme *pouvoir adhérent* donnera *pouvoir d'adhérer*.

À la suite d'observations effectuées sur notre corpus, nous avons été en mesure d'identifier un cas particulier d'adjectif dérivé de verbe. En effet, nous avons identifié une occurrence de l'adjectif *alimentaire*, où l'adjectif est lié au verbe pronominal *s'alimenter* au lieu d'être lié aux noms *aliment* ou *alimentation*, comme dans les autres occurrences. Dans le syntagme *refus alimentaire* l'adjectif peut être décrit par *refus d'ingérer des aliments*, ce qui signifie la présence d'une double paraphrase, soit : *refus d'ingérer + ingérer des aliments*. Pour respecter les patrons syntaxiques, nous devons trouver une expression qui nous permettra d'utiliser une seule paraphrase, donc nous avons décrit le syntagme par *refus de s'alimenter*.

1.2.1.2 Adjectifs dérivés de noms

L'adjectif dérivé de nom est un adjectif dont la morphologie et le sens sont apparentés à ceux d'un nom de base. En d'autres mots, il s'agit d'un adjectif qui ressemble généralement au nom ou utilise une composition savante (ex. *hépatique*, préfixe *hépat-*, grec pour *foie*). À première vue, les adjectifs dérivés de noms se reconnaissent par leur ressemblance morphologique avec un nom (Daille 2001). Dans la plupart des cas, il est facile de trouver le nom de base dont est dérivé l'adjectif. Par exemple, si l'adjectif *viral* est observé, nous dirons intuitivement que la base nominale est *virus*.

Dans cette section nous verrons d'abord les principales caractéristiques qui sont propres aux adjectifs dérivés de noms. Ensuite, nous nous attarderons sur le concept de « relation sémantique ». Enfin, nous examinons un groupe d'adjectifs

dérivés de noms un peu particulier : ils sont souvent le produit d'une conversion (voir la section 1.1.3) et ils ont un lien sémantique ténu avec un nom.

1.2.1.2.1 *Caractéristiques des adjectifs dérivés de noms*

Dans le cadre de ses travaux, Monceaux (1997 : 39-40) s'est penchée sur la description de l'adjectif dérivé de nom dans une perspective de traitement automatique des langues. Dans ces travaux, Monceaux le nomme *adjectif de relation*, ou *pseudo-adjectif dénominal*, et elle le présente comme étant un adjectif « qui détermine, plutôt que qualifie, le nom recteur. C'est-à-dire qu'il vaut pour un complément du nom, même s'il n'en est pas un véritable équivalent fonctionnel, plutôt que pour une qualification » (Monceaux 1997 : 40). Elle énumère, entre autres, une liste des principales caractéristiques de ces adjectifs et des critères permettant de les distinguer, notamment des adjectifs qualificatifs (ces critères sont repris par Daille (2001), voir la section 2.3.3).

- a) **La parenté morphologique avec un nom** : l'adjectif est dérivé d'un nom au moyen d'un suffixe (ex. *-aire*, *-al*, *-el*; voir la section 2.3.1 pour une liste plus complète). Comme pour l'adjectif dérivé de verbe, il arrive que la base nominale provienne de la racine latine ou grecque. Par exemple, l'adjectif *hépatique* est construit à partir de la racine grecque *hépatos* qui signifie « foie ». L'adjectif exprime alors une relation sémantique entre le nom exprimé par la racine grecque et le nom qu'il modifie. Par exemple, dans le syntagme *pathologie hépatique*, l'adjectif peut-être remplacé par la paraphrase *qui affecte le foie*.
- b) **La capacité d'être remplacé par un complément prépositionnel se rattachant au nom qu'il modifie** : une relation simple entre le nom de base et le nom que l'adjectif modifie peut être exprimée par une préposition (ex. *acidité sanguine – acidité du sang*) (Monceaux 1997 : 39). Nous verrons à la section suivante que la relation peut parfois être plus complexe et exiger une expression sous forme de proposition subordonnée plutôt qu'une simple préposition.

- c) **Le fait de ne jamais être épithète antéposé** : il est parfois possible de placer une épithète devant le nom (ex. *une grande beauté*) lorsque celle-ci est un adjectif qualificatif. Il n'est pas possible d'en faire autant avec l'adjectif dérivé de nom. En effet, il est inusité de lire *l'abdominale douleur*. L'antéposition peut également servir de critère de distinction entre deux formes homonymes. Par exemple, il est inusité d'utiliser *les infernales puissances* (puissances de l'enfer) alors qu'il est juste d'écrire *cette infernale voiture* (voiture faisant beaucoup de bruit) (Monceaux 1997 : 40).
- d) **Le fait d'être inusité comme attribut** : lorsqu'un adjectif dérivé de nom est observé dans une construction attributive, *nom + verbe d'état + adjectif*, il sous-entend la répétition du nom de tête et, par le fait même, la présence d'un complément d'objet. De plus, une construction attributive contraint généralement à une interprétation sous-classificatrice (partie-tout), contrastive (opposition) ou typique (normalité). Par exemple, *le traitement est symptomatique* signifie réellement « le traitement est un traitement symptomatique ». Dans l'exemple, nous observons la présence du complément et le caractère typique de celui-ci, car l'adjectif accepte la modification par l'adverbe *normalement*.
- e) **L'incompatibilité avec l'expression d'un degré** : contrairement à l'adjectif qualificatif, l'expression de la gradation est rare avec les adjectifs dérivés de noms. En effet, il est inusité d'écrire *une infection très pulmonaire*, car une infection affecte les poumons ou elle ne les affecte pas. Par contre, ces adjectifs acceptent les adverbes d'exclusion (ex. *strictement, uniquement*, etc.) et de restriction (ex. *essentiellement, surtout*, etc.). Par exemple, il est tout à fait correct d'écrire *une infection strictement pulmonaire*.
- f) **Le refus de la nominalisation avec le verbe avoir** : l'adjectif ne peut pas être remplacé par un segment de phrase qui contient le verbe *avoir* et l'adjectif transformé en nom (ex. *sanguin-sanguinité*). Par exemple, il n'est pas possible de remplacer *ce groupe sanguin* par *ce groupe a une certaine sanguinité* tout en gardant le même sens (Monceaux 1997 : 40).
- g) **Le refus de la nominalisation au moyen de l'article générique** : dans la plupart des cas cette règle s'applique, car il est impossible de nominaliser l'adjectif

cardiaque, par le simple ajout d'un article, sans sous-entendre la présence d'un nom antéposé. Par exemple, *le cardiaque* sous-entend *le patient cardiaque*. Ce critère s'applique moins bien aux adjectifs par conversion. Par exemple, *antibiotique* est à la fois un nom (ex. *administrer l'antibiotique* [substance]) et un adjectif (ex. *crème antibiotique* [propriété de la crème]).

Ces critères sont généraux et il est préférable de considérer l'application de plus d'un critère lors du classement. Comme nous l'avons relevé au dernier point, il arrive de rencontrer certaines exceptions. Outre le cas des adjectifs par conversion, il arrive également que la valeur sémantique de l'adjectif repose sur un concept connexe au nom de base, ce qui explique la possibilité d'observer plus d'une base nominale. Par exemple, reprenons l'adjectif *alimentaire*. Il renvoie intuitivement au nom *aliment*. Par exemple, dans le syntagme *denrée alimentaire*, l'adjectif peut être remplacé par la paraphrase *qui est composée d'aliments*. Toutefois, dans le syntagme *enquête alimentaire*, il renvoie au concept connexe d'*alimentation*, car l'adjectif peut être remplacé par la paraphrase *qui porte sur l'alimentation*.

1.2.1.2.2 Adjectifs « relationnels »

La complexité de l'adjectif dérivé de nom ne tient pas seulement dans la reconnaissance de sa base nominale. Elle repose également sur la quantité, parfois impressionnante, de relations sémantiques qui unissent le nom de tête du syntagme nominal et le nom de base de l'adjectif (voir la section 3.5.2.1). Dans un grand nombre de publications (Mélis-Puchulu 1991; Monceaux 1997; Noailly 1999; Daille 2001; Maniez 2005), ce type d'adjectif est appelé *adjectif relationnel*. Il possède, entre autres, la caractéristique de pouvoir être remplacé par une paraphrase prépositionnelle dont la forme générale est *préposition + déterminant + nom de base* (Monceaux 1997 : 42). Tamba-Mecz (1980 : 118) dit de l'adjectif relationnel :

« qu'il marque une relation entre deux notions distinctes, au lieu d'indiquer une qualité comme l'adjectif qualificatif [...]. Les relations que signifient ces adjectifs sont synonymes de celles que l'on peut identifier entre deux noms reliés par une préposition [...]. »

Cela dit, nous croyons que le statut relationnel de l'adjectif ne se limite pas à la dérivation nominale, car la nature de la relation repose sur la base sémantique de l'adjectif au sens large. Comme le soutient Melis-Puchulu (1991 : 34), il est possible d'apporter des nuances à la définition de l'adjectif relationnel. L'analyse en contexte des syntagmes nominaux a permis d'observer que la relation sémantique qui unit le nom de tête du syntagme et le nom de base des adjectifs dérivés de noms doit parfois être plus élaborée qu'une simple préposition.

En admettant qu'un syntagme prépositionnel contenant le nom de base ne parvienne pas à toujours exprimer toutes les acceptions des adjectifs dérivés de noms, nous reconnaissons qu'ils possèdent souvent plusieurs acceptions. Au moment de différencier les acceptions de l'adjectif, il est préférable, de viser le remplacement de l'adjectif par un complément du nom qui explique la nature de la relation sémantique au lieu de se limiter à la paraphrase prépositionnelle (Riegel, Pellat et al. 2004 : 357).

Nous retenons donc deux cas de figure. Le sens de l'adjectif dérivé de nom peut s'expliquer au moyen d'une préposition suivie du nom de base. Par exemple, le syntagme *culture bactérienne* peut être paraphrasé par *culture de bactéries*, car la relation sémantique est simple, c'est-à-dire qu'elle vise à unir les deux notions, *culture* et *bactéries*, sans qu'il y ait d'ajout de sens. Toutefois, les adjectifs dérivés de noms, peuvent exprimer un apport sémantique d'ordre définitionnel (ex. *qui affecte x, qui a le caractère de y*) ou perceptuel (ex. *qui ressemble à x*) (Mélis-Puchulu 1991 : 44). Par exemple, reprenons l'adjectif *bactérien*. Nous avons vu qu'il peut exprimer le simple renvoi au nom de base, mais combiné à un autre nom, il ajoute un élément de sens : le syntagme *infection bactérienne* ne se définit pas par *infection de bactéries*, mais par *infection qui est causée par une bactérie*.

1.2.1.2.3 Cas particuliers

Un certain nombre d'adjectifs apparentés au nom possèdent un lien sémantique ténu avec celui-ci. Parmi eux, nous comptons des adjectifs par conversion qui possèdent une forme identique à celle d'un nom. Toutefois, dans ces cas particuliers, le sens du nom ne peut être retrouvé tel quel dans le sens de l'adjectif. Même si le sens est parfois connexe, il est difficile de construire la paraphrase en utilisant la base nominale comme nous le faisons lorsque l'adjectif possède une forme semblable à celle du nom de base. De manière générale, l'adjectif partage un lien étymologique avec le nom, mais le lien direct avec celui-ci a été perdu. La paraphrase explique alors la relation unissant le nom de tête du syntagme à l'élément de sens de l'adjectif tout en expliquant cet élément. Par exemple, observons l'adjectif *adulte*. Il possède la même forme que le nom. Par contre, la notion d'adulte en médecine varie de la notion d'âge adulte telle que nous la connaissons généralement. Dans la société, l'âge adulte est généralement associé à un chiffre. Par exemple, au Canada, l'âge adulte (majorité) est fixé à 18 ans et aux États-unis, il est fixé à 21 ans. En médecine, la notion d'âge adulte est plutôt associé au développement d'un organisme, soit à la maturité physique. Par exemple, sur les monographies de médicaments, comme nous pouvons le lire dans le CPS (Compendium de produits et spécialités pharmaceutiques), sous la rubrique du *Tylenol* contre la douleur (Leesa 2000 : 1858), la posologie pour adulte vise les humains de 12 ans et plus. En d'autres mots, l'âge auquel le corps atteint sa maturité sexuelle, c'est-à-dire qu'il devient apte à se reproduire. Cela dit, tenter de définir l'adjectif par une paraphrase qui contient le nom *adulte* comme nous le faisons pour les adjectifs dérivés de noms donne une définition qui reste ambiguë, elle ne fait pas la distinction entre la majorité d'un être humain et la maturité d'un organisme vivant. Dans les occurrences observées en infectiologie, il s'agit principalement de la maturité. Nous décrivons donc, dans le syntagme *animal adulte*, le sens de l'adjectif par la paraphrase *qui a atteint la maturité*.

Dans un deuxième temps, observons l'adjectif *clinique*. Il possède la même forme que le nom désignant l'« établissement privé destiné aux soins des malades, hospitalisés ou non » (Garnier et Delamare 2002 : s.v.). Par contre, lorsqu'il est utilisé

comme adjectif, *clinique* ne renvoie pas seulement à l'établissement, mais à un élément il comprend des éléments de sens connexe de celui-ci, soit le malade. C'est pourquoi un *examen clinique* ne signifie pas un « examen effectué en clinique », mais plutôt « un examen effectué sur un malade ».

1.2.2 Critères de classification des adjectifs

Dans ce chapitre, nous avons vu que l'adjectif est un élément syntaxiquement facultatif qui décrit ou détermine le nom qu'il modifie. De plus, nous avons présenté différentes façons de classer les adjectifs. Dans le cadre de notre étude, nous avons adopté une approche lexico-sémantique, fondée sur les travaux de Arrivée (1986), Méli-Puchulu (1991) et Monceaux (1997). Nous avons proposé quatre classes d'adjectifs : les adjectifs qualificatifs, les adjectifs déterminatifs, les adjectifs dérivés de verbes et les adjectifs dérivés de noms. En outre, les adjectifs sont classés en fonction d'une acception donnée. Par exemple, l'adjectif *alimentaire* est un adjectif dérivé de nom lorsqu'il est utilisé avec *denrée* (« qui est composée d'aliments ») ou avec *étude* (« qui porte sur l'alimentation »). Par contre, il est un adjectif dérivé de verbe lorsqu'il est combiné au nom *refus* (« de s'alimenter ») (voir la section 1.2.1.1.3).

Le tableau 1.1 résume les critères que nous avons choisis d'appliquer pour classer les adjectifs dans les différentes classes.

Ces critères sont le point de départ du classement des adjectifs de notre corpus qui nous permet d'isoler ceux qui font l'objet de notre étude à savoir les adjectifs dérivés de noms. Toutefois, nous souhaitons profiter du chapitre suivant pour effectuer un survol des principaux travaux sur l'analyse et le traitement automatique des adjectifs en corpus spécialisés.

Chapitre 2 : État de l'art

Dans ce chapitre, nous effectuons d'abord un survol de la place occupée par les différentes parties du discours dans les ressources terminologiques (ex. bases de données et dictionnaires spécialisés). Par la suite, nous nous concentrons sur les principales études terminologiques qui ont été faites sur les adjectifs : notamment l'étude de L'Homme (2004a) portant sur les adjectifs en corpus spécialisés, les études de Normand et Bourigault (2001), Maniez (2001) et Bae (2006) qui proposent des méthodes de description des adjectifs spécialisés et, enfin, les travaux de Grabar et Zweigenbaum (2003), Zweigenbaum, Hadouch et Grabar (2003), Daille (2001) et Maniez (2002) qui ont mis au point des méthodes de traitement automatique de la langue appliquées aux adjectifs.

2.1 Place des différentes parties du discours en terminologie

Avant de nous attaquer à l'étude des adjectifs spécialisés, nous évoquons les raisons pour lesquelles l'intérêt pour les adjectifs est si récent en terminologie. Par la suite, nous effectuons un rapide survol des travaux terminologiques qui ont été faits sur les verbes et les adjectifs dérivés de verbes.

2.1.1 Prédominance du nom dans les terminologies

Les auteurs de travaux terminologiques s'entendent pour dire que la présence des noms est nettement plus importante dans les dictionnaires spécialisés que celle des autres parties du discours (Sager 1990; Daille 2001; L'Homme 2003). Parmi ces derniers, Sager (1990 : 58) a observé que les verbes et les adjectifs utilisés dans les langues techniques sont généralement présents sous leur forme nominale dans les dictionnaires terminologiques. Pour expliquer ce phénomène, Cabré (1998 : 73) mentionne notamment que la terminologie part des concepts pour découvrir les dénominations. Contrairement aux ouvrages lexicographiques « qui contiennent des mots de toutes les catégories grammaticales » (Cabré 1998 : 76), les ouvrages

terminologiques comptent principalement des noms, car ils ont comme objectif de nommer des notions spécialisées et non des faits de la vie quotidienne qui renvoient souvent à des qualités. Toutefois, lorsque nous abordons la question d'un point de vue traductionnel, il convient d'admettre que le fait de connaître les noms donnés aux notions ne suffit pas toujours pour traduire un texte spécialisé correctement.

Le travail terminologique fournit des éléments descriptifs qui permettent une compréhension de la notion. Dans le cas des langues de spécialité, il est fondamental de pouvoir utiliser les noms de concepts. Pour y arriver, nous utilisons d'autres parties du discours, comme les verbes et les adjectifs. Par contre, ces derniers sont souvent observés dans les termes complexes et ils possèdent rarement leur propre entrée dans les bases de données et les thésaurus (Daille 2001 : 161) et, souvent, lorsqu'ils figurent dans ces répertoires, ils ne sont pas définis.

Au cours des dix dernières années, l'émergence d'une terminologie qui prend davantage en compte le fonctionnement des termes en contexte a permis de voir que les termes sont étroitement liés au texte dans lequel ils apparaissent. Cette observation a soulevé nombre de questions à propos de la définition de *terme*. Sans entrer dans ce sujet particulièrement complexe, nous devons toutefois mettre en évidence un des aspects qui ont été soulevés : le concept de « terme » ne se limite plus aux noms (L'Homme, Heid et al. 2003).

2.1.2 Prise en compte du verbe dans les descriptions terminologiques

La première partie du discours, autre que le nom, qui a suscité de l'intérêt en terminologie a été celle du verbe. L'analyse des verbes dans les textes techniques a permis de mettre en évidence la ou les acceptions qui sont spécialisées. Pour y parvenir, L'Homme (2004b : 64-66) présente une série de critères :

- 1) « **L'unité lexicale a un sens qui est lié à un domaine de spécialité** » (L'Homme 2004b : 64) : Lorsqu'il s'agit d'un nom, il est généralement facile d'établir le lien entre l'unité lexicale et le domaine qui est préalablement délimité selon le projet

(ex. *programme* dans le domaine de l'informatique). Lorsqu'il s'agit d'autres parties du discours, la délimitation n'est pas toujours aussi claire et il est préférable de combiner ce critère aux trois critères lexico-sémantiques qui suivent.

- 2) « **La nature des actants sémantiques peut servir d'indice pour confirmer le sens spécialisé d'une unité lexicale à sens prédicatif** » (L'Homme 2004b : 64-65) : Lorsque les actants nominaux sont considérés comme des termes (voir le critère 1, il y a de fortes chances que l'unité prédicative le soit également. Par exemple, le verbe *soigner* est lié au domaine médical lorsqu'il possède un terme médical comme actant, à savoir *patient*, *malade*. Par contre, le sens du verbe varie lorsque les actants changent de domaine. En effet, *soigner* combiné à un actant lié au domaine général, comme une action (ex. *travail*, *tâche*) ne renvoie pas au même sens que *soigner* lorsqu'il est utilisé avec un actant qui est lié au domaine médical comme *patient* ou *malade*. En effet, *soigner une action* signifie « effectuer quelque chose avec soin » et *soigner un patient*, « assurer le bien-être de quelqu'un » (Rey et Rey-Debove 2004 : s.v.).
- 3) « **La parenté morphologique** » (L'Homme 2004b : 65) : Lorsque l'unité lexicale possède une ressemblance morphologique, avec une unité définie comme terme en fonction des critères 1 et 2, il s'agit d'un indice confirmant le sens spécialisé de celle-ci. Si nous reprenons l'exemple de *soigner*, nous pouvons soutenir le fait qu'il est un terme en évoquant la relation morphologique et la parenté sémantique qui existe avec le nom *soin*, tel qu'il est vu dans l'expression *soins de santé*, liée au domaine médical.
- 4) « **Toute autre relation paradigmatique** » (L'Homme 2004b : 66) : Lorsqu'une unité lexicale est reconnue comme terme selon les trois premiers critères, il est fort possible que les unités qui possèdent une relation paradigmatique (autre que morphologique) avec celle-ci soient également spécialisées. Prenons l'exemple du verbe *réveiller*. Il sera juste de parler d'un verbe médical puisqu'il entre dans le même paradigme que les verbes *opérer* et *anesthésier* qui sont facilement identifiables comme termes. En effet, il s'agit de trois verbes utilisés lors de procédures chirurgicales.

Ces critères d'identification des unités lexicales spécialisées s'appliquent aux différentes parties du discours. En effet, les critères ont été utilisés avec les adjectifs spécialisés dans les travaux de Bae (2006).

L'étude des verbes spécialisés a également mis en évidence que le verbe constitue un point de départ pour repérer des termes dérivés (nominaux et adjectivaux). L'Homme (2003 : 410) donne l'exemple du verbe *compiler* qui est à l'origine des dérivations *compilation*, *compilable*, *précompiler*, *recompiler* et *recompilable*. Elle précise également que la relation sémantique qui unit le verbe et l'unité lexicale dérivée s'exprime au moyen d'une paraphrase dans laquelle se trouve le verbe (ex. *compilable* s'exprime par *qui peut être compilé*).

L'étude des dérivés de verbes a mis en évidence le rôle des adjectifs dérivés dans le discours spécialisé. En effet, même s'ils apparaissent rarement seuls dans les bases de données terminologiques et les dictionnaires, ils occupent une place importante lorsqu'ils sont compris dans les termes complexes. De même, leur présence dans certains corpus est telle que les adjectifs ont suscité l'intérêt des terminologues et autres spécialistes travaillant sur les corpus spécialisés (Daille (2001), Grabar et Zweigenbaum (2003), L'Homme (2004), Maniez (2001), Zweigenbaum, Hadouch et Grabar (2003)).

2.1.3 Place des adjectifs en corpus spécialisés

Dans le cadre d'une étude récente, L'Homme (2004a) s'est intéressée aux « adjectifs dérivés sémantiques » et leur rôle dans la structuration des terminologies. Cette étude permet d'observer l'importance de la contribution des adjectifs à la structuration des terminologies. Elle vise également à vérifier si cette contribution se définit en fonction des liens sémantiques qu'ils partagent avec leur base sémantique.

Pour atteindre ces objectifs, L'Homme a utilisé trois corpus spécialisés appartenant au groupe de terminologie de l'Observatoire de linguistique Sens-Texte (OLST) qui compte environ 500 000 mots chacun : un corpus informatique, un corpus juridique et un corpus médical. Les corpus ont d'abord été traités au moyen de l'étiqueteur WinBrill pour permettre l'extraction automatique de concordances contenant un adjectif. Ces dernières ont alors été importées dans une base de données. Par la suite, les adjectifs ont été étudiés pour être classés dans les catégories suivantes : adjectifs dérivés de noms, adjectifs dérivés de verbes, adjectifs qualificatifs, adjectifs numériques, adjectifs indéfinis³, adjectifs régissant des syntagmes prépositionnels et adjectifs qui possèdent un lien sémantique ténu avec un nom.

Ce classement a permis d'observer la répartition des 1000 adjectifs qui ont été sélectionnés de façon aléatoire. Le tableau 2.1 présente la proportion d'adjectifs dérivés de manière générale et les adjectifs dérivés de noms et de verbes qui sont majoritaires selon le corpus. Ces données indiquent que les adjectifs dérivés sont très présents dans les trois corpus spécialisés (ils représentent de 20 à 40 % des adjectifs), que le corpus médical compte le plus d'adjectifs dérivés de noms (respectivement 85 % des adjectifs dérivés) et que le corpus informatique compte le plus d'adjectifs dérivés de verbes (39,4 % des adjectifs dérivés).

Tableau 2.1 Proportion des adjectifs dérivés dans trois corpus spécialisés

Type d'adjectif	Corpus	Proportion (%)
<u>Adjectifs dérivés</u>	Informatique	20,8 %
	Juridique	37,7 %
	Médical	40,8 %
<u>Adjectifs dérivés de noms</u>	Médical	85,3 %
	Juridique	78,2 %
	Informatique	60,6 %
<u>Adjectifs dérivés de verbes</u>	Informatique	39,4 %
	Juridique	21,8 %
	Médical	14,7 %

³ Les adjectifs numériques et indéfinis ont été classés dans les adjectifs déterminatifs à la section 1.2.1.1.2.

L'Homme (2004a) s'est ensuite attardée à l'étude des adjectifs dérivés. L'analyse des adjectifs dérivés sémantiques identifiés au cours de cette étude a attiré l'attention sur la relation qui existe entre l'adjectif et sa base nominale ou verbale. Celle-ci a également montré que la relation en question permet d'exprimer un ajout de sens à la base de l'adjectif. Par exemple, l'adjectif *infectieux* peut signifier « qui provoque une infection ». De plus, il a été observé que la nature de la relation sémantique change parfois selon la classe du nom que l'adjectif modifie. Par exemple, l'adjectif *symptomatique* décrit l'état d'un être animé (*un patient qui présente des symptômes*) alors qu'il renvoie à l'action d'une substance (*un médicament qui s'attaque aux symptômes*).

Les données obtenues dans le cadre de cette étude font état de l'importance quantitative et du caractère polysémique des adjectifs dérivés de noms dans les corpus médicaux. Cette conclusion soutient le besoin d'approfondir l'étude de ce groupe d'adjectifs.

2.2 Études portant sur les adjectifs spécialisés en médecine

L'importance des adjectifs dans les textes médicaux a soulevé l'intérêt d'un certain nombre de chercheurs (Normand et Bourigault (2001), Maniez (2001) et Bae (2006)). Dans cette section, nous présentons des études qui portent sur les outils servant à l'enrichissement des ressources terminologiques dans le domaine médical en français et sur la traduction du nom adjectival en anglais. Enfin, nous décrivons une analyse très semblable à la nôtre qui prend en compte le sens des termes adjectivaux médicaux en coréen.

2.2.1 Analyse des adjectifs en corpus au moyen d'outils de traitement des langues naturelles

Dans le cadre d'une étude visant à enrichir et uniformiser la terminologie utilisée en histopathologie, Normand et Bourigault (2001) ont voulu montrer qu'il est possible d'acquérir une terminologie qui facilite la rédaction des observations histopathologiques⁴ à partir d'un corpus. De plus, l'étude vise à explorer de nouvelles méthodes pour mettre en forme des glossaires et élaborer une terminologie qui permet la description de concepts et des observations faites par les experts du domaine. Cette étude vise enfin à vérifier si l'utilisation des bons outils informatiques permet d'acquérir une terminologie utilisée lors de l'évaluation des pathologies.

Pour y parvenir, Normand et Bourigault ont utilisé un corpus de l'équipe d'informatique médicale de l'Hôpital Broussais qui compte 250 000 mots. L'analyse des adjectifs a exigé l'utilisation de deux logiciels. Le premier est un extracteur de termes nommé LEXTER qui traite un corpus préalablement étiqueté. Le logiciel fractionne d'abord les chaînes de mots en syntagmes (ex. *fragment de tissu mammaire*). Ces syntagmes sont également nommés *MLNP (Maximal-Length Noun Phrases)*. Ensuite, un second module décompose les *MLNP* en paire de dépendances sémantiques, soit une tête et une expansion. L'exemple qui est présenté à la figure 2.1 montre que l'expansion d'un syntagme peut être constituée d'une simple unité lexicale ou d'un syntagme :

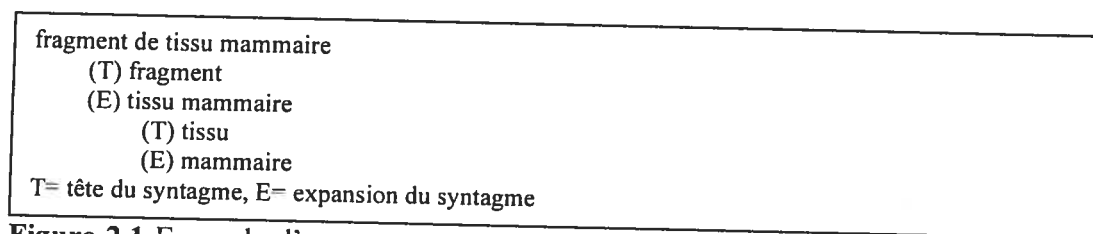


Figure 2.1 Exemple d'un syntagme nominal qui contient un autre syntagme nominal en expansion (Normand et Bourigault 2001 : 157)

⁴ Histopathologie : étude microscopique des tissus et des organes malades. (Garnier et Delamare 2002 : 389).

Le second logiciel utilisé dans le traitement des adjectifs est LEXICLASS. Il s'agit d'un logiciel qui regroupe les termes d'après leur contexte terminologique, soit le lien que possèdent les termes dans le réseau sémantique (ici, le lien entre les noms de tête). Par exemple, les noms de tête *tissu*, *fragment* et *cellule* seront regroupés, car il s'agit de *structures tissulaires*. Pour interpréter les listes de combinaisons générées par le logiciel, le terminologue examine les noms de tête pour avoir ensuite accès, entre autres, aux syntagmes nom-adjectif et aux termes complexes formés de paires nom-adjectif.

Cette façon de décomposer les segments extraits du corpus a permis d'identifier 437 adjectifs distincts dont 284 apparaissent plus d'une fois. Il a été possible de constater que des liens sémantiques sont établis entre les groupes de noms et d'adjectifs qui entretiennent une dépendance syntaxique. De plus, l'acquisition d'un glossaire à partir de l'extraction a été réussie, car la plupart des groupes de termes extraits ont été jugés pertinents par les spécialistes qui ont révisé les listes. Enfin, en utilisant des méthodes d'extraction et de regroupement automatiques, il devient possible d'éviter le piège de la subjectivité qui est souvent associé au traitement manuel des adjectifs.

2.2.2 Traduction du nom adjectival en anglais médical

Pour étudier la traduction en français du nom adjectival anglais (nom qui remplit la fonction grammaticale de l'adjectif (communication personnelle)), Maniez (2001) a effectué une étude comparative visant à analyser des adjectifs anglais et français dans un corpus anglais, un corpus français et un corpus bilingue de façon à voir de quelle façon ceux-ci peuvent être traduits. Son étude vise d'abord à montrer de quelle façon l'utilisation de corpus spécialisés peut contribuer à mettre en lumière certaines caractéristiques des langues de spécialité. De plus, l'étude permet de formuler quelques principes qui peuvent guider un traducteur confronté à un groupe nominal anglais contenant un nom adjectival.

Pour procéder à l'analyse des adjectifs et des noms adjectivaux, Maniez a mis en forme trois corpus. Le corpus anglais est constitué de textes publiés dans *Internal Medicine* paru en 1993 et il contient près de 13,5 millions de mots. Le corpus français est constitué de textes repêchés dans Internet et il totalise 750 000 mots. Enfin, le corpus bilingue est composé d'articles parus dans le *Journal of the American Medical Association* et de leur traduction, parue dans la même revue. La méthode utilisée pour l'obtention des résultats semble être en partie manuelle et centrée sur certains adjectifs spécifiques⁵. Il reste que la taille des corpus laisse sous-entendre qu'un traitement automatique a permis de recenser les adjectifs.

D'après Maniez, le corpus français compte plus de 400 adjectifs dérivés (il s'agit plus précisément d'adjectifs construits par suffixation). Parmi ceux-ci, très peu possèdent un équivalent adjectival dans le corpus anglais. L'adjectif est soit absent, soit rarement utilisé. Par exemple, l'adjectif *plasmatic* existe dans la langue anglaise, mais il est rarement utilisé. Dans le corpus étudié, l'adjectif *plasmatic* est identifié trois fois alors que le nom adjectival *plasma* apparaît près de 4000 fois.

Par ailleurs, pour illustrer les contraintes lexicales qu'il a observées, Maniez utilise l'exemple des adjectifs français dont le suffixe est *-ionnel*. Il souligne :

- 1) que nombre de ces adjectifs ne possèdent pas d'équivalent anglais dans les dictionnaires;
- 2) qu'ils sont souvent plus fréquents en français;
- 3) que les adjectifs ayant un préfixe prépositionnel (ex. *prétransfusionnel*) ne possèdent pas d'équivalent en anglais;
- 4) que leur utilisation est parfois trop rare pour justifier une entrée dans un dictionnaire; et

⁵ La méthode utilisée n'est pas présentée dans l'article consulté. Nous avons donc déduit cet élément méthodologique à partir des autres éléments présents dans l'article et de l'absence de renvois à des logiciels.

- 5) qu'un emploi des adjectifs relationnels qui est plus rare en français qu'en anglais peut signifier une tendance à l'anglicisme (ex. l'utilisation grandissante de *occupationnel* au lieu de *professionnel* pour traduire *occupational*).

L'auteur précise que la proximité entre deux noms ou un nom et un adjectif ne garantit pas la présence d'un syntagme nominal. En utilisant l'exemple du nom *cell*, Maniez montre que dans 43 % des occurrences, celui-ci possède une fonction adjectivale, car, en anglais, la fonction grammaticale change selon l'organisation du patron syntagmatique. Le tableau 2.2 présente un échantillon d'exemples de patrons syntagmatiques où le nom *cell* est traduit par un syntagme prépositionnel ayant une fonction de complément du nom.

« L'adjectif *cellulaire* est employé dans presque tous les cas où *cell* est l'unique prémodificateur du groupe nominal. Dans tous les autres cas de figure, c'est-à-dire lorsque *cell* est lui-même précédé d'un nom ou d'un adjectif, la traduction par le complément de nom est employée » (Maniez 2001 : 65).

Tableau 2.2 Échantillon des patrons syntagmatiques du nom adjectival *cell* (Maniez 2001 : 64)

Structure	Syntagme nominal anglais	Syntagme nominal français
ADJ ADJ <i>cell</i> N	gastric epithelial cell surface	surface des cellules épithéliales gastriques
ADJ ADJ N <i>cell</i> N	vascular smooth muscle cell prolifération	prolifération des cellules musculaires lisses des vaisseaux
ADJ <i>cell</i> N	endothelial cell lysis	lyse des cellules endothéliales
ADJ N <i>cell</i> N	natural killer cell activity	activité des cellules tueuses naturelles
<i>cell</i> N	cell viability	viabilité de la cellule
N <i>cell</i> N	foam cell formation	formation de cellules spumeuses

Afin d'identifier quelques tendances dans la traduction de termes adjectivaux (nom adjectival ou adjectif) anglais vers le français, Maniez utilise d'abord l'exemple

de l'adjectif *hepatic* dans le corpus bilingue. Il a observé que 59 des 77 occurrences de l'adjectif *hepatic* n'apparaissent pas dans la phrase originale anglaise où il est remplacé par le nom adjectival *liver*, même si en français, il avait été traduit par l'adjectif *hépatique*. De plus, la présence de l'adjectif en anglais mène généralement à l'utilisation de l'adjectif en français, soit dans 92 % des cas. Enfin, il a remarqué qu'un syntagme peut donner différentes traductions en français selon le contexte. Par exemple, *liver disease* donnera les traductions : *hépatopathie*, *atteinte hépatique*, *pathologie hépatique*, *maladie hépatique*.

2.2.3 Analyse du sens des termes adjectivaux en coréen⁶

L'étude de Bae (2006) est probablement celle qui se rapproche le plus de notre travail. En effet, l'auteure propose une méthode d'analyse des adjectifs qui prend en compte les différents sens des termes adjectivaux. Elle a recours, pour le vérifier, à une modélisation du sens des adjectifs qui utilise les classes sémantiques des actants. Dans ce travail, Bae s'est concentrée sur l'analyse du sens des adjectifs qualificatifs en corpus coréen.

Pour y parvenir, Bae a mis en forme un corpus médical comptant 1 500 000 *eojeols* (groupe de caractères séparés par un espace ou une ponctuation). Pour sélectionner les termes spécifiques au domaine médical, Bae a utilisé l'indice de spécificité fourni par *TermoStat* (voir la section 3.2.1) et elle applique les critères lexico-sémantiques présentés par L'Homme (2004b : 64-66, voir section 2.1.2).

Après avoir appliqué ces critères, Bae a retenu 48 adjectifs ou termes adjectivaux. À partir de cette liste, elle a cherché à savoir si le sens du terme adjectival était le même d'une occurrence à une autre. Pour vérifier le sens des termes elle a

⁶ Les exemples cités dans cette section sont des traductions en français de gloses du coréen qui sont proposées par l'auteure.

appliqué une autre série de critères aux contextes contenant ces termes (Bae 2006 : 32-36) :

1) **La cooccurrence** : il s'agit de vérifier si la cooccurrence des actants produit une phrase naturelle. Si deux actants donnent une phrase naturelle lorsqu'ils sont pris indépendamment, et non lorsqu'ils sont coordonnés, nous observons que l'actant possède deux sens. Par exemple :

a) *La souche bactérienne est sensible à la vancomycine*

b) *Le test FAT est plus sensible que le test VDRL à tous les stades de la syphilis.*

c) **La souche bactérienne et le test FAT sont sensibles à la vancomycine.*

2) **La substitution synonymique** : lorsqu'il est possible de substituer un actant par un même synonyme dans deux contextes différents, nous pouvons déduire qu'il a le même sens dans les deux contextes. Si, au contraire, le synonyme ne peut pas remplacer l'actant dans les deux contextes, il sera question de sens différents. Bae (2006 : 34) donne l'exemple de l'adjectif *sensible* qui peut être remplacé par *détruit* :

a) *Le streptocoque est détruite par {sensible à} l'antibiotique.*

b) *La souche bactérienne est détruite par {sensible à} la vancomycine.*

c) **Les cellules de mammifère sont détruites par {sensibles à} l'infection virale.*

Parmi ces énoncés, nous observons que les deux premiers exemples peuvent accepter la substitution sans observer de changement dans le sens de l'énoncé. Toutefois, le troisième exemple est inusité lorsque *sensible* est remplacé par *détruit*. En effet, le sens réel de la phrase est conservé lorsque *sensible* est remplacé par *exposé*.

3) **La dérivation morphologique différentielle** : Il est d'abord possible d'observer des dérivations différentes d'un domaine à un autre en observant la liste de celles-ci. Par exemple, en médecine, le nom *virus* est la base de *virale* et de *virologie*, alors qu'en informatique, *virus* est la base d'*antivirus*

(L'Homme 2004). Au sein d'un même domaine, il a été observé que les sens peuvent être regroupés selon la forme des dérivations. Par exemple, *sensibilisé*, *sensibilisation* renvoient à un traitement qui augmente la sensibilité, mais *sensibilité* et *hypersensibilité* renvoient à la sensation ou à l'influx nerveux.

- 4) **Groupe de composés** : lorsqu'une langue n'est pas riche en dérivés, comme le coréen et l'anglais, le critère cité en 3) est difficile à vérifier. Dans ce cas, il est possible d'utiliser les termes composés qui contiennent la composante adjectivale pour distinguer le sens des celles-ci. L'auteure donne l'exemple du terme adjectival *être fin* qui est rarement employé seul. Par contre, elle souligne que la composante adjectivale est présente dans une trentaine de termes composés. Bae compare un échantillon de ces termes dont voici les traductions données dans l'article :

- a) *microcirculation*, *microfilaire*, *micromanipulation* et *microradiographie*.
- b) *aiguille fine*, *cytoponction*, *cytoponction par aiguille fine* et *tremblement fin*.

Ici les deux sens observés pour la composante adjectivale sont « petit » et « fin » (Bae 2006 : 36).

- 5) **Relations paradigmaticques autres que la synonymie** : L'application du critère ressemble à celle du critère cité au point 2). Il s'agit cependant de substituer l'adjectif par une unité lexicale qui possède un lien paradigmaticque avec celui-ci. L'auteure reprend terme adjectival *être sensible*. L'exemple de l'auteure montre que l'antonyme de l'adjectif (entre parenthèses) présent dans le premier exemple est *inefficace* et non dans le second.

- a) *L'utilisation du préservatif est sensible (inefficace) à prévenir des infections à condylome acuminés.*
- b) **La souche bactérienne est sensible (inefficace) à la vancomycine.*

Dans le deuxième exemple, l'antonyme de *être sensible* est *résistant* et non *inefficace*. La présence de deux antonymes différents aide à mettre en évidence la présence de deux sens (Bae 2006 : 36).

L'application des critères pour distinguer les différents sens des termes adjectivaux (d'après Bae, les termes étudiés possèdent une moyenne de 1,875 sens), a permis d'établir un lien entre le sens de l'adjectif et la classe sémantique des actants. Bae a toutefois observé que le réseau sémantique consulté pour obtenir cette classe, *CoreNet*⁷, est principalement conçu pour le domaine général. Pour cette raison, l'utilisation des outils existants peut difficilement appuyer la différenciation des sens.

Suite à cette étude, Bae reconnaît que l'application des critères fait souvent appel à l'intuition du terminologue. Par contre, en réussissant une meilleure automatisation du regroupement des classes sémantiques, la méthode pourra mener à une distinction automatique des sens des termes adjectivaux.

2.3 Traitement automatique des adjectifs spécialisés

Dans cette section, nous présentons des études qui ont eu recours à des traitements automatiques pour caractériser les adjectifs. La première porte sur la productivité des adjectifs selon le genre et le domaine; la seconde vise l'étude des relations morphologiques dans les paires nom-adjectif; la troisième cherche à identifier les adjectifs dérivés de noms lors de l'extraction automatique; enfin, la quatrième veut établir une distinction entre les termes complexes et les collocations. Il s'agit de quatre travaux qui se penchent sur l'automatisation de certaines étapes du traitement des adjectifs dans les domaines de l'agriculture (Daille 2001) et de la médecine (Maniez 2002; Grabar et Zweigenbaum 2003; Zweigenbaum, Hadouche et Grabar 2003).

⁷ *CoreNet* est un réseau sémantique multilingue (coréen, japonais et chinois) qui a été développé par KAIST (Korea Advanced Institute of Science and Technology) – KORTERM.

2.3.1 Productivité des adjectifs selon le genre et le domaine

L'étude de Grabar et Zweigenbaum (2003) vise à évaluer la portée des variations morphologiques des adjectifs dans un corpus lorsque le domaine ou le genre changent. À l'aide de différents procédés d'identification des adjectifs, ils souhaitent démontrer que le type d'adjectifs dérivés varie d'un groupe de textes à l'autre. Il tente de vérifier l'affirmation de Baayen (2001 : 208) selon laquelle : « La productivité des catégories morphologiques peut varier de façon substantielle avec des dimensions telles que l'auteur, le style, le registre, le domaine et le lectorat visé ».

La première étape de la méthodologie mise au point par les chercheurs consiste à identifier les dérivés adjectivaux. À partir de six corpus issus de différents domaines médicaux (signes, hématologie, neurologie, néphrologie et cardiologie) et d'un corpus général composé d'extraits du journal *Le Monde*, ils ont d'abord appliqué le principe informatique selon lequel « les mots qui partagent une chaîne de caractères initiale commune suffisamment longue (ex. *sténose/sténotique*) et qui sont rencontrés dans un contexte sémantiquement contraint sont également proches morphologiquement » (Grabar et Zweigenbaum 2003 : 110). Ce principe a servi à créer une liste de couples base/dérivé. Le fait de travailler sur des contextes sémantiquement contraints a permis d'écarter le bruit causé par des paires issues de domaines différents (ex. *tronc/trouble*).

Par la suite, ils ont dressé une liste de suffixes allomorphes qui participent souvent à la construction des adjectifs dérivés (tableau 2.3) de façon à les comparer avec les adjectifs contenus dans les couples préalablement extraits. En fait, ils ont procédé à une double comparaison au moyen de la liste d'allomorphes et de deux lexiques de noms : *Répertoire d'anatomopathologie de la SNOMED internationale* (Côté 1996) et la *Classification statistique internationale des maladies et des problèmes de santé connexes* (OMS 1995). Cette comparaison a permis d'effectuer le dénombrement des dérivés adjectivaux à partir d'une recherche automatique des couples base/dérivé selon le principe énoncé plus tôt. Ils ont ensuite effectué une vérification manuelle des couples pour éliminer les couples erronés. Enfin, ils ont

procédé au décompte des adjectifs d'après leur dérivation, et plus précisément, selon leur suffixe.

Tableau 2.3 Allomorphes regroupés automatiquement puis complétés manuellement (Grabar et Zweigenbaum 2003 : 111)

Suffixe	Allomorphes regroupés	Suffixe	Allomorphes regroupés
-aire	-naire, -ulaire, -culaire	-é	-né, -gé
-al	-ial, -inal, -péritonéal, -arial	-ique	-tique, -atique
-el	-iel, -ionnel, -tiel	-in	-guin, -atin, -llin
-ien	-en, -gien, -arien, -idien, -éen	-ais	-çais, -elais, -nnais, -gais, -lais
-eux	-ineux	-ois	-cois, -çois, -geois

L'étude de Grabar et Zweigenbaum a permis d'observer deux groupes de dérivés. Le premier regroupe les dérivés qui possèdent des sens différents. Par exemple, le nom *accident* a produit les adjectifs *accidentel* (cause) et *accidenté* (résultat). Le deuxième groupe comprend les dérivés qui possèdent des sens proches. Ces dérivés comptent parfois des anglicismes que certains experts du domaine utilisent. Par exemple, pour le nom *bronche*, il est possible de lire l'adjectif *bronchial*, qui est un anglicisme, au lieu de l'adjectif français *bronchique* (Grabar et Zweigenbaum 2003 : 115).

Deuxièmement, les chercheurs ont repris les résultats obtenus lors du dénombrement pour évaluer la productivité des adjectifs dérivés. Dans le cas présent, Grabar et Zweigenbaum mesurent la productivité des procédés morphologiques en appliquant une formule statistique qui s'appuie sur le nombre d'occurrences des dérivés produits au moyen des suffixes allomorphes énumérés précédemment, par rapport au corpus de référence. Le calcul de productivité a servi à mettre en opposition trois types de textes :

- 1) **Les corpus médicaux et un corpus général** : Le corpus général est composé d'articles du journal *Le Monde*. Les résultats recueillis ont permis de constater une plus grande productivité avec des suffixes en *-el*, *-ique*, *-al*, et *-oide* dans les corpus médicaux par rapport au corpus *Le Monde*.
- 2) **Les différents sous-domaines de la médecine** : L'exemple présenté porte sur les sous-domaines de l'hématologie et de la cardiologie. Les résultats montrent que les suffixes en *-ique*, *-if*, *-el*, *-aire* et *-ien* sont plus productifs dans le sous domaine de l'hématologie.
- 3) **Les genres de documents** : L'étude compare les textes trouvés dans Internet et les comptes rendus médicaux. En général, les résultats ont montré une plus grande variété dans les textes du Web. En observant les suffixes de façon plus attentive, Grabar et Zweigenbaum ont observé que : a) les suffixes en *-ique*, *-if*, et *-aire* sont en proportion comparables dans les deux corpus; b) que les suffixes en *-al* sont deux fois moins fréquents dans les compte rendus; et c) que les suffixes en *-el* et *-eux* sont plus fréquents dans les textes du Web.

Malgré les conclusions qui ont pu être tirées du calcul de la productivité, Grabar et Zweigenbaum reconnaissent que leurs travaux doivent être approfondis. Ils ont observé que « le repérage semi-automatique des dérivés adjectivaux [...] induit un biais dans la constitution des données ». Ils proposent alors de réduire ce biais au moyen de règles supplémentaires, règles qui contribuent également à réduire les « silences [au moment de] la collecte des dérivés » (Grabar et Zweigenbaum 2003 : 122). De plus, ils soulignent l'efficacité de la validation manuelle pour réduire le bruit, mais ils reconnaissent également qu'il leur faudra compléter la validation pour assurer des résultats fiables (Grabar et Zweigenbaum 2003 : 122).

2.3.2 Relations morphologiques des adjectifs en corpus médicaux

Le travail de Zweigenbaum, Hadouche et Grabar (2003) a pour but d'enrichir les terminologies médicales grâce à l'analyse de la relation morphologique qui existe

entre les noms et les adjectifs rencontrés dans les textes médicaux. Pour procéder à l'apprentissage des relations morphologiques en corpus, les auteurs proposent une méthode qui vise à palier le manque qui est observé dans les bases de connaissances en français. Avant de procéder à l'analyse des couples nom-adjectif, ils ont mis en forme un corpus de médecine à partir du catalogue CISMeF. Ce site recense plus de 10 000 sites Internet médicaux francophones. Leurs recherches dans le catalogue ont été faites à partir des mots-clés : *signes* et *symptômes*. Le corpus analysé compte 4627 documents pour un total de 2 041 627 occurrences, dont 54 324 lemmes différents. De plus, les auteurs ont analysé une liste de noms tirés du *Répertoire d'anatomopathologie de la SNOMED internationale* (Côté 1996) et appartenant au champ sémantique de l'anatomie.

La méthode proposée par Zweigenbaum, Hadouche et Grabar repose principalement sur deux principes. Le premier consiste à repérer les mots qui partagent une même chaîne de caractères initiaux. Par exemple, *opérer* et *opération* partagent les mêmes quatre premiers caractères, soit *opér-*. Ce principe peut être exprimé par l'énoncé suivante : « recenser les mots qui partagent la même chaîne de caractères initiale de longueur supérieure ou égale à c et qui se trouvent souvent dans une même fenêtre de M mots » (Zweigenbaum, Hadouche et Grabar 2003 : 287). Dans l'énoncé, c représente le nombre de caractères devant être identiques et M , le nombre maximal de mots pouvant séparer un nom et son dérivé pour que le couple soit jugé correct. Le deuxième principe n'est pas d'ordre morphologique. Il vise plutôt à repérer les mots qui sont unis par des liens sémantiques. Par exemple, il est juste d'affirmer que les mots *hôpital*, *médecin* et *opérer* partagent un lien, car ils appartiennent au même champ sémantique.

À ces principes, les auteurs ont ajouté des critères qui permettent de réduire le bruit généré lors de l'extraction des couples nom-adjectif. Ces critères de filtrage présentent également quelques caractéristiques des relations morphologiques qui sont observées :

- 1) **Rapport de vraisemblance (valeur d'association)** : Comparaison de « la probabilité d'observer le nombre de cooccurrences du mot m dans l'hypothèse H_1 , où les mots sont indépendants du champ sémantique et la probabilité d'observer leur nombre de cooccurrences dans l'hypothèse H_2 , où les mots sont dépendants » (Zweigenbaum, Hadouche et Grabar 2003 : 287). Un seuil est fixé sous lequel les couples cooccurrents sont rejetés.
- 2) **Rejet des dérivations régressives** : Rejet des dérivés produits par l'ajout un suffixe à la racine du mot comptant moins de caractères que le mot de base (ex. *sacrum/sacré*) (Zweigenbaum, Hadouche et Grabar 2003 : 288).
- 3) **Rejet des adjectifs composés** : Rejet des adjectifs construits à partir de compositions savantes, la plupart du temps grecques ou latines (ex. *bronche/bronchopneumonique*) (Zweigenbaum, Hadouche et Grabar 2003 : 288).
- 4) **Fréquence des « règles »** : « Fréquence des règles constituées d'un même opérateur morphologique qui s'applique souvent à plus d'un mot » (ex. *-e/-ique* dans *kyste/kystique*) (Zweigenbaum, Hadouche et Grabar 2003 : 288).

Une fois l'application des filtres terminée, les auteurs ont procédé à l'évaluation de la liste dans le but d'observer la précision de la méthode de sélection des couples nom-dérivé. Dans un premier temps, ils ont évalué la précision générale de la liste obtenue. Pour y parvenir, ils ont trié les couples en ordre décroissant selon leur valeur d'association (voir le premier critère de filtrage). Ensuite, ils ont examiné la proportion cumulée de couples corrects par rapport au nombre de couples examinés et au pourcentage local (par tranche de 200 couples) de ces couples. Dans un deuxième temps, les auteurs se sont penchés sur la précision et le rappel des dérivés adjectivaux. Pour ce faire, ils ont observé les noms extraits du champ sémantique de l'anatomie dans *SNOMED* qui commencent par *a*. Ils ont enfin procédé à la recherche des dérivés et des couples générés à partir de ces noms, dans le corpus.

En utilisant une fenêtre de 150 mots ($m=150$ mots) et en s'appuyant sur la méthode présentée précédemment, les auteurs ont obtenu une liste de 48 002 couples.

L'analyse de cette liste a permis d'évaluer la précision de la méthode selon trois angles (Zweigenbaum, Hadouche et Grabar 2003 : 290-291) :

- 1) **Proportion cumulée** : La précision est de 78 %. Parmi les 5000 occurrences ayant les valeurs d'association les plus élevées, 3778 étaient correctes.
- 2) **Proportion locale** : Pour la tranche des occurrences 4801 à 5000, la précision est de 71,5 %. Il est intéressant d'observer que les courbes de résultats (proportions cumulée et locale) diminuent de façon proportionnelle à la valeur d'association.
- 3) **Échantillon d'anatomie commençant par a** : Sur les 45 noms contenus dans cet échantillon, 17 ont produit un dérivé et 13 ont été jugés corrects. Parmi les adjectifs qui ne sont pas considérés comme étant attendus (qui ne relèvent pas spécifiquement du domaine médical), les auteurs donnent l'exemple de *sang/sanglant (sanguin)*, ou *poil/poilu (pileux)*.

À la lumière de ces résultats, les auteurs ont observé que la méthode permet de repérer un grand nombre de couples. Le taux de précision est de 71 % pour les 9000 premiers couples (tri selon la valeur d'association) et de 43,9 % pour la totalité, soit 21 000 pour 48 000 couples obtenus. Les critères de filtrage permettent d'intercepter un grand nombre d'erreurs. Sur l'échantillon d'anatomie tiré de *SNOMED*, la précision était de 85 à 91 %. Par contre seuls 32 à 34 % d'entre eux ont été trouvés dans le corpus. Enfin, l'étude a permis de mettre en évidence les limites de la méthode : 1) l'analyse manuelle des 48 000 couples est une tâche fastidieuse, qui risque de laisser passer des erreurs; et 2) la contrainte posée par les critères de filtrage écarte certains couples qui sont corrects (ex. *ampoule/ampullaire*).

2.3.3 Identification automatique des adjectifs dérivés de noms

L'étude que nous présentons ici vise à identifier les syntagmes nominaux qui contiennent des adjectifs relationnels⁸, ainsi que les syntagmes prépositionnels qui peuvent les paraphraser. De plus, Daille (2001) souhaite vérifier le caractère de dénomination de ces adjectifs et évaluer ce caractère dans les syntagmes nominaux où ils apparaissent.

D'après l'auteure, l'identification des adjectifs relationnels et des paraphrases prépositionnelles peut servir : 1) au regroupement des formes synonymiques dans l'acquisition de terminologies; 2) à la mise à jour des thésaurus par le remplacement des propositions prépositionnelles par l'adjectif relationnel; et 3) à la désambiguïsation du sens des adjectifs relationnels dans les dictionnaires unilingues et bilingues, au moyen de paraphrases prépositionnelles. Par cette étude, Daille tente de confirmer que les syntagmes de forme *Nom-Adjectif relationnel* sont synonymes des syntagmes *Nom1-préposition-Nom2*, dans lesquels le *Nom2* est la base nominale de l'adjectif relationnel.

Le corpus qui a été mis en forme dans le cadre de l'étude est composé de 7272 résumés de textes portant sur l'agriculture. La première étape dans l'identification des adjectifs relationnels consiste à en définir les propriétés linguistiques. Les caractéristiques présentées par Daille sont morphologiques, paraphrastiques, syntaxiques et sémantiques. La liste et la description des caractéristiques utilisées sont similaires à celles qui ont été présentées à la section 1.2.1.2.1.

En ce qui concerne le traitement informatique, Daille a procédé au traitement du corpus pour en extraire les adjectifs relationnels et les paraphrases prépositionnelles. Pour y parvenir, elle procède à l'étiquetage du corpus au moyen de

⁸ Les adjectifs relationnels tels que les définit Daille correspondent aux adjectifs dérivés de noms que nous étudions. Par fidélité au travail de l'auteure, nous avons conservé la terminologie utilisée au cours de l'étude présentée.

WinBrill puis à l'extraction automatique de termes de base au moyen de trois patrons (Daille 2001 : 155) :

- 1) **Nom1 – Adj** : ex. *emballage biodégradable*
- 2) **Nom1 – Prép (dét.) Nom2** : ex. *protéine de poisson*
- 3) **Nom1 – à Vinf** : ex. *viande à griller*

Par la suite, l'auteure vérifie si les adjectifs étaient construits à partir d'une base nominale suffixée (tableau 2.4) pour éliminer, entre autres, les adjectifs verbaux. La troisième étape consiste à identifier les bases nominales des adjectifs retenus pour ensuite jumeler les adjectifs, ainsi que leur base nominale, précédemment identifiée, aux occurrences du patron *Nom1 – Prép – Nom2*. Enfin, il s'agit de vérifier si l'adjectif peut remplacer la proposition prépositionnelle avant de vérifier l'occurrence des adjectifs dans le thésaurus AGROVOC (*Thésaurus agricole plurilingue développé par l'Organisation des Nations Unies pour l'alimentation et l'agriculture (FAO)*).

Tableau 2.4 Suffixes utilisés dans les règles de transformation (Daille 2001 : 157)

Suffixe	Règle	Exemple
-al	-al/	<i>national/nation</i>
-aire	-aire/	<i>dentaire/dent</i>
-atif	-atif/+e	<i>normatif/norme</i>
-é	-é/+e	<i>carboné/carbone</i>
-el	-el/+e	<i>industriel/industrie</i>
-er	-er/+e	<i>paysager/paysage</i>
-eux	-eux/+e	<i>veineux/veine</i>
-ien	-ien/+ie	<i>bactérien/bactérie</i>
-ier	-ier/	<i>fruitier/fruit</i>
-if	-if/	<i>sportif/sport</i>
-in	-in/	<i>porcin/porc</i>
-ique	-ique/+ie	<i>graphique/graphie</i>
-iste	-iste/+isme	<i>nationalisme/nationaliste</i>
-oire	-oire/ion	<i>inflammatoire/inflammation</i>

L'extraction automatique des différents patrons a permis de recenser les occurrences de ces derniers (tableau 2.5). Sur un échantillon de 124 adjectifs relationnels correctement identifiés, 68 apparaissent dans le thésaurus comme épithète du nom et 73 apparaissent dans les entrées de leur base nominale. Parmi ces derniers, 43 possèdent un caractère relationnel et non les 30 autres. Enfin, seuls 11 adjectifs

n'apparaissent pas du tout dans le thésaurus. En s'attardant à l'étude des patrons extraits, Daille constate que le patron syntaxique *Nom-AdjRel* (1) est beaucoup plus stable que le patron *Nom1 – prép – Nom2* (2), car seules 10 % des occurrences du patron (1) acceptent les variations syntaxiques, du type coordination avec un autre adjectif relationnel (ex. *produit alimentaire/produit alimentaire et agricole*), contrairement à 60 % pour le patron (2) (ex. *acide vinique = acide (prép) vin – acide organique du vin, acide aminé du vin, acide tartrique dans le vin, etc.*).

Tableau 2.5 Dénombrement des patrons extraits du corpus d'agriculture (AGRO)
(Daille 2001 : 159 [trad. libre])

Nombre d'occurrences par patron	1	≥ 2	Total
Nom – Adj	12 344	4778	17 122
Nom1 – prép – Nom2	17 232	5949	23 181
Nom – à – Vinf	203	16	219
Total	29 779	10 743	40 522

À partir des résultats obtenus au cours de cette étude, Daille conclut que les syntagmes nominaux qui contiennent un adjectif relationnel véhiculent souvent plus d'informations sur le sens qu'un syntagme construit au moyen d'un simple patron prépositionnel. De plus, elle observe que la méthode utilisée s'est révélée efficace pour l'extraction des adjectifs relationnels et des formes pouvant éventuellement être remplacées par des adjectifs relationnels.

2.3.4 Distinction entre les termes complexes et les collocations

La dernière étude que nous présentons vise à distinguer les séquences construites selon le patron syntaxique *adjectif-nom* en anglais médical, à savoir les termes complexes de la langue médicale des simples collocations. En parvenant à distinguer ces deux types d'expressions, Maniez (2002) tente de vérifier si les expressions de patrons syntaxiques identiques peuvent être repérées automatiquement

grâce à un étiquetage morpho-syntaxique. La distinction des deux groupes repose, quant à elle, sur des caractéristiques sémantiques.

Pour mener son étude, Maniez a mis en forme un corpus anglais qui regroupe 58 articles du *Journal of the American Medical Association* et qui compte 134 000 mots. Après avoir étiqueté le corpus, il a procédé à l'acquisition automatique de séquences d'après les patrons proposés par Clas (1994 : 578) :

- 1) Verbe et nom, ou le verbe a un contenu très général proche de « faire » :
ex. prononcer un discours;
- 2) Nom et adjectif : *ex. rude épreuve, marque distinctive;*
- 3) Adverbe et adjectif : *ex. vachement bon;*
- 4) Verbe et adverbe : *ex. boire goulûment;*
- 5) Nom (sujet) et verbe : *ex. la cloche sonne, le chat miaule;*
- 6) Séquence qui marque de la quantité (unité collective) du nom : *ex. essaim d'abeilles.*

Par la suite, il a effectué le calcul automatisé des occurrences des séquences à partir du patron 2), à savoir *nom-adjectif*. Le dénombrement a permis d'isoler 1057 combinaisons. Toutefois, pour isoler les collocations, Maniez a choisi d'éliminer manuellement certains adjectifs dérivés de noms spécialisés (tels qu'identifiés par Daille (2001)), comme l'adjectif *metastatic* et les combinaisons dont les adverbes (*ex. therefore, perhaps*) ne modifient pas l'adjectif. Après ce filtrage, il est resté 869 occurrences, dont 596 combinaisons distinctes et 342 adjectifs différents. Une fois les collocations restantes vérifiées manuellement, une recherche en corpus des adjectifs préalablement identifiés a permis de repérer 3121 occurrences de la combinaison *adjectif-nom* qui regroupaient les 342 adjectifs et dont 2034 sont des combinaisons distinctes.

À partir des données obtenues dans le cadre de cette étude, Maniez a été en mesure de différencier les termes complexes, qui contiennent souvent des adjectifs dérivés et dans lesquels peuvent difficilement se glisser des adverbes, des collocations correspondant au patron *adjectif-nom* dont le nombre d'occurrences témoignent de la place importante qu'elles tiennent dans les corpus analysés.

2.4 Contribution de la présente étude

Les travaux antérieurs que nous venons de passer en revue soulignent l'intérêt de tenir compte de la partie du discours de l'adjectif (Cabré 1998; Daille 2001; L'Homme 2004a) et, *a fortiori*, de l'adjectif médical (Maniez 2001; Normand et Bourigault 2001; Grabar et Zweigenbaum 2003; Zweigenbaum, Hadouche et Grabar 2003; Bae 2006). Ces travaux soulignent également l'intérêt d'envisager cette partie du discours en corpus, de procéder à la distinction des types d'adjectifs et de désambiguïser le sens des adjectifs dérivés de noms.

Dans un premier temps, nous avons présenté les éléments théoriques qui ont contribué au repérage et à l'analyse des adjectifs dérivés de noms. Nous pouvons d'abord souligner les travaux de Zweigenbaum, Hadouche et Grabar (2003) et de Grabar et Zweigenbaum (2003) sur l'apprentissage des relations morphologiques qui ont permis de clarifier les modes de production des adjectifs. Nous pouvons également mentionner les travaux de Daille (2001) sur la différenciation des adjectifs dérivés noms qui ont été particulièrement utiles lors du classement des adjectifs.

Dans un deuxième temps, nous avons voulu mettre en évidence des travaux qui nous ont permis de nous sensibiliser au travail sur corpus et au traitement automatique de nos données. Notons les travaux de Normand et Bourigault (2004) sur l'analyse syntaxique des syntagmes nominaux. Enfin, nous avons appuyé certains éléments de notre méthodologie et de l'analyse des adjectifs dérivés de noms sur les travaux de Daille (2001) sur l'identification des adjectifs dits relationels et de L'Homme (2004) sur le rôle des adjectifs dans la structuration des terminologies.

Comme Bae (2006) le fait dans son étude des termes adjectivaux coréens (les adjectifs qualificatifs), nous voulons désambiguïser des adjectifs spécialisés de façon à créer de nouvelles ressources terminologiques et à enrichir celles déjà existantes. Nos études sont presque parallèles, mais elles se différencient principalement par les caractéristiques linguistiques de nos langues de travail respectives, à savoir le coréen et le français. De plus, Bae travaille sur les adjectifs qualificatifs utilisés en médecine alors que nous travaillons sur les adjectifs dérivés de noms. Cela dit, notre étude se veut une suite aux travaux déjà effectués sur les adjectifs dérivés de nom en français. En effet, notre travail est :

- 1) la mise en application des caractéristiques des adjectifs, tels que présentés dans les travaux cités;
- 2) l'étude approfondie des adjectifs dérivés de noms spécifiques à notre corpus pour permettre une compréhension en contexte de leur sens et de leur utilisation;
- 3) l'élaboration d'un modèle d'encodage qui utilise les observations faites au cours des dernières années sur ce groupe d'adjectifs et qui se veut un outil d'aide à la rédaction pour les traducteurs et les rédacteurs médicaux.

Chapitre 3 : Méthodologie

Le cadre méthodologique qui est présenté au fil des prochaines pages constitue le fondement du présent projet de recherche. La mise en forme de notre corpus et son traitement automatique sont des étapes essentielles qui nous permettent de dresser une liste d'adjectifs. Nous classons ces adjectifs pour ensuite procéder à l'analyse des adjectifs dérivés de noms. Cette analyse mène enfin à l'élaboration d'un modèle d'encodage.

3.1 Mise en forme du corpus

Dans cette section, nous verrons les étapes qui nous ont permis de mettre en forme notre corpus. Dans un premier temps, nous expliquons le concept de « corpus » tel que vu par Bowker et Pearson (2002). Par la suite, nous énumérons les critères de départ qui ont guidé nos recherches de textes médicaux puis nous présentons les critères de sélection des textes qui composent notre corpus.

3.1.1 Définition

Depuis quelques années, la place qu'occupent les corpus dans les travaux terminologiques est devenue suffisamment importante pour constituer un thème de recherche. Bowker et Pearson définissent le corpus comme suit : « a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria » (2002 : 9). Le corpus repose donc sur quatre points importants : l'authenticité, la taille, le format électronique et les critères de sélection.

D'abord, l'authenticité du corpus spécialisé se traduit par le choix d'un échantillon représentatif de la langue qui est utilisée dans la présentation des connaissances auprès des professionnels d'un domaine donné. Par la suite, le corpus doit être suffisamment volumineux pour permettre de tirer des conclusions d'analyse satisfaisantes sur l'objet étudié. Le nombre de textes, de pages ou de mots n'est donc pas une valeur figée, car il varie selon l'ampleur du projet et la richesse des textes

trouvés. De plus, le traitement informatique des corpus au moment d'effectuer la recherche terminologique rend essentiel l'accès à un corpus en format électronique. Enfin, le dernier point fondamental à considérer au moment de mettre en forme un corpus est l'établissement des critères de sélection. Ces critères assurent la pertinence des textes sélectionnés tout en facilitant et en organisant les recherches de textes électroniques dans Internet (Bowker et Pearson 2002).

Dans les travaux de terminologie et de linguistique, l'utilisation du corpus prend sans cesse de l'importance. Il est évident que le corpus ne renferme pas toutes les données notionnelles des dictionnaires, mais il est essentiel à la mise en contexte des termes. Le corpus sert donc de complément aux ressources dictionnairiques traditionnelles tout en confirmant ou en infirmant parfois les pistes de recherches avancées (Bowker et Pearson 2002).

3.1.2 Critères de sélection

Parmi les critères qui sont proposés par Bowker et Pearson (2002 : 12-13, 50-51), nous avons privilégié ceux qui conviennent le plus au contexte de notre étude. Dans un premier temps, nous exposons quelques critères de départ qui ont été fixés avant de procéder à la recherche des textes. Par la suite, nous présentons les critères qui nous ont guidée dans la sélection définitive des textes.

3.1.2.1 Critères de départ

Le premier critère que nous avons choisi, est la langue. Dans le cadre de la présente étude, nous avons choisi d'analyser les adjectifs en corpus français. En nous concentrant sur la désambiguïsation des adjectifs médicaux en corpus unilingue, nous croyons être en mesure de produire un outil qui permettra aux traducteurs de produire des textes qui conviennent davantage à un lectorat expert. Ce dernier point nous mène au deuxième critère qui consiste à choisir un corpus spécialisé. Nous avons choisi des

textes médicaux qui s'adressent principalement à des experts dans le domaine médical, soit à des médecins ou à des étudiants en médecine.

Avant de compléter la mise en forme de notre corpus, nous avons procédé au traitement informatique d'un échantillon au moyen de TermoStat (voir section 3.2.1) pour extraire les adjectifs et observer le nombre et la fréquence des adjectifs dérivés de noms. Cet exercice nous a permis de définir la taille du corpus à mettre en forme pour obtenir de données en quantité suffisante, soit 200 000 mots.

3.1.2.2 Critères de sélection des textes

Au moment de sélectionner les textes constituant notre corpus, nous nous sommes d'abord attardée sur la forme du texte (Bowker et Pearson 2002 : 49). Premièrement, nous avons jugé essentiel de travailler sur les textes complets pour pouvoir placer les candidats termes en contexte réel. En fait, il s'agit de choisir des textes autonomes (article entier ou section de manuel traitant d'un sujet dans son entier). La taille et le nombre de mots de chaque texte sont parfois fixés dès le début de l'étude, mais nous avons choisi de ne pas soumettre nos recherches à de telles limites, car nous nous doutons que les textes spécialisés en médecine rédigés en français sont généralement beaucoup moins fréquents en anglais. Nous avons donc recueilli 67 textes comptant de 1000 à 10 000 mots chacun. Le corpus que nous analysons compte près de 220 000 mots.

Au moment de sélectionner les textes, il a d'abord fallu choisir un sous-domaine de la médecine pour concentrer nos recherches et éviter d'obtenir un corpus qui serait sans pertinence dans le cadre d'une analyse fine des adjectifs. Après avoir prélevé des échantillons dans un certain nombre de sous-domaines, notre choix s'est arrêté sur l'infectiologie. Ce dernier est défini par Delamare comme « l'étude des maladies [qui sont provoquées par un agent infectieux] » (2002 : 435). Dans le cadre du présent travail, nous avons regroupé des textes portant sur la description, le diagnostic et le traitement de quelques maladies infectieuses d'origine virologique,

bactériologique et fongique. Le choix du sous-domaine est soutenu par sa multidisciplinarité, car l'infectiologie est à la fois descriptive (diagnostic) et pratique (traitement). Enfin, il s'agit d'une discipline dont la terminologie compte un nombre important d'adjectifs dérivés de noms.

Toujours selon Bowker et Pearson (2002 : 50-51), il est important de faire des choix dans le type de textes analysés et nous avons concentré nos recherches sur trois types de textes. Premièrement, nous avons cherché des articles qui ont été publiés dans les revues scientifiques dans les domaines de la médecine, de l'infectiologie, de la virologie et de l'immunologie. Pour être retenus, les articles doivent être rédigés par des experts et être destinés à un lectorat expert. Parmi les ouvrages consultés, nous comptons, entre autres, *Médecine et maladies infectieuses*, *EMC - Maladies Infectieuses*, *EMC - Médecine*.

Deuxièmement, nous avons choisi les notes de cours mises en ligne par les facultés de médecine de certaines universités françaises. Les notes portent sur les maladies infectieuses ainsi que sur les méthodes de diagnostic et le traitement de celles-ci. Les textes ont été rédigés par des experts du domaine médical et ils sont destinés à un lectorat universitaire, particulièrement à des étudiants de médecine.

Troisièmement, nous avons retenu un certain nombre de fiches techniques émises, principalement, par l'Organisation mondiale de la santé (OMS). Dans ce cas, les textes ont été rédigés par des experts et ils s'adressent aux médecins tout en étant accessibles au public. Les textes qui constituent notre corpus sont rédigés par des auteurs français, québécois ou suisses (OMS).

Enfin, il est intéressant de préciser que notre corpus est principalement composé de textes parus après 2000. De cette façon, notre corpus contient une terminologie actuelle.

À ces critères, nous avons ajouté deux autres critères qui nous permettent d'assurer l'authenticité de notre corpus. Nous avons d'abord privilégié les textes qui sont signés. Bien que certains textes non signés soient intéressants, nous avons systématiquement rejeté les textes non signés, à l'exception des fiches de l'OMS. Dans le cas de l'OMS, il s'agit de fiches rédigées par des collectifs pour le compte de l'Organisation. De plus, l'authenticité des textes est soutenue par la notoriété de l'OMS dans le domaine médical. Enfin, nous avons porté une attention particulière à la traduction des textes. Même si certains auteurs de langue française ont un nom d'apparence anglaise, et vice versa, nous avons privilégié les auteurs dont le nom était visiblement français. Aussi, nous avons choisi des articles qui ont paru dans des revues scientifiques dont la langue principale de publication est le français.

Le Tableau 3.1 rappelle les caractéristiques de notre corpus qui est composé de 67 textes variant de 1000 à 10 000 mots chacun. Une liste des textes qui composent le corpus a été mise à l'annexe A.

Tableau 3.1. Critères de sélection des textes du corpus

	Critère	Particularité pour notre étude
Critères de départ	Langue	Français
	Spécialisation	1) Domaine médical 2) Lectorat expert
	Nombre de mots	200 000 mots
Critères de sélection des textes	Sous-domaine	Infectiologie
	Forme	1) Textes complets 1) Taille variable (1000 à 10 000 mots)
	Type	1) Articles scientifiques 2) Notes de cours universitaires 3) Fiches techniques
	Année de parution	Principalement ≥ 2000
	Source	1) Auteurs de langue française 2) Revues de langue française

3.2 Production d'une liste préliminaire de candidats adjectifs

Cette section porte sur la méthode d'extraction et les critères de sélection des candidats adjectifs (forme lexicale qui a été extraite comme étant un adjectif) qui ont menés à l'obtention de la liste d'adjectifs étudiés. Dans un premier temps, nous avons procédé à l'extraction automatique des adjectifs de notre corpus, pour ensuite nettoyer la liste générée par l'extracteur, puis effectuer une lemmatisation des adjectifs pour éliminer les doublons de la liste de candidats adjectifs.

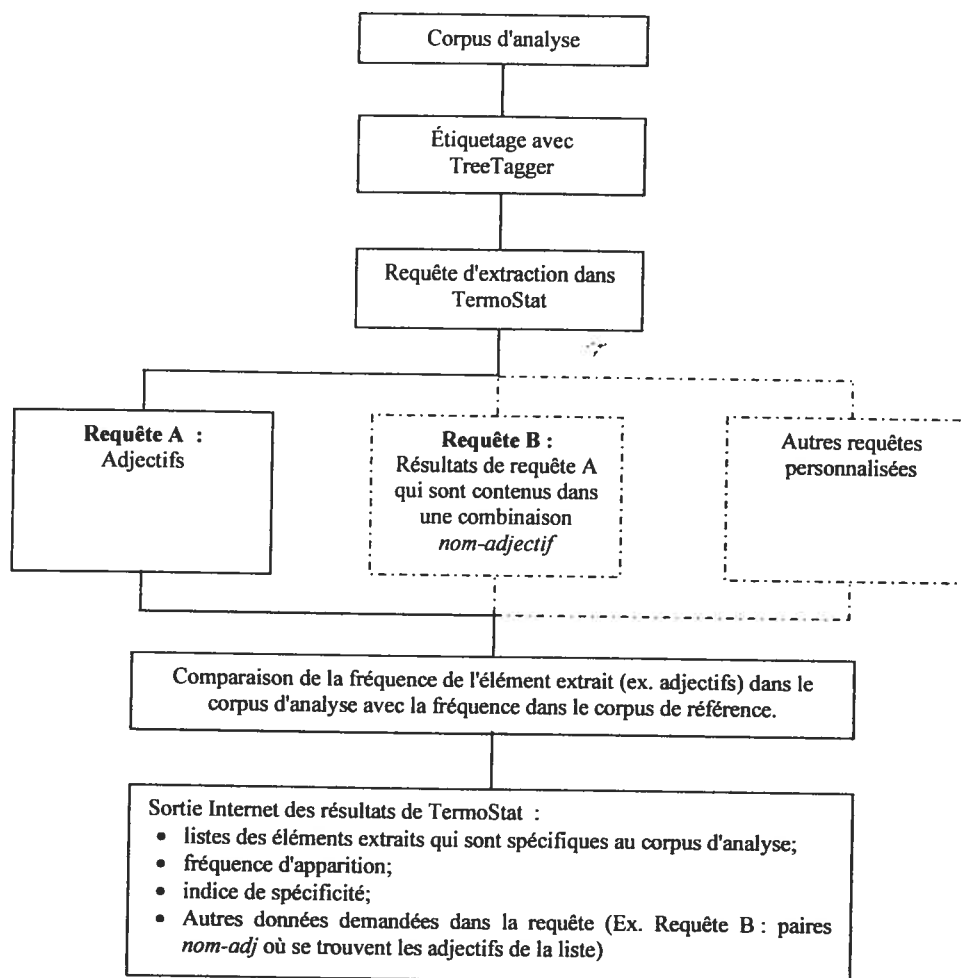
3.2.1 Extraction et spécificité des adjectifs

Une fois les textes sélectionnés, ils ont été regroupés dans un fichier en format texte (.txt) de façon à pouvoir être traités au moyen du logiciel TermoStat conçu par Drouin (2003). Le logiciel étiquette les parties du discours du corpus au moyen de l'étiqueteur morpho-syntaxique TreeTagger (Schmid 1994). Par la suite, TermoStat extrait les adjectifs et leur attribue un indice de spécificité. Ce dernier paramètre est utilisé lors de l'évaluation des adjectifs qui seront sélectionnés pour l'analyse de leur sens.

Dans un nombre grandissant de travaux terminologiques, la spécificité est utilisée comme point de départ pour l'acquisition automatique de la terminologie (Lemay 2003; Drouin 2004; L'Homme 2005; Racine 2006). Au moment d'attribuer un indice de spécificité aux candidats adjectifs extraits, TermoStat compare le « lexique d'un corpus technique (corpus d'analyse) à celui d'un corpus non technique (corpus de référence) » (Carreno Cruz 2005 : 65). Dans le cas présent, le corpus de référence utilisé lors du traitement automatique est composé des textes parus dans le journal *Le Monde* pour l'année 2002, soit près de 30 millions de mots. Une telle comparaison permet d'obtenir un indice de spécificité au moyen de calculs statistiques. Ces calculs reposent sur la comparaison des fréquences des éléments recherchés dans le corpus d'analyse par rapport à la fréquence de ces mêmes éléments dans le corpus de référence. En d'autres mots, l'indice de spécificité correspond à l'écart entre la fréquence estimée, par rapport à la fréquence dans le corpus de référence (*Le Monde*), et la fréquence observée dans le corpus d'analyse.

En plus de nous informer sur la spécificité des adjectifs par rapport au corpus, il est intéressant de soulever la flexibilité des requêtes pouvant être effectuées en vue de l'extraction. En effet, il est possible de personnaliser la requête d'extraction de façon à mieux répondre aux besoins d'un projet donné. Dans notre cas, la requête a visé exclusivement les adjectifs, car nous voulions d'abord avoir une liste complète des adjectifs spécifiques à notre corpus (tous types confondus) pour ensuite nous attarder aux adjectifs contenus dans les combinaisons *nom-adjectif*. Par contre, si nous avions choisi de n'étudier que ces derniers adjectifs, nous aurions pu effectuer une requête qui nous aurait donné les adjectifs, ainsi que les syntagmes dont ils font partie.

La figure 3.1 permet de visualiser les étapes de traitement du corpus qui ont mené à l'obtention de la liste de candidats adjectifs.



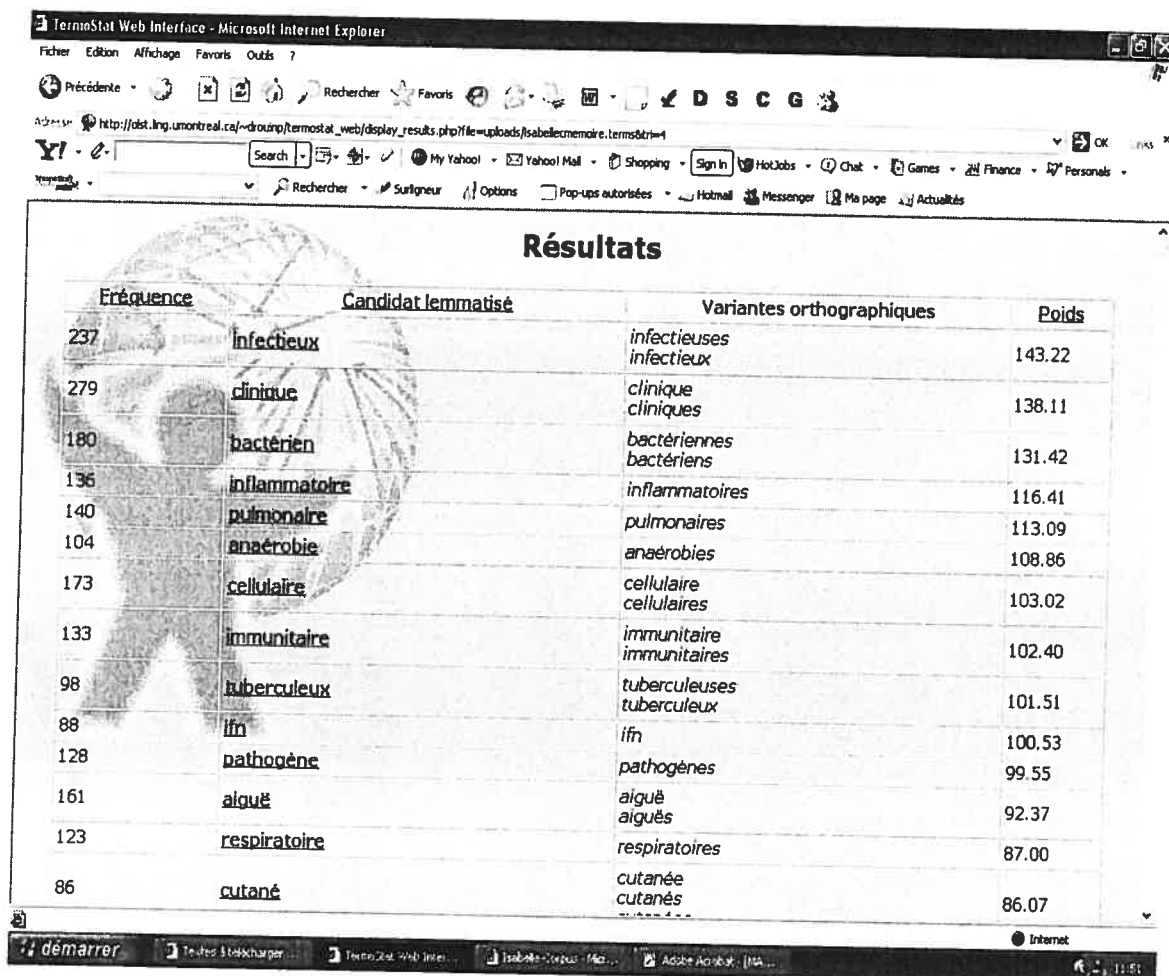
Légende

- Traitement effectué dans le cadre du projet
- - - - Autres requêtes possibles.

Figure 3.1 Rappel des étapes de traitement du corpus en vue de l'extraction des candidats adjectifs

Une fois le traitement terminé, TermoStat produit un tableau de résultats (figure 3.2) qui contient les données qui nous permettent de préparer la liste de candidats adjectifs. La première colonne donne la fréquence de chaque adjectif dans le corpus d'analyse, la deuxième présente la forme lemmatisée du candidat adjectif, la troisième présente les formes fléchies pour un même lemme qui sont présentes dans le corpus d'analyse et la quatrième donne l'indice de spécificité qui a été attribué à l'adjectif. L'interface de présentation des résultats est un module qui est consultable

par Internet et qui permet de trier les données selon la fréquence, la forme lemmatisée ou l'indice de spécificité. L'échantillon qui suit est trié en ordre décroissant d'indice de spécificité (sous le titre *Poids*).



Résultats

Fréquence	Candidat lemmatisé	Variantes orthographiques	Poids
237	infectieux	infectieuses infectieux	143.22
279	clinique	clinique cliniques	138.11
180	bactérien	bactériennes bactériens	131.42
136	inflammatoire	inflammatoires	116.41
140	pulmonaire	pulmonaires	113.09
104	anaérobie	anaérobies	108.86
173	cellulaire	cellulaire cellulaires	103.02
133	immunitaire	immunitaire immunitaires	102.40
98	tuberculeux	tuberculeuses tuberculeux	101.51
88	ifn	ifn	100.53
128	pathogène	pathogènes	99.55
161	aiguë	aiguë aiguës	92.37
123	respiratoire	respiratoires	87.00
86	cutané	cutanée cutanés	86.07

Figure 3.2 Liste des candidats adjectifs, triés selon leur spécificité, générée par l'interface Internet de TernoStat

Pour consulter les contextes qui contiennent l'un ou l'autre adjectif, il suffit de cliquer sur l'entrée du tableau apparaissant dans la colonne *Forme lemmatisée*. Il devient alors possible de voir les phrases qui contiennent l'adjectif, comme le présente la figure 3.3.

Contextes

Fermer la fenêtre

5-3 - Immunologie clinique TP4 : Thyroïdite autoimmune , PAN , rejet de greffe rénale Le système immunitaire est omniprésent dans l'organisme et intervient donc dans beaucoup de pathologies .

Chaque virus a un tropisme particulier pour certains types de cellules , qui est responsable des manifestations cliniques .

Il s'y associe toujours une réaction exsudative , très développée au départ , puis en grande partie régressive (ce qui explique la régression des lésions cliniques ; leur stabilisation correspond à la nécrose neuronale vraie) .

Le tableau clinique est celui d'une angiocholite fébrile .

3.6 Les signes cliniques de l'inflammation Ils dépendent du type de l'inflammation : la rougeur , l'œdème , l'augmentation de la chaleur locale , la douleur constituent les signes cardinaux connus depuis Galien (' rubor , tumor , dolor , calor ') .

3.9 Les formes cliniques de l'inflammation La réaction inflammatoire revêt des aspects particuliers qui dépendent de la prédominance d'une des composantes de l'inflammation .

Les avantages de ce test sont : précocité (dès le début des signes) , simplicité (sur urine) , rapidité (en + 15 minutes) , diagnostic tardif (> 2 mois après les signes cliniques) , même après un traitement antibiotique adapté . Donc Bonne valeur prédictive mais ce test ne détecte pas les autres sérogroupes de L .

L'intérêt de ces diverses méthodes se résume d'une part par un bénéfice clinique lié : gain de sensibilité (X 2 par rapport aux méthodes classiques) pour la recherche notamment des Chlamydia génitaux .

Les signes cliniques décrits ci - dessus ainsi qu'un interrogatoire anamnestique soigneux suffiront bien

Terminé

demarrer

Textes à télécharger...

TernoStat Web Inter...

TernoStat Web Inter...

Isabelle-Corpus - FIC...

Adobe Acrobat - [MA...

Internet

12:05

Figure 3.3 Contextes fournis par TernoStat pour l'adjectif *clinique*

Nous avons utilisé le logiciel Syntex (Bourigault *et al.* 2004) pour dresser la liste des syntagmes nominaux et consulter les contextes. Ce logiciel sera présenté à la section 3.5 portant sur l'analyse des adjectifs dérivés de noms.

3.2.2 Nettoyage de la liste de candidats adjectifs

Les résultats du traitement ont été importés dans une base de données Access pour en faciliter le nettoyage et la préparation en vue de l'analyse. La liste importée dans la base de données compte 2597 candidats adjectifs. Nous avons d'abord construit une table présentant la forme canonique des adjectifs, la fréquence

d'apparition dans le corpus, l'indice de spécificité et les formes fléchies (figure 3.4). La table contient également des champs réservés aux commentaires.

Forme canonique	fréquence	indice	forme fléchie	lemmatisation manuelle	Comme
viral	420 197 113224082514		virales_viraux		
infectieux	237 143 224116900643		infectieuses_infectieux		
clinique	279 138 109692767387		clinique_cliniques		
bactérien	180 131 41799843362		bactériennes_bactériens		
inflammatoire	136 116 405140718082		inflammatoires		
pulmonaire	140 113 088471450809		pulmonaires		
anaérobie	104 108 858617719297		anaerobies		
cellulaire	173 103 016404286331		cellulaire_cellulaires		
immunitaire	133 102 40307453405		immunitaire_immunitaires		
tuberculeux	98 101 513146070349		tuberculeuses_tuberculeux		
pathogène	128 99 5541128438415		pathogène_pathogènes	indices vanés, fréq totale	
aigüe	178 92 3655320292529		aigu_aigus_aigüe_aigües		
respiratoire	123 87 0004447171984		respiratoires		
cutané	86 86 0691184681168		cutanée_cutanés_cutanées		
grippal	69 85 7724146296361		grippaux_grippal_grippale_grippales		
antituberculeux	64 84 8789590944516		antituberculeux_antituberculeuse		
digestif	98 84 5502937362455		digestif_digestives_digestifs		
antigénique	61 83 4892110183993		antigéniques		
hémorragique	66 80 3369515330299		hémorragique_hémorragiques		
botulique	61 80 207853953276		botuliques		
immun	55 78 4821629421348		immune_immunes_immuns		
urinaire	91 77 6187431834214		urinaire_urinaires		
pancréatique	53 76 2731913765522		pancréatiques		
hépatique	63 75 6236320127037		hépatique_hépatiques		
aviaire	50 73 9582736009471		aviaire_aviaires		
lymphocytaire	49 73 9115903238401		lymphocytaire_lymphocytaires		
spécifique	240 73 7607941954727		spécifique_spécifiques		
antiviral	49 72 450766099773		antivirale_antiviraux_antivirales		
osseux	71 70 1669082130634		osseuse_osseuses_osseux		

Figure 3.4 Table de la base de données utilisée au moment de nettoyer et de lemmatiser la liste de candidats adjectifs

De cette liste, nous avons retiré les éléments qui ne sont pas des adjectifs pour obtenir une liste 1645 adjectifs. Parmi les éléments qui causent du bruit dans le traitement informatique du corpus et qui ont été extraits par TermoStat comme candidats adjectifs, nous comptons :

1. des abréviations et des acronymes : ex. *IFN, VIH, IgG, IRM*;
2. des lettres seules : ex. *b), h*;

3. des noms, des verbes, des noms propres et des marques de commerce :
ex. *Potomac*, *céphalée*, *Prévost*, *Kikwit* (ville);
4. des valeurs numériques et des unités de mesure : ex. *16 s.*, *100 mg*;
5. des mots grecs et latins (la plupart constituaient des noms d'organismes pathogènes) : ex. *Vibrio cholera*, *Clostridium argentinense*⁹;
6. des symboles : ex. *, = ;
7. des mots anglais : ex. *invest*, *mycotic*, *strain*, *tumor*;
8. des participes présents et passés : ex. *provenant*, *donnant*;
9. des préfixes et des suffixes (seuls dans le texte dont le sens était général) : ex. *semi*, *alpha*, *bêta*.

Le bruit constitue une erreur causée principalement par l'étiquetage. En effet, l'étiqueteur morpho-syntaxique effectue des calculs de probabilité permettant d'attribuer une partie du discours à une unité lexicale. Dans un corpus spécialisé comme celui de cette étude, il arrive que le logiciel rencontre une unité lexicale qui lui est inconnue, ce qui risque de provoquer une erreur d'étiquetage (L'Homme, L'Homme et al. 2002). De plus, lors de l'utilisation d'un étiqueteur morpho-syntaxique, il est possible d'observer des erreurs d'étiquetage dues à la nominalisation de certains adjectifs (ex. *la femme a fait le plein d'essence*) (Arrivé, Gadet et al. 1986). Ce type d'erreur est alors dû aux données qui se trouvent dans le dictionnaire de l'étiqueteur. Dans le cas de l'adjectif *plein*, si le dictionnaire de l'étiqueteur ne l'identifie qu'à titre d'adjectif, sa fonction nominale (ex. *plein d'essence*) risque d'être ignorée et il recevra l'étiquette *adjectif*.

Il arrive parfois que l'adjectif présent dans la liste soit ambigu. L'interface de résultats de TermoStat nous permet alors de consulter le corpus pour en vérifier le contexte et justifier ainsi sa présence ou son retrait de la liste de candidats adjectifs à l'étude. La possibilité de voir l'adjectif en contexte s'est révélée particulièrement utile

⁹ TermoStat a extrait des termes simples. Les noms de bactéries et de virus étaient donc présents dans la liste de façon séparée et la consultation du contexte nous a permis de reconstituer les noms scientifiques tels

pour différencier les participes présents et passés (verbes conjugués) des adjectifs verbaux (voir la description à la section 1.2.1.1.3). La vérification du terme en contexte a également permis d'éliminer les noms comme *céphalée* dont la forme se rapproche de celle d'un participe passé adjectival. Enfin, lorsque le statut d'adjectif restait incertain après une vérification dans le texte, nous avons eu recours à des outils comme des bases de données terminologiques (Grand dictionnaire terminologique (GDT (OQLF 2006) et *TERMIUM Plus* (Bureau de la traduction 2006)) ou à des dictionnaires médicaux et généraux (Garnier, Delamare et al. (2001), Manuila, Manuila et al. (2001) et Petit Robert (Rey et Rey Debove 2004)).

3.2.3 Lemmatisation des adjectifs

« La lemmatisation consiste à ramener les formes fléchies des mots variables à une forme canonique » (L'Homme 2004b). L'adjectif étant une partie du discours qui s'accorde à la fois en genre et en nombre, cette étape est particulièrement importante pour éviter le dédoublement dans les entrées d'un outil terminologique qui pourrait être créé au moyen de la base de données. Dans le cas présent, TreeTagger a effectué une lemmatisation correcte pour la majorité des candidats adjectifs. Toutefois, une centaine d'adjectifs ont dû être lemmatisés manuellement. Parmi ceux-ci, nous avons remarqué que les adjectifs qui n'ont pas été lemmatisés par TreeTagger sont très spécialisés (ex. *toxicogène, épithélioïde*) et ils ne se trouvent certainement pas dans le dictionnaire du module de lemmatisation du logiciel.

Au moment de lemmatiser manuellement les adjectifs situés dans la colonne *forme canonique* de notre base de données (figure 3.4), il ne suffit pas de ramener l'adjectif à sa forme neutre, soit masculin singulier. Nous souhaitons garder une trace de la fréquence totale des formes fléchies. Pour y arriver, nous avons parcouru la liste d'adjectifs et lorsqu'un adjectif était présent sous sa forme fléchie, nous le notions pour ensuite effectuer une recherche afin de trouver les autres flexions présentes dans

la liste. Nous avons pris soin de noter la fréquence de chaque flexion¹⁰ dans le Rapport de lemmatisation manuelle (figure 3.5).

Adjectif lemmatisé	fréq. tot.	Fréquences par flexion				indice. sup.
		m.s. ¹¹	m.p.	f.s.	f.p.	
polymicrobien	9		1	3	5	21
toxigène	46	11	35			62
épithélioïde	28	7	21			48
extrapulmonaire	33	14	19			45
immunodéprimé	30	10	17	1	2	40
subaigu	17	2		13	2	32
mononucléé	12	2			10	30
asymptomatique	34	24	10			50
...						

Figure 3.5 Échantillon du *Rapport de lemmatisation manuelle* (le Rapport est présenté en entier à l'annexe B)

Nous avons ensuite procédé aux corrections des données présentes dans le tableau de la figure 3.4 en effaçant d'abord les entrées des différentes flexions pour ne garder que l'entrée qui possédait l'indice de spécificité le plus élevé. Nous avons également lemmatisé manuellement l'adjectif qui se trouvait dans la colonne *forme canonique*, nous avons remplacé la fréquence de la flexion par la fréquence totale des flexions présentes dans le corpus, puis nous avons ajouté les différentes formes fléchies présentes dans le corpus. Enfin, nous avons ajouté une note, dans la colonne commentaires de la table Access, qui contient les manipulations engendrées par la lemmatisation manuelle :

¹⁰ À la figure 3.5, les flexions sont indiquées de la manière suivante : m.s. (masculin singulier), m.p. (masculin pluriel), f.s. (féminin singulier) et f.p. (féminin pluriel).

¹¹ Puisqu'il s'agit de lemmatiser la forme des unités lexicales dans le but d'éliminer les doublons pouvant se trouver dans la liste de candidats adjectifs, les formes épécènes ont été traitées comme des flexions masculines.

1. **Même indice, même fréquence** : le corpus ne contenait que la forme fléchie et la lemmatisation ne consistait qu'à neutraliser l'adjectif pour obtenir la forme canonique.
2. **Même indice, fréquence totale** : le corpus contenait plus d'une forme fléchie, mais l'indice de spécificité était le même. La fréquence est la somme des fréquences des différentes flexions.
3. **Indices variés, fréquence totale** : le corpus contenait plus d'une forme fléchie et l'indice de spécificité variait d'une flexion à l'autre. L'indice qui apparaît dans la base de données est la valeur supérieure. La fréquence est la somme des fréquences des différentes flexions.

Une fois cette étape complétée, la base de données compte 1645 candidats adjectifs lemmatisés extraits de notre corpus de départ.

Nous sommes consciente que la lemmatisation manuelle a probablement changé quelques indices de spécificité, mais nous ne croyons pas que les corrections apportées à la fréquence d'apparition des formes fléchies dans le corpus aient un impact marqué sur la liste finale des adjectifs étudiés.

3.3 Présentation de la base de données

Pour procéder au classement et à l'analyse des adjectifs, nous avons utilisé la base de données qui a servi au nettoyage des candidats adjectifs et nous y avons créé des formulaires Access. Dans cette section, nous effectuons d'abord un survol des champs de la base de données. Nous nous attardons sur les formulaires Access intitulés *Types d'adjectifs* et *Contextes*. Par la suite, nous présentons la numérotation que nous avons choisie pour identifier les types d'adjectifs et les différentes acceptions. Enfin, nous expliquons pourquoi nous avons choisi d'utiliser la paraphrase comme outil pour distinguer les acceptions.

3.3.1 Description des champs de la base de données utilisée lors de l'analyse

À partir du tableau présenté à la figure 3.4, nous avons produit un formulaire Access (figure 3.6) qui nous permet de réunir les différents éléments servant à l'analyse des adjectifs : la forme canonique (Adjectif), la base nominale (Base), la caractéristique générale qui soutient la classification choisie (Commentaires), les différentes acceptions trouvées dans le corpus (Acception) et la paraphrase qui explique le lien qui unit la base nominale (NomB) au nom situé en tête de syntagme. Enfin, le formulaire donne accès à un deuxième formulaire (figure 3.7) qui présente un ou plusieurs contextes accompagnés de la substitution (Application) de l'adjectif par la paraphrase explicative rédigée pour cette acception de l'adjectif.

The screenshot shows a Microsoft Access form titled "Types d'adjectifs". The form is organized into several sections:

- Top Section:** Contains fields for "Adjectif", "Fréquence", "type", "Index", and "Commentaires".
- Base Section:** A section labeled "Base" with an associated input field.
- Main Section:** A section titled "Acception" and "Paraphrase" containing a table with columns for "Tête(s) de syntagme" and "Contexte(s)".
- Navigation:** A navigation bar at the bottom showing "Enr: 1 sur 1".

Figure 3.6 Formulaire *Types d'adjectifs*

Avant d'aborder le processus d'analyse des adjectifs à proprement parler, nous procédons à une description détaillée du contenu des champs rencontrés dans les formulaires de la base de données :

Adjectif : La forme canonique de l'adjectif est numérotée d'après son type et sa base nominale (voir la section 3.4.2 pour la numérotation).

Fréquence : La fréquence représente le nombre total d'occurrences de l'adjectif dans le corpus. En d'autres mots, la valeur comprend les occurrences des adjectifs peu importe le type et la différence de base nominale. La fréquence nous informe de l'importance de l'adjectif dans le corpus et non de la répartition des différentes acceptions dans le texte.

Indice : L'indice de spécificité nous permet d'évaluer à quel point l'adjectif est spécifique à notre corpus médical. Comme nous l'avons précisé à la section 3.2.3, l'indice représente la spécificité la plus élevée qui a été attribuée à l'une ou l'autre des formes fléchies de l'adjectif lors du traitement du corpus.

Type : Le type d'adjectif renvoie à la classification de l'adjectif étudié, telle qu'elle est décrite à la section 1.2.

Commentaire : Le commentaire reprend le critère général qui sert de squelette à la paraphrase définitoire expliquant le sens de l'adjectif. Ce champ permet également d'inscrire des notes sur l'analyse de l'adjectif. Par exemple, l'explication du lien entre le nom de base de l'adjectif et la forme de l'adjectif (ex. du grec *hepatos* pour rein).

Base : Le nom de base correspond au nom dont le sens est présent dans celui de l'adjectif dérivé de nom. Même si les adjectifs dérivés de verbes ne font pas partie de l'étude proprement dite, nous avons également noté la base verbale de ces derniers au moment de la sélection des adjectifs. S'il s'agit d'un autre type d'adjectif, nous n'avons rien inscrit dans ce champ.

Acception : Le champ permet de distinguer les acceptions puisque nous reprenons la forme canonique et nous y ajoutons une numérotation (voir section 3.4.2).

Paraphrase : La paraphrase est l'énoncé définitoire expliquant l'acception de l'adjectif dont il est question. La structure de la paraphrase renvoie au modèle paraphrastique qui se trouve dans le champ commentaire.

Tête(s) de syntagme : Les têtes de syntagmes sont les noms avec lesquels l'adjectif est combiné pour une acception donnée. La liste ne représente qu'un échantillon

des noms de tête pouvant se combiner avec l'acception, soit les noms rencontrés dans notre corpus.

Contexte(s) : Le bouton *Contexte(s)* renvoie à un autre formulaire (illustré à la figure 3.7) qui présente un ou plusieurs contextes tirés de notre corpus.

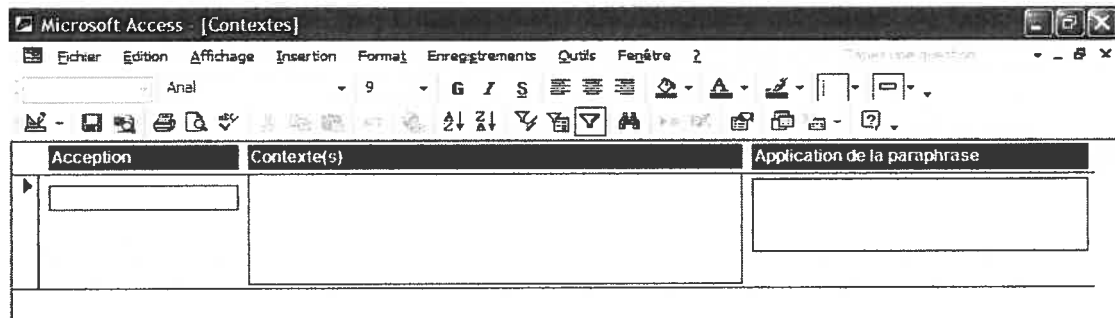


Figure 3.7 Formulaire *Contextes*

Acception : L'adjectif qui correspond à l'acception de la fiche du formulaire *types d'adjectifs*.

Contexte(s) : Le ou les contextes qui permettent d'illustrer l'acception.

Application de la paraphrase : La substitution de l'adjectif par la paraphrase dans un extrait du contexte adjacent dans le but de vérifier le jumelage du contexte avec l'acception (voir la section 3.5).

La base de données telle qu'elle est décrite ici, contient les champs nécessaires à l'analyse des adjectifs. Toutefois, il ne s'agit pas du modèle d'encodage que nous proposons plus loin. Premièrement, la base de données d'analyse contient des champs, comme la fréquence et l'indice de spécificité, qui ne sont pas nécessaires dans la désambiguïsation des adjectifs telle que décrite dans les objectifs du projet. Deuxièmement, nous avons joint la base de données d'analyse et la sortie Access de l'analyseur syntaxique SYNTAX qui nous sert à consulter les contextes dans lesquels sont observés les adjectifs (voir la section 3.6). Troisièmement, quelques champs ne servant pas à l'analyse, comme la définition de la base nominale et l'équivalent anglais, seront ajoutés au modèle d'encodage (voir la section 4.2). Nous aurions pu utiliser une seule base de données, mais celle-ci aurait été inutilement lourde et le

menu des formulaires aurait été particulièrement complexe pour un utilisateur qui n'est pas familier avec l'interface du logiciel Access.

3.4 Classement et sélection des candidats adjectifs

Dans le but d'identifier les adjectifs dérivés de noms qui feront l'objet de notre étude, nous avons procédé à une classification des adjectifs. Dans cette section nous effectuons un retour sur les critères de classification selon les types d'adjectifs tels qu'ils ont été présentés à la section 1.2. Nous présentons l'application des critères aux différents types d'adjectifs de façon à produire un patron paraphrastique lorsqu'il est possible de le faire. Par la suite, nous procédons à la présentation de la liste finale des adjectifs qui subiront une analyse fine de leur sens.

Au moment de procéder à la classification des adjectifs, nous ne faisons pas appel à l'ensemble des critères d'identification des termes évoqués à la section 2.1.2, car : 1) les adjectifs sont retenus peu importe leur type; et 2) la plupart des adjectifs apparaissant au début de la liste de candidats produite par TermoStat sont médicaux, donc il n'est pas nécessaire d'en confirmer l'appartenance à un domaine de spécialisation.

3.4.1 Classification des candidats adjectifs

Comme nous l'avons présenté à la section 1.2, nous avons choisi d'adopter une typologie essentiellement lexico-sémantique. Parmi les types d'adjectifs, nous comptons les qualificatifs, les déterminatifs, les dérivés de verbes et les dérivés de noms (auxquels nous greffons les adjectifs ayant un lien ténu avec un nom). Les adjectifs sont classés en fonction de leur sens et non de leur forme. En d'autres mots, la base et le type d'adjectif peuvent varier d'après le nom de tête auquel il est combiné. Reprenons l'exemple de l'adjectif *alimentaire*. Nous avons vu qu'il possède deux bases nominales et une base verbale. Il apparaît donc dans trois fiches (voir la section 3.4.2), dont deux classées dans les dérivés de noms et une classée dans les dérivés de verbes. Même si l'identification des bases se fait généralement de façon intuitive, car celle-ci est souvent formellement proche de l'adjectif, nous avons élaboré un énoncé

auquel doit s'appliquer l'adjectif pour confirmer son appartenance à une classe. La méthode d'identification des bases nominales est présentée à la section 3.5.2.1.1 pour la méthode d'identification.

a) Adjectifs qualificatifs

Comme nous l'avons vu à la section 1.2.3, les trois critères que nous évaluerons au moment d'analyser et de classer les adjectifs sont la fonction d'attribut et d'épithète, la capacité de coordination de l'adjectif qualificatif avec un autre adjectif qualificatif et sa capacité à accepter un adverbe de degré ou un complément. Toutefois, malgré les exemples donnés précédemment, ces critères souffrent de nombreuses exceptions et ne sont pas suffisants s'ils sont pris séparément. Cela dit, Maniez (2005) a observé le même phénomène avec les adjectifs anglais. Il a toutefois constaté que si deux critères, parmi une liste qui comprend nos trois critères, sont appliqués, la classification des adjectifs qualificatifs est juste dans 94 % des cas. Nous nous sommes donc inspirée de cette remarque pour la classification des adjectifs.

Au moment d'analyser l'adjectif, ce dernier doit répondre à deux des trois éléments de l'énoncé suivant pour être classé à titre d'adjectif qualificatif : 1) adjectif attribut/épithète; 2) qui peut se coordonner avec un adjectif qualificatif; ou 3) être accompagné d'un complément. Les figures 3.8 et 3.9 présentent un exemple des fiches *Types d'adjectifs* et *Contextes* qui est tiré de notre base de données.

The screenshot shows a Microsoft Access form titled 'Types d'adjectifs'. The form is divided into several sections:

- ADJECTIF:** absent 1
- FRÉQUENCE:** 16
- type:** Qualificatif
- INDICE:** 4.49116458624437
- COMMENTAIRE:** Adjectif attribut/épithète qui peut se coordonner avec un adjectif qualificatif ou être accompagné d'un complément
- BASE:** aucune base
- Table with 2 columns: Acception and Paraphrase**
 - Row 1: Acception: absent 1a; Paraphrase: fièvre, polynucléaire, œdème, protéines, notion, manifestation, raideur, capsule; Contexte(s): absent 1
 - Row 2: Acception: *; Paraphrase: ; Contexte(s): absent 1

The status bar at the bottom indicates 'Enr: 14 | 1 | sur 1'.

Figure 3.8 Exemple d'une fiche *Types d'adjectifs* pour un adjectif qualificatif

The screenshot shows a Microsoft Access form titled 'Contextes'. The form is divided into three columns: Acception, Contexte(s), and Application de la paraphrase.

Acception	Contexte(s)	Application de la paraphrase
absent 1a	En fait, les premiers sont en règle prédominants sur les seconds qui peuvent faire défaut ; ainsi la fièvre est absente dans 50 % des cas environ, tandis qu'une polynucléose n'est retrouvée que dans 60 % des cas.	être + absente + dans 50 % des cas (complément)
absent 1a	L'œdème périlésionnel est absent ou discret.	être + absent + ou (coordination) + discret (adj qual)

Two callout boxes point to the 'Application de la paraphrase' column:

- Élément 1** points to the first row's application: "être + absente + dans 50 % des cas (complément)".
- Élément 2** points to the second row's application: "être + absent + ou (coordination) + discret (adj qual)".

Figure 3.9 Exemple d'une fiche *Contextes* pour un adjectif qualificatif

b) Adjectifs déterminatifs

Comme nous l'avons mentionné précédemment (voir la section 1.2.1.1.2), ce groupe d'adjectifs joue souvent le rôle de déterminant et il possède un statut essentiel dans la structure de la phrase. De plus, la fonction grammaticale de ces adjectifs fait en sorte que les outils de traitement automatique des corpus vont plutôt les étiqueter

comme étant des déterminants, ce qui signifie qu'ils ne seront pas extraits au même titre que les autres adjectifs. Ce phénomène ne pose toutefois pas un problème, car les adjectifs déterminatifs ne possèdent pas de valeur sémantique suffisante pour justifier une analyse approfondie dans le cadre de cette étude. Cela dit, parmi les adjectifs déterminatifs, seuls les adjectifs numériques et ordinaux feront l'objet d'une entrée dans notre base de données. Ils ne seront toutefois que classés, sans être analysés.

c) Adjectifs dérivés de verbes

Pour identifier les adjectifs dérivés de verbes nous nous sommes attardée à la ressemblance morphologique (ex. *adhérent* – *adhérer*) ainsi qu'au sens de la base et au sens général que ces adjectifs partagent avec le verbe dont ils sont dérivés (*alimentaire* – *s'alimenter*) (voir la section 1.2.1.1.3).

Au moment de la classification, nous avons utilisé l'énoncé suivant pour valider la relation qui existe entre le *NomT* du syntagme nominal et le verbe de base (*VerbeB*) : *NomT* + *relation avec VerbeB*. Les figures 3.10 et 3.11 présentent un exemple de fiches *Types d'adjectifs* et *Contextes* tirés de notre base de données.

The screenshot shows a Microsoft Access form titled 'Types d'adjectifs'. The form is divided into several sections:

- Header:** 'Adjectif' (adhérent 1), 'Fréquence' (6), 'type' (Dérivé de verbe), 'Index' (3.20651818025784), and 'Commentaires' (NomT + relation avec VerbeB).
- Base:** 'adhérer'.
- Table:** A table with two columns: 'Acception' and 'Paraphrase'.

Acception	Paraphrase
adhérent 1b	NomT d'adhérer
Tete(s) de syntagme	pouvoir
adhérent 1a	NomT qui adhère
Tete(s) de syntagme	cellule, enduit, membrane

Figure 3.10 Exemple d'une fiche *Types d'adjectifs* pour un adjectif dérivé de verbe

The screenshot shows a Microsoft Access form titled 'Contextes'. The form is divided into three columns:

Acception	Contexte(s)	Application de la paraphrase
adhérent 1a	Un enduit fibrinoleucocytaire est parfois adhérent au niveau de la séreuse et atteste alors d'une péritonite associée.	enduit + qui adhère
adhérent 1a	La colite pseudo-membraneuse se caractérise par une infiltration inflammatoire intense de la muqueuse du côlon et du rectum ainsi que par la présence de fausses membranes fibrineuses adhérentes à la muqueuse.	membrane + qui adhère

Figure 3.11 Exemple d'une fiche *Contextes* pour un adjectif dérivé de verbe

d) Adjectifs dérivés de noms

Pour identifier les adjectifs dérivés de noms nous nous sommes attardée à la ressemblance morphologique (ex. *viral* - *virus*) ainsi qu'au sens de la base (ex. *hépatique*, base *hépat*- grec pour *foie* (Garnier, Delamare et al. 2002 : s.v.)) et au sens général que ces adjectifs partagent avec le nom dont ils sont dérivés (*alimentaire* - *aliment* et *alimentaire*) (voir la section 1.2.6).

Au moment de la classification, nous avons utilisé l'énoncé suivant pour valider la relation qui existe entre le nom de tête (NomT) du syntagme nominal et le nom de base (NomB) : *NomT + relation avec NomB*. Les figures 3.12 et 3.13 présentent un exemple de fiches *Types d'adjectifs* et *Contextes* tirés de notre base de données.

Microsoft Access - [Types d'adjectifs]

Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre 2

Arial 9 G I S

Adjectif	Fréquence	420	typ>	Dérivé de nom
viral 1	Indice	197.113224082514	Commentaires	
Base	NomT + relation avec NomB			
virus				

Acception	Paraphrase	
viral 1a	NomT qui est causé par un virus	viral 1
Tête(s) de syntagme	infection, maladie (hépatite, méningite, etc.), fièvre (hémorragique), inflammation, affection, atteinte, lésion	Contexte(s)
viral 1b	NomT qui est une partie d'un virus	viral 1
Tête(s) de syntagme	protéine, génome, gène, peptide, acide nucléique, ARN, glycoprotéine, molécule, élément, enveloppe, déterminant	Contexte(s)
viral 1c	NomT qui contient un virus	viral 1
Tête(s) de syntagme	suspension, charge	Contexte(s)
viral 1d	NomT qui est constitué d'un virus	viral 1
Tête(s) de syntagme	particule, matériel, agent	Contexte(s)
viral 1e	NomT qui provient d'un virus	viral 1
Tête(s) de syntagme	antigène	Contexte(s)
viral 1f	NomT qui vise un virus	viral 1

Enr : 14 1638 sur 1651

Figure 3.12 Exemple d'une fiche *Types d'adjectifs* pour un adjectif dérivé de nom

Acception	Contexte(s)	Application de la paraphrase
abdominal 1b	Le purpura rhumatoïde est une vascularite systémique de l'enfant ou de l'adulte caractérisée par une atteinte cutanée (purpura), digestive (douleurs abdominales), articulaire (arthralgies) et rénale (glomérulonéphrite à dépôts d'IgA).	douleur + qui se manifeste dans l'abdomen
abdominal 1b	* Clinique : Le début est très souvent insidieux pouvant se résumer à de vagues douleurs rachidiennes au long cours ou à des algies abdominales évoluant sur des semaines, des mois, voire des années.	algie + qui se manifeste dans l'abdomen

Figure 3.13 Exemple d'une fiche *Contextes* pour un adjectif dérivé de nom

Comme nous l'avons souligné à la section 1.2.1.2, le corpus contient un petit nombre d'adjectifs ayant un lien sémantique ténu avec un nom. Lorsque l'adjectif possède une forme identique à celle d'un nom, mais que son sens est connexe à celui-ci, nous devons être en mesure de remplacer l'adjectif en contexte par une paraphrase qui expliquera la relation qui unit le NomT du syntagme et le concept auquel renvoie le NomB, sans utiliser ce dernier. Donc, en reprenant l'exemple de l'adjectif *adulte* qui possède la même forme que le nom *adulte*, un *animal adulte* pourra être paraphrasé par *un animal qui a atteint la maturité*.

Au moment de la classification, nous les avons d'abord différencié des adjectifs dérivés de noms et nous avons utilisé la structure suivante pour appuyer le classement des adjectifs ayant un lien sémantique ténu avec un nom : *NomT + prop. subordonnée explicative (explication sans contenir le NomB)*. Une telle structure nous offre une flexibilité qui est parfois nécessaire lorsque le sens des différentes acceptions touche à des aspects différents du NomB (voir l'exemple des fiches aux figures 3.14 et 3.15).

ADJECTIF	FRÉQUENCE	type
adulte 1	46	lien sémantique tenu
BASE	INDICE	COMMENTAIRE
adulte	17.9023701593354	NomT + prop. subordonnée explicative (finalité sans contenir le NomB)

Acception	Paraphrase
adulte 1a	NomT + qui a atteint la maturité
Tête(s) de syntagme	animal (toutes les espèces), âge, ver (espèces), insecte (espèce), système nerveux, specimen, état, tissu, cerveau
adulte 1b	NomT + qui est destiné à un être animé mature
Tête(s) de syntagme	dose

Figure 3.14 Exemple d'une fiche *Types d'adjectifs* pour un adjectif ayant un lien sémantique tenu avec un nom

Acception	Contexte(s)	Application de la paraphrase
adulte 1a	Ce qui peut expliquer que le matériel génétique de si nombreux virus soient communément retrouvés dans le système nerveux adulte sain.	système nerveux + qui a atteint la maturité
adulte 1a	Le diagnostic repose essentiellement sur l'aspect des vers adultes retrouvés au sein d'un segment digestif.	ver + qui a atteint la maturité

Figure 3.15 Exemple d'une fiche *Contextes* pour un adjectif ayant un lien sémantique tenu avec un nom

Après avoir rédigé les différents critères qui nous permettent de classer les adjectifs selon leur type, nous avons été en mesure d'isoler les adjectifs dérivés de noms qui font l'objet de notre étude.

3.4.2 Numérotation des entrées

Lors de la classification et de l'analyse des adjectifs, nous avons numéroté les types d'adjectifs et les acceptions de façon à bien les distinguer. Nous avons utilisé trois niveaux de numérotation soit le type (chiffre romain), la base sémantique (chiffre arabe) et l'acception (lettre) de l'adjectif. Le type d'adjectif est généralement unique : nous avons donc choisi d'inscrire le chiffre romain seulement si l'adjectif possédait plus d'un type pour éviter d'alourdir la numérotation. Par contre, nous avons systématiquement numéroté les adjectifs selon la base, car une fiche est créée pour chaque base sémantique et nous avons tenu à ce que toutes les fiches soient identifiées au moyen d'une forme canonique numérotée. Enfin, chaque acception est numérotée au moyen d'une lettre minuscule. La figure 3.16 illustre la numérotation utilisée pour l'adjectif *alimentaire*.

- Type d'adjectif : chiffres romains (I)
- Base sémantique : chiffres arabes (1)
- Acception : lettre (a)

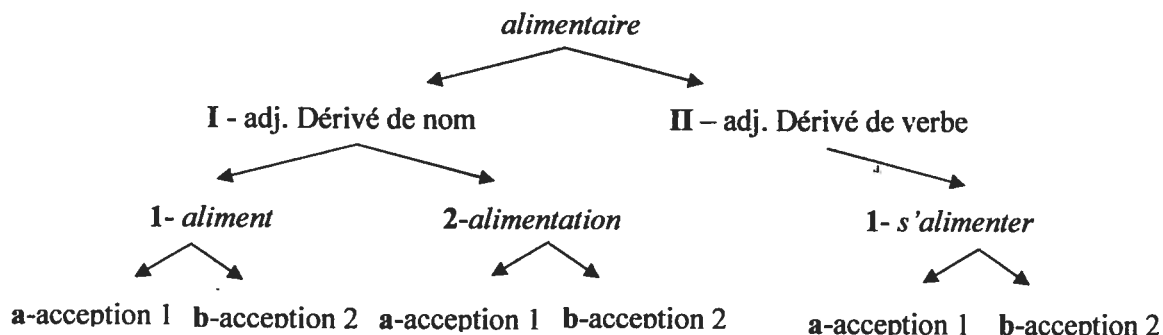


Figure 3.16 Exemple de la numération des acceptions de l'adjectif *alimentaire*

Dans la base de données qui sert à l'analyse, chaque base nominale possède une fiche sur laquelle sont regroupés les différentes acceptions de l'adjectif. Toujours d'après l'exemple *alimentaire*, l'adjectif fait donc l'objet de trois fiches, soit une pour

alimentaire I1, une pour *alimentaire I2* et une pour *alimentaire III*. Chaque acception est alors énumérée et numérotée sur la fiche. Par exemple, *alimentaire I2* possède deux acceptions donc la fiche possède deux sous-entrées : *alimentaire I2a* et *alimentaire I2b*.

3.4.3 Sélection des adjectifs analysés

À cette étape, nous avons repris la liste d'adjectifs obtenue après le nettoyage et nous l'avons triée en ordre décroissant selon l'indice de spécificité. Rappelons que la liste compte 1645 adjectifs. Puisque l'application des paraphrases qui vérifie l'appartenance à un type d'adjectif demande l'analyse du sens des adjectifs, nous avons choisi de mener de front la classification et l'analyse. Nous avons donc procédé à l'analyse sémantique de tous les adjectifs. Au lieu de fixer un nombre initial d'adjectifs dérivés de noms à analyser, nous avons préféré commencer l'analyse et évaluer le nombre d'adjectifs nécessaires à notre étude après l'analyse des 100 adjectifs les plus spécifiques, tous types confondus. En cours de route, nous avons constaté que les paraphrases qui expriment la relation unissant le nom de tête au nom de base de l'adjectif se répétaient d'un adjectif à un autre. Nous avons décidé d'arrêter l'analyse après avoir analysé 5 adjectifs dérivés de noms sans observer de nouvelle relation sémantique. Nous avons fixé le seuil à cette valeur, car nous ne pouvons savoir si une nouvelle relation sémantique sera observée tous les 10 ou les 100 adjectifs dérivés de noms. Nous avons également jugé que le nombre d'adjectifs analysés était suffisant compte tenu du nombre de syntagmes analysés (1300) dans le cadre du présent projet. De plus, les données amassées à ce point nous permettent de tirer des conclusions répondant à nos objectifs. Au total, nous avons analysé 86 adjectifs, dont 75 adjectifs dérivés de noms (voir l'annexe D pour la liste complète).

3.5 Analyse des adjectifs dérivés de noms

Comme nous l'avons vu à la section précédente, l'analyse des différentes acceptions des adjectifs de notre corpus a été faite au moment de leur classement. En

effet, pour être en mesure de classer les adjectifs, il faut d'abord identifier la relation qui existe entre le nom de tête du syntagme dans lequel il se trouve et l'adjectif ou sa base, nominale. Cet exercice constitue le fondement de l'analyse fine des adjectifs. Pour y arriver, nous avons dressé la liste des noms de tête pour chaque adjectif, puis nous avons mis en évidence la relation sémantique. La vérification de la typologie et de la relation sémantique a été possible en appliquant au contexte la paraphrase définitoire qui exprime la relation sémantique (voir la section 3.5.2.2). L'exercice consiste à remplacer l'adjectif dérivé de nom par la paraphrase définitoire qui correspond à son acception. Par exemple, l'adjectif *viral* *1a* (voir les figures 3.12 et 3.13) est défini par la paraphrase *NomT qui est causé par un virus*. En appliquant la paraphrase au syntagme *hépatite virale*, nous obtenons *hépatite qui est causée par un virus*. Enfin, l'application des paraphrases a permis, non seulement, de vérifier la paraphrase, mais également de regrouper les noms de tête selon la relation sémantique qu'ils appellent. Une telle analyse est possible au moyen d'une mise en contexte de l'adjectif.

À la section suivante, nous présentons l'analyseur syntaxique SYNTEX, dont les sorties permettent de consulter le corpus selon différents critères basés sur les relations syntaxiques. Par la suite, nous verrons la méthode qui nous a permis de distinguer les acceptions des adjectifs analysés.

3.5.1 Étude des contextes au moyen de SYNTEX

SYNTEX est un analyseur qui a été mis au point par Bourigault (2004) et son équipe. Un premier module d'étiquetage, qui utilise TreeTagger¹² (voir la section 3.2.1), appose une partie du discours aux mots à partir du corpus en format texte (.txt). Par la suite, un second module établit une liste de « dépendances syntaxiques » (Bourigault, Fabre et al. 2004) dans les phrases du corpus, mais aussi au sein des

¹² Bourigault utilise une version de TreeTagger qui est plus complète, soit plus avancée que la version utilisée dans TermoStat. Cette différence s'explique par l'utilisation de la version distribuée gratuitement dans TermoStat.

différents syntagmes qui constituent ces phrases. Les résultats de l'analyse sont présentés dans une base de données Access.

Dans un contexte terminologique, le logiciel nous permet d'extraire les syntagmes qui contiennent un lemme donné. Il est possible d'effectuer des recherches selon différents groupes de critères (Figure 3.17).

Figure 3.17 Formulaire de recherche dans les sorties Access de SYNTEX'

a) La catégorie

Une recherche d'après la *catégorie* permet de trouver une unité lexicale en sélectionnant la *catégorie* (partie du discours (nom, adjectif, etc) ou son appartenance à un syntagme (nominal, verbal)), la fréquence, ou une chaîne de caractères. Nous pouvons utiliser un ou plusieurs critères de recherche. Dans le cadre du projet, nous effectuons les recherches à partir de chaînes de caractères sans préciser la catégorie,

car nous voulons voir toutes les occurrences de l'adjectif *et*, par le fait même, contourner les erreurs d'étiquetage (voir la section 3.2.2).

b) La structure

D'après le fonctionnement de l'analyseur syntaxique, la structure renvoie à la construction des syntagmes. En effet, le syntagme comprend une tête, auquel renvoie le type de syntagme (ex. la tête d'un syntagme nominal est un nom) et des expansions. Par exemple, dans le syntagme nominal *infection virale*, la tête est *infection* et l'expansion, *virale*. Dans le cadre de notre projet, nous pouvons effectuer une recherche pour obtenir la liste de tous les syntagmes dont la tête est un nom et l'expansion est un adjectif. Aussi, nous pouvons rechercher tous les syntagmes dont l'expansion est, par exemple, *virale*. Nous obtenons alors une liste de 165 syntagmes avec la fréquence de chacun en corpus.

c) La phrase

Il est possible de chercher une phrase pour en consulter le contexte élargi. En d'autres mots, lorsque nous souhaitons voir les 3 phrases, ou segments, qui sont situés avant et après la phrase recherchée. Nous pouvons également effectuer une recherche à partir du numéro que le logiciel attribue aux segments (phrases ou fragments de phrase qui sont numérotés) du corpus. Toutefois, dans les deux cas, les recherches effectuées à l'aide de ce module échouent lorsque la chaîne de caractères ou le numéro ne correspond pas exactement au segment contenu dans le corpus.

Une recherche produit une liste de résultats qui nous informe de la partie du discours de l'unité lexicale recherchée. Par exemple, effectuer une recherche de la chaîne de caractères *virale* sans contrainte catégorielle émet une liste d'occurrences (figure 3.18) regroupées par catégorie (voir le point *a* à la page précédente). La liste nous informe, entre autres, que *virale* est un adjectif. Dans le cadre de notre étude, nous consultons les occurrences de *virale* dans lesquelles l'adjectif est placé en expansion. En sélectionnant ce groupe d'occurrences (double clic sur la valeur en expansion (E)

de la colonne *productivité*), nous obtenons une liste de syntagmes où l'adjectif *viral* est en expansion (figure 3.19).

The screenshot shows a Microsoft Access window with a table named 'Liste'. The table has the following columns: productivité (T, E), nb voisins (T, E), nb var, nbdoc, freq, fsc1, fsc2, cat, terme, and validité. The data is as follows:

productivité	T	E	nb voisins	T	E	nb var	nbdoc	freq	fsc1	fsc2	cat	terme	validité
7	165		2			1	400	0.5	0	0	Adj	viral	@cccccc
0	0					1	1	0	0	0	Nom	virales	ccccccc

Figure 3.18 Liste de résultats de la recherche de l'adjectif *viral*

The screenshot shows a Microsoft Access window with a table named 'descendants Expansion'. The table has the following columns: productivité (T, E), nb voisins (T, E), nb var, nbdoc, freq, fsc1, fsc2, cat, terme, and validité. The data is as follows:

productivité	T	E	nb voisins	T	E	nb var	nbdoc	freq	fsc1	fsc2	cat	terme	validité
24	54		1			1	98	0	0	0	SNom	infections virales	@cccccc
9	24					1	280	2	0	0	SNom	particules virales	ccccccc
5	8					1	24	10	0	0	SNom	hépatite virale	ccccccc
10	6					1	20	0	0	0	SNom	protéines virales	ccccccc
7	10					1	43	0	0	0	SNom	antigènes viraux	ccccccc
1	12					1	44	0	0	0	SNom	génomme virale	ccccccc
2	8					1	64	0	0	0	SNom	réplication virale	ccccccc
1	9					1	42	0	0	0	SNom	persistence virale	ccccccc
3	4					1	9	0	0	0	SNom	méningites virales	ccccccc
5	3					1	7	0	0	0	SNom	souches virales	ccccccc
2	4					1	8	0	0	0	SNom	gènes viraux	ccccccc
3	2					1	8	0	0	0	SNom	maladie virale	ccccccc
0	7					1	8	0	0	0	SNom	immunologie virale	ccccccc
0	6					1	6	0	0	0	SNom	infections virales persistantes	ccccccc
0	8					1	6	0	0	0	SNom	population virale	ccccccc
0	3					1	6	0	0	0	SNom	fièvre hémorragique virale	ccccccc
0	3					1	6	0	0	0	SNom	variants viraux	ccccccc
0	2					1	6	0	0	0	SNom	hépatite virale aiguë	ccccccc
3	1					1	6	0	0	0	SNom	peptides viraux	ccccccc
0	4					1	6	0	0	0	SNom	inflammation virale	ccccccc
4	0					1	4	0	0	0	SNom	épitopes viraux	ccccccc
1	3					1	4	0	0	0	SNom	origine virale	ccccccc

Figure 3.19 Liste de résultats de *viral* comme adjectif en expansion de syntagme.

Cette liste nous permet de voir les syntagmes nominaux qui sont présents dans notre corpus et qui contiennent l'adjectif analysé. À partir de cette première liste, nous pouvons dresser une seconde liste énumérant les noms de tête qui peuvent être combinés à l'adjectif analysé. À partir de cette liste, nous pouvons également consulter les contextes dans lesquels est utilisé l'adjectif lorsqu'il est combiné avec l'un ou l'autre des noms de tête (double clic sur la fréquence du syntagme choisi). La consultation du contexte (segment qui contient l'adjectif ainsi que les trois segments qui précèdent et qui suivent ce dernier (figure 3.20)) nous permet de dégager la relation sémantique qui unit le nom de tête à la base nominale de l'adjectif et d'appliquer la paraphrase (voir section 3.4.4).

The screenshot shows the Microsoft Access interface with a table of occurrences. The table has two columns: an ID and a description. Entry 7395 is selected, and its context is displayed in a separate window. The context window shows the text: 'Les infections virales surviennent surtout lorsque les réactions de défense de l'organisme sont insuffisantes (enfants, sujets âgés, immunodéprimés par cancer ou chimiothérapie...)'.

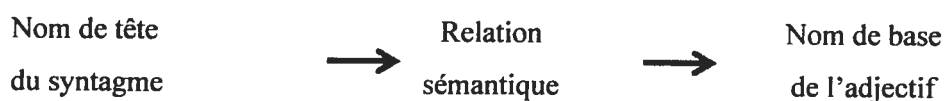
ID	Description
7382	Les virus sont des microorganismes qui se reproduisent obligatoirement dans les cellules dont ils sont les " parasites obligatoires " et dont ils vont utiliser les mécanismes de synthèse pour se répliquer et former de
7383	Les modes de pénétration du virus dans l'organisme sont nombreux : voie digestive (poliomyélite), respiratoire (grippe), épidermique (herpès), etc...
7384	Le placenta peut être franchi et le fœtus contaminé dans la rubéole.
7395	Les infections virales surviennent surtout lorsque les réactions de défense de l'organisme sont insuffisantes (enfants, sujets âgés, immunodéprimés par cancer ou chimiothérapie...).
7396	L'infection comporte trois périodes :
7397	- Période de latence, parfois très longue, seulement reconnue par la présence d'anticorps dans le sang.
7398	- Période de stimulation pendant laquelle l'effet cytopathogène est visible histologiquement.

Figure 3.20 Contexte qui contient le syntagme *infection virale*

Nous avons répété cette recherche pour les 75 adjectifs dérivés de noms et les 1300 syntagmes dont ils font partie de façon à mettre en évidence toutes les acceptions de chaque adjectif.

3.5.2 Méthode de distinction des acceptions

Dans la présentation de la base de données, nous avons adopté une numérotation qui permet d'identifier clairement les différentes acceptions. Dans le présent travail, les acceptions sont définies en fonction de la relation sémantique qui unit le nom de base de l'adjectif et le nom de tête dans le syntagme formé selon le patron *nom-adjectif*. En nous appuyant sur cette explication, nous avons choisi d'expliquer les acceptions des adjectifs au moyen de paraphrases définitoires qui sont rédigées à partir de trois éléments : le nom de base, le nom de tête et la relation sémantique.



Dans cette section, nous présentons d'abord les types de noms sur lesquels reposent les paraphrases définitoires, soit les noms de base et de tête. Dans un deuxième temps, nous utilisons ces concepts pour expliquer les patrons que nous avons choisis dans l'élaboration des paraphrases. Enfin, nous expliquons les principes qui ont dirigé l'élaboration du système de paraphrases définitoires issu de l'analyse des adjectifs.

3.5.2.1 Types de noms

Pour désambiguïser les adjectifs qui sont dérivés de noms, il est essentiel de bien saisir ce que nous entendons par *nom de base* de l'adjectif et la manière avec laquelle nous les identifions. Ensuite, il est tout aussi important de préciser la notion de *nom de tête*, car l'analyse des adjectifs repose sur l'ajout de sens qu'apporte la relation qui existe entre le nom placé en tête du syntagme nominal dans les combinaisons *nom-adjectif* et le nom de base.

3.5.2.1.1 Nom de base (NomB)

Le nom de base est le nom qui constitue l'élément de sens principal de l'adjectif et qui est au cœur de la paraphrase définitoire. La plupart du temps, celui-ci partage des caractéristiques formelles avec l'adjectif. Par exemple, le nom de base de l'adjectif *viral* est *virus*. Toutefois, puisque notre étude est sémantique, nous rencontrons certains cas particuliers où le nom de base varie du nom qui possède la forme graphique la plus près de l'adjectif. Si nous prenons l'exemple de l'adjectif *clinique*, le nom dont la forme est la plus près est *clinique* (l'établissement) toutefois, en consultant les contextes dans lesquels est utilisé l'adjectif, il est clair qu'il ne renvoie pas à l'établissement. La lecture des segments qui précèdent et suivent le segment où se trouve le syntagme analysé et la consultation de certaines bases de données (ex. *GDT* (OQLF 2006) et *Termium Plus* (Bureau de la traduction 2006)) nous permettent de comprendre que l'élément de sens qui émane de la description des syntagmes est le malade. Dans un tel cas, il est question d'un adjectif ayant un lien ténu avec un nom et le nom de base est alors *malade*. Par exemple, un *examen clinique* n'est pas un *examen qui est effectué en clinique*, mais bien un *examen qui est effectué sur un malade*¹³.

Un deuxième cas particulier est la mise en évidence de deux ou plusieurs bases nominales pour un même adjectif. Par exemple, en observant l'adjectif *vaccinal* nous pouvons affirmer que l'adjectif est dérivé du nom *vaccin* (produit). Ainsi, dans le syntagme *souche vaccinale*, le nom de base est *vaccin*, car l'adjectif peut être remplacé par la paraphrase *qui est contenue dans le vaccin*. Toutefois, l'analyse des contextes consultés au moyen de SYNTAX nous a permis de constater que l'adjectif renvoie parfois au nom *vaccination* (action). Nous convenons que le nom *vaccination* se définit pas « administration d'un vaccin » (Manuila, Manuila et al. 2001) et que, par le fait même, il renvoie au vaccin, la structure paraphrastique utilisée pour décrire nos

¹³ À l'entrée *examen clinique*, le *Dictionnaire des termes de médecine* de Garnier et Delamare (2002) donne la définition suivante : « Premier temps de l'examen médical : il comprend l'interrogatoire, la recherche de signes généraux et l'examen physique. » De plus, à l'entrée de l'adjectif *clinique*, il précise que « l'examen clinique peut être effectué ou constaté par le médecin au lit du malade sans le secours d'appareils ou de méthode de laboratoire » (2002 : 169). À la lecture de ces définitions et en gardant en mémoire le patron paraphrastique de nos définitions (*NomT* + *relation sémantique unique* + *NomB*) l'élément de sens qui est commun au deux énoncés est le malade et non l'établissement.

acceptions rejette l'utilisation de la double paraphrase (construction qui utilise deux compléments prépositionnels ou deux propositions relatives). Cela dit, dans le syntagme *prévention vaccinale*, l'adjectif pourrait être remplacé par la paraphrase *au moyen de l'administration d'un vaccin*. Mais la relation exprimée par la paraphrase contient deux compléments, soit *au moyen de l'administration* et *l'administration d'un vaccin*. Pour pallier ce problème, nous nous arrêtons au premier terme sémantique, c'est-à-dire au nom *vaccination*. Une telle substitution permet d'obtenir une paraphrase définitoire claire et concise : *au moyen de la vaccination*.

Le cas de *tuberculeux* se démarque par le fait qu'il renvoie à une maladie (*tuberculose*) et à une structure anatomique (*tubercule*). Comme dans les cas mentionnés précédemment, il s'agit ici de trouver la base nominale qui permet d'éviter l'utilisation d'une double paraphrase. Dans la plupart des contextes, l'adjectif renvoie à la tuberculose. Toutefois, dans le contexte « *Aspergillus* s'apparente à un saprophyte lorsqu'il colonise une cavité naturelle (sinus de la face) ou néoformée pulmonaire (**séquelleuse tuberculeuse**, bulle d'emphysème ou cancer nécrosé) et détermine alors un aspergillome. », *tuberculeux* renvoie un « nodule arrondi et saillant situé à la surface d'un organe ou d'un os » qui se nomme *tubercule* (Manuila, Manuila et al 2001 : s.v. 1). Cela dit, le raisonnement expliqué plus tôt justifie l'identification de *tubercule* comme base sémantique possible pour l'adjectif *tuberculeux*.

En général, nous avons eu recours à ce type de raisonnement dans les cas de partie-tout, soit lorsque l'adjectif renvoie à un élément (ex. *lymphe*) et à ce qu'il contient (ex. *lymphocyte*) ou une substance (ex. *bile*) et la structure anatomique où il se situe (ex. *voie biliaire*). De plus, son utilisation est également justifiée dans cas d'un produit (ex. *vaccin*) et de l'action d'utiliser ce produit (ex. *vaccination*).

Pour ce qui est des structures anatomiques dont la dénomination contient l'adjectif (ex. *voie biliaire*, *appareil respiratoire*, etc.) il est important de préciser que le description du syntagme décrivant celles-ci se trouve dans la base de données. Par exemple, une *maladie biliaire* est décrite par *une maladie qui affecte la voie biliaire* et

la *voie biliaire* est décrit comme *une voie où circule la bile*. Pour éviter d'avoir une double paraphrase (ex. *une maladie qui affecte la voie où circule la bile*), nous avons choisi d'utiliser le premier terme sémantique comme base sémantique, soit la dénomination courante de la structure anatomique, soit *voie biliaire*.

3.5.2.1.2 Nom de tête (NomT)

Ce nom est placé en tête du syntagme *nom-adjectif* étudié et il est modifié par l'adjectif. Dans la présente étude, il constitue un élément essentiel, car les relations dont nous avons fait état à la section 1.2.1.2.2 varient selon la nature du NomT. Par exemple, l'adjectif *épithélial* est remplacé par la paraphrase *qui est une partie de l'épithélium* lorsqu'il est précédé du NomT *cellule* et par la paraphrase *qui est constituée de l'épithélium* lorsqu'il est combiné avec le NomT *tissu*. Ce rôle nous a incité à dresser la liste des noms de tête auxquels sont combinés les adjectifs dans le corpus de médecine que nous analysons et nous les avons regroupés selon la relation qu'ils appellent (voir le système de paraphrases à l'annexe C).

3.5.2.2 Utilisation de la paraphrase définitoire

Pour analyser un adjectif, notre travail consiste à identifier la ou les acceptions observées dans notre corpus. Cet exercice nous permet de mettre en évidence les variations de sens observées d'un contexte à un autre. Pour définir les acceptions des adjectifs, nous avons utilisé des paraphrases inspirées des patrons syntaxiques des *adjectifs de relation* présentés par Monceaux (1997 : 42-45). En effet, parmi les caractéristiques des adjectifs dérivés de noms que nous avons énumérées à la section 1.2.1.2, nous présentons leur capacité à être remplacés par une paraphrase illustrant une relation prépositionnelle ou une relation sémantique plus complexe (voir la section 1.2.1.2.2). Le tableau 3.2 présente les patrons syntaxiques qui ont servi à la construction des paraphrases définitoires de la présente base de données.

Tableau 3.2 Patrons syntaxiques des paraphrases définitives des adjectifs dérivés de noms

Type de relation	Patron syntaxique	Exemple de paraphrase
Relation simple	NomT + préposition + NomB	<i>NomT de bactéries</i> (ex. <i>culture bactérienne</i>) <i>NomT en bactéries</i> (ex. <i>concentration bactérienne</i>)
Relation complexe	NomT + proposition relative + NomB qui contient un verbe conjugué	<i>NomT qui est causé par une bactérie</i> (ex. <i>infection bactérienne</i>) <i>NomT qui contient une bactérie</i> (ex. <i>aérosol bactérien</i>) <i>NomT qui est une bactérie</i> (ex. <i>microorganisme bactérien</i>)

Note : NomT = Nom de tête du syntagme; NomB = Nom de base.

Une relation simple entre deux concepts, peut être observée lorsque l'adjectif précise la nature d'un nom ou que ce dernier caractérise la base nominale de l'adjectif (Monceaux 1997 : 56). Par exemple, dans le syntagme *culture cellulaire*, le concept de *culture* renvoie simplement à la *cellule* (base nominale de *cellulaire*). L'adjectif *cellulaire* ne fait que préciser la nature de la *culture*¹⁴. La relation entre les deux noms étant directe nous n'avons pas besoin de la préciser davantage. Nous observons le même type de relation dans le syntagme *fonction cellulaire*, où *fonction* est une caractéristique de la base nominale *cellule*. Dans les deux cas, nous pouvons alors remplacer l'adjectif par la paraphrase *de la cellule*¹⁵.

¹⁴ Le nom *culture* peut signifier (1) « l'action de cultiver » (ex. *faire une culture de cellules*) et (2) « le résultat de la culture 1 » (ex. *la culture confirme le diagnostic*). Au moment de l'analyse, si nous avons rencontré le sens 1, nous avons utilisé la paraphrase *NomT + de la cellule*, et si nous avons observé le sens 2, nous avons utilisé *NomT qui est constitué de cellules*.

¹⁵ Nous soulignons la possibilité d'une variation sémantique pour la préposition *de*. Pour pouvoir atteindre l'objectif de désambiguïsation des adjectifs, nous avons pris soin de distinguer les acceptions où la préposition *de* lie l'adjectif à 1) une action (ex. *torsion gastrique*); 2) une propriété (ex. *fonction cellulaire*), et 3) lorsqu'elle permet d'éviter la présence de paraphrases maladroites (ex. *carcasse animale* : « carcasse qui appartient à l'animal »). Les différents sens de la préposition *de* sont, respectivement, numérotés *de(1)*, *de(2)*, *de(3)* dans la base de données et dans le système de paraphrases.

Le modèle que nous avons adopté pour la description des acceptions repose sur la structure paraphrastique *NomT + relation sémantique + NomB*. Puisque la présence de préfixes force parfois la coordination de 2 noms de base (ex. *maladie cardiovasculaire* = maladie qui affecte le cœur (*cardio~*) et les vaisseaux (*vascul~*)) ou une modulation de la paraphrase (ex. *patient asymptomatique* = patient que ne présente pas (*a~*) de symptôme) (voir la section 3.5.2.2.1), la structure ne permet pas la présence d'une double paraphrase, soit le besoin d'utiliser deux compléments pour atteindre la base. Pour pallier le problème, nous avons choisi d'utiliser le premier terme qui nous permet l'utilisation d'une seule paraphrase (ex. *administration d'un vaccin* = *vaccination*) L'application de cette restriction donne lieu à la présence de bases multiples (l'exemple de l'adjectif *vaccinal* (base : *vaccin* et *vaccination*) est présenté à la section 3.5.2.1.1).

Nous avons observé que le nom de tête demande parfois l'utilisation d'une préposition différente à *de* pour exprimer une relation directe. Par exemple, dans le syntagme *interaction cellulaire*, l'adjectif précise la nature de l'*interaction*. Toutefois, une interaction est possible **entre** deux entités. Cela dit, au moment de rédiger les paraphrases définitoires, nous avons pris soin de respecter les prépositions régies. Nous nous sommes appuyée sur celles apparaissant dans les dictionnaires de langue (ex. *Le Petit Robert* (Rey et Rey-Debove 2004)) de façon à donner une information idiomatique à l'utilisateur de la base de données. C'est pourquoi la paraphrase définitoire utilisée pour une *interaction cellulaire* est *entre les cellules* et non *des cellules*.

Dans le cas d'une relation complexe, nous avons adopté un patron syntaxique souvent utilisé par les dictionnaires pour définir les adjectifs : pronom relatif (généralement *qui*) + construction verbale explicative (Schneidecker 2002 : 11). Nous nous démarquons toutefois des dictionnaires existants par la précision des verbes utilisés dans ce patron. Par exemple, dans le syntagme nominal *maladie hépatique*, nous expliquons *hépatique* (base nominale : *foie*) par *NomT qui affecte le foie*. Dans un dictionnaire plus général, la définition adoptera souvent la forme *qui + se rapporte*

à *NomB*. Dans un tel ouvrage, qui répond à des besoins généraux, l'adjectif *hépatique* serait donc défini par *qui se rapporte au foie*.

Dans le cadre de l'analyse des adjectifs, nous nous sommes fondée sur deux principes de base pour la rédaction de nos paraphrases définitives :

1- Le nom de base de l'adjectif : Le point de départ de nos paraphrases est le nom de base de l'adjectif qui est lié à un autre nom (NomT) au moyen d'une relation sémantique. Par exemple, toutes les paraphrases de l'adjectif *abdominal* seront construites à partir du nom *abdomen* (*qui se situe dans l'abdomen, qui se manifeste dans l'abdomen, qui affecte dans l'abdomen, etc.*).

2- Les collocatifs verbaux appelés par les noms placés en tête de syntagme pour formuler la relation sémantique : La recherche de combinaisons *nom-verbe* dans SYNTAX (voir section 3.5.1) nous a permis d'observer des collocatifs de la langue médicale. Par exemple, la recherche du nom *maladie* nous informe qu'elles *sont causées par x, affectent y ou provoquent z*.

Lorsque nous avons rédigé les paraphrases définitives, nous avons assuré une certaine uniformité dans l'expression des relations en répertoriant les paraphrases neutralisées dans un tableau (annexe C). Nous nous sommes non seulement assurée de respecter une certaine structure dans la formulation, mais le tableau nous a permis de bien voir les relations qui étaient appelées à la fois par les NomT et par les NomB. De cette façon, nous pouvions rapidement choisir la paraphrase à utiliser et la vérifier en remplaçant l'adjectif en contexte par la paraphrase neutralisée dans laquelle le NomB était remplacé par la base nominale de l'adjectif. Par exemple, pour analyser le syntagme *chirurgie thoracique*, nous avons vérifié dans le tableau si nous avions déjà analysé un syntagme contenant le NomT *chirurgie*. Nous avons alors observé que la combinaison du nom *chirurgie* avec un adjectif renvoyant à une structure anatomique, en l'occurrence *intestin* (*chirurgie intestinale*), appelait la paraphrase *qui est effectué sur NomB*. Nous avons donc testé la paraphrase définitive en contexte pour obtenir *chirurgie qui est effectuée sur le thorax*, ce qui a donné une expression naturelle et sémantiquement juste dans le contexte issu de notre corpus.

Un des objectifs de notre modèle d'encodage est de permettre la compréhension du contexte pour permettre une meilleure utilisation des adjectifs. Dans cette optique, ces principes prennent une certaine importance, car leur application permet de mettre en évidence l'acceptation de l'adjectif qui apparaît dans un contexte donné et l'utilisation du collocatif verbal approprié.

Le système de paraphrases regroupe une liste de formulations des relations sémantiques qui peuvent servir à expliquer le sens des adjectifs analysés au cours de notre étude. Toutefois, au moment de rédiger les paraphrases définitives, nous avons adapté la phraséologie en fonction des collocatifs verbaux appelés par les noms de tête rattachés aux acceptations répertoriées dans la base de données terminologiques. Il est donc intéressant de soulever la différence entre l'énoncé qui apparaît dans le système de paraphrases et la paraphrase qui utilise la bonne préposition (en gras dans l'exemple) dans un contexte donné. Par exemple, dans le système de paraphrases, l'énoncé qui apparaît est : *NomT qui se manifeste dans/sur/chez NomB*. Dans la base de données, nous trouvons les acceptations qui sont inscrites de la façon suivante :

- a) *Douleur abdominale* « NomT qui se manifeste **dans** l'abdomen »
- b) *Signe osseux* « NomT qui se manifeste **sur** l'os »
- c) *Forme clinique* « NomT qui se manifeste **chez** le malade »

3.5.2.2.1 Les adjectifs préfixés

Dans la liste des adjectifs analysés, nous observons, chez certains, la présence d'un préfixe (ex. *antituberculeux*). Dans bien des cas, le préfixe oriente le choix du verbe qui sera utilisé dans la paraphrase définitive. Par exemple, dans l'adjectif *antituberculeux*, *anti-* signifie « contre » (Garnier, Delamare et al. 2002 : xxiii) alors la paraphrase définitive est *NomT qui combat la tuberculose*. Ici, le verbe *combattre* est le cooccurrent utilisé en médecine pour exprimer « qui agit contre ».

Enfin, nous avons observé que les adjectifs peuvent également posséder deux préfixes. Dans une telle situation, nous avons procédé à l'addition du sens de chaque préfixe de façon à expliquer l'adjectif avec le plus de justesse. Par exemple, dans le cas de l'adjectif *antianaérobie* (voir la figure 4.5), nous observons la présence de l'adjectif *aérobie* (« NomT qui nécessite la présence d'air ») qui est précédé des suffixe *an-* (« absence ») et *anti-* (« contre »). Si nous procédons à la reconstruction de l'adjectif à partir de la racine, ici *aérobie*, nous obtenons :

- 1) *aérobie* : « NomT qui nécessite la présence d'air »
- 2) *anaérobie* : « NomT qui nécessite l'absence d'air »
- 3) *antianaérobie* : « NomT qui combat ce qui nécessite l'absence d'air »

Dans le cadre d'études ultérieures, il sera important d'approfondir le rôle des affixes dans la construction des adjectifs dérivés, de façon à observer le lien qui existe entre l'ajout sémantique qu'ils apportent et la place de celui-ci dans la rédaction de la paraphrase définitoire (voir la section 2.3.1).

La démarche méthodologique que nous venons de présenter a permis de brosser un tableau illustrant la place qu'occupent nos adjectifs dans le corpus médical que nous avons analysé. Le chapitre qui suit présente les différentes données que nous avons recueillies et il présente le modèle d'encodage que nous avons élaboré.

Chapitre 4 : Données obtenues et modèle d'encodage

Dans ce chapitre, nous brossons d'abord un tableau de notre étude en décrivant les données obtenues à la suite de l'application de la méthodologie décrite au chapitre 3. Nous commençons par présenter la répartition des adjectifs que nous avons extraits de notre corpus et nous terminons en nous concentrant sur les caractéristiques des adjectifs dérivés de noms. Dans un deuxième temps, nous présentons le modèle d'encodage qui a servi à la construction de la base de données terminologiques destinée aux traducteurs que nous avons nommée *MédiTerm*. Nous y verrons les champs de la base de données d'analyse qui ont servi de point de départ à l'élaboration des fiches de *MédiTerm* et les nouveaux champs que nous jugeons essentiels pour assurer une utilisation efficace de la ressource que nous avons mise au point.

4.1 Données obtenues

Même si notre travail vise l'analyse du sens et l'encodage des résultats de cette analyse, nous avons été en mesure de recueillir un certain nombre de données nous informant sur la répartition des types d'adjectifs dans notre corpus et sur d'autres valeurs relatives aux acceptions observées, aux noms de tête et de base et au système de paraphrase qui nous a servi à uniformiser nos définitions.

4.1.1 Portrait détaillé des adjectifs dans le corpus d'analyse

Des études antérieures (L'Homme 2004a, voir la section 2.1.3) ont montré que les adjectifs extraits du corpus médical analysé par L'Homme (2004) étaient souvent dérivés et que ces derniers étaient majoritairement dérivés de nom. Les données que nous avons ne sont pas directement comparables, car nous avons analysé les adjectifs spécifiques à notre corpus, ce qui n'avait pas été le cas de cette dernière étude où les adjectifs ont été sélectionnés aléatoirement. Toutefois, nous voyons clairement que les adjectifs spécifiques de notre corpus médical appartiennent surtout au groupe des

adjectifs dérivés de noms (voir le tableau 4.1). Cette proportion confirme les résultats avancés par L'Homme (2004a) qui avait observé une tendance similaire dans son corpus médical.

Tableau 4.1 Répartition des adjectifs analysés

<u>Adjectifs analysés</u>	
total	86
Adjectifs qualificatifs	6
Adjectifs déterminatifs	1
Adjectifs dérivés de verbes	4
Adjectifs dérivés de noms	75

Dans le cadre de notre méthodologie, nous avons procédé à l'analyse des adjectifs les plus spécifiques de notre corpus, soit 86 adjectifs dont 6 qualificatifs, 1 déterminatif, 4 dérivés de verbes et 75 dérivés de noms. Le tableau 4.2 présente la liste des adjectifs dérivés de noms qui sont contenus dans la base de données. Cette liste est reprise en annexe D où le tableau présente également la fréquence, l'indice de spécificité et la ou les bases nominales.

Tableau 4.2 Liste des adjectifs dérivés de noms analysés

abdominal	ganglionnaire	parasitaire
anaérobie	grippal	parenchymateux
antigénique	hématogène	pathogène
antituberculeux	hémorragique	péricardique
antiviral	hépatique	pleural
asymptomatique	histologique	polymicrobien
aviaire	immun	protéolytique
bactérien	immunitaire	pulmonaire
bactériologique	infectieux	purulent
biliaire	inflammatoire	pyogène
botulique	interhumain	rachidien
cellulaire	interstitiel	rénal
clinique II	intestinal	respiratoire
cutané	intracellulaire	septicémique
diagnostique	intraveineux	septique
digestif	lombaire	sérologique
encéphalique	lymphatique	structural
endothélial	lymphocytaire	thoracique
éosinophile	lymphoïde	tissulaire
épidémique	méningé	toxinogène
épidural	microbiologique	tuberculeux
épithélial	muqueux	urinaire
épithélioïde	neurologique	vaccinal
étiologique	osseux	vasculaire
extrapulmonaire	pancréatique	Viral

En observant la liste, telle qu'elle est présentée en annexe D, nous voyons la fréquence de chaque adjectif, soit le nombre de contextes dans lesquels ils sont rencontrés. Nous avons donc consulté plus de 5500 contextes. Nous avons observé que les adjectifs les plus spécifiques à notre corpus sont également les plus fréquents. En effet, l'adjectif *viral* est à la fois le plus spécifique (indice de 197,1132) et le plus fréquent (420 occurrences). Un tel nombre d'occurrences par adjectif explique le nombre important de contextes consultés lors de l'analyse.

4.1.2 Bases nominales

Lors de l'analyse des 75 adjectifs dérivés de noms, nous avons rencontré 81 bases nominales différentes. Le tableau 4.3 présente la liste des bases nominales recensées lors de l'analyse des adjectifs.

Tableau 4.3 Liste des bases nominales

abdomen	hémorragie	protéolyse
air	histologie	pus
antigène	humain	rein
appareil digestif	immunisation	respiration
appareil respiratoire	immunité	sang
appareil urinaire	infection	sepsie
bactérie	inflammation	septicémie
bactériologie	interstice	sérologie
bile	intestin	structure
botulisme	lombe	symptôme
cellule	lymphe	système lymphatique
cellule épithéliale	lymphocyte	système nerveux
clinique	malade	système vasculaire
Clostridium botulinum	maladie	thorax
colonne vertébrale (rachis)	méninges	tissu
diagnostic	microbe	toxine
digestion	microbiologie	tubercule
dure-mère	mucus	tuberculose
encéphale	muqueuse	urine
endothélium	oiseau (principalement volaille)	vaccin
éosine	os	vaccination
épidémie	pancréas	vaisseau
épithélium	parasite	vaisseau lymphatique
étiologie	parenchyme	veine
foie	peau	virus
ganglion	péricarde	voie biliaire
grippe	poumon	

Parmi les adjectifs dérivés de noms, la plupart ne possèdent qu'une seule base nominale, mais 11 adjectifs en possèdent 2 ou plus (tableau 4.4). Par exemple, l'adjectif *digestif* possède deux bases nominales, soit *digestion* et *appareil digestif*. Les bases nominales appartiennent à deux classes sémantiques bien distinctes. En effet, l'adjectif renvoie à l'acte de digérer (*digestion*) et à la structure anatomique (*appareil digestif*). De la même façon, l'adjectif *biliaire* renvoie à la structure anatomique (*voie biliaire*) ainsi qu'à la substance qui y circule (*bile*) (voir la section 3.5.2.1.1).

Tableau 4.4 Bases nominales des adjectifs dérivés de noms

<u>Bases nominales par adjectif</u>	
Minimum	1
Maximum	3
<u>Adjectifs possédant 2 bases nominales et plus :</u>	
<ul style="list-style-type: none"> • <i>biliaire (bile/voie biliaire)</i> • <i>botulique (Clostridium botulinum/ botulisme)</i> • <i>digestif (digestion/appareil digestif)</i> • <i>hémorragique (hémorragie/sang)</i> • <i>lymphoïde (lymphe/système lymphatique/lymphocyte)</i> • <i>lymphatique (lymphe/système lymphatique)</i> 	<ul style="list-style-type: none"> • <i>muqueux (mucus/muqueuse)</i> • <i>respiratoire (respiration/appareil respiratoire)</i> • <i>tuberculeux (tuberculose/tubercule)</i> • <i>urinaire (urine/appareil urinaire)</i> • <i>vaccinal (vaccin/vaccination)</i> • <i>vasculaire (vaisseau/système vasculaire)</i>

En observant la liste en annexe, nous sommes en mesure de retrouver les bases nominales énumérées précédemment, mais nous pouvons également constater qu'une même base peut servir à plus d'un adjectif. En effet 9 noms constituent la base nominale de deux adjectifs. Par exemple, nous remarquons que le nom *lymphe* est identifié comme base pour les adjectifs *lymphatique* et *lymphoïde*. Le tableau 4.5 présente la liste des bases nominales qui ont produit deux adjectifs.

Tableau 4.5 Liste des bases nominales produisant deux adjectifs

Nom de base	Adjectifs dérivés
<i>cellule</i>	<i>cellulaire, intracellulaire</i>
<i>lymphe</i>	<i>lymphatique, lymphoïde</i>
<i>lymphocyte</i>	<i>lymphocytaire, lymphoïde</i>
<i>poumon</i>	<i>pulmonaire, extrapulmonaire</i>
<i>pus</i>	<i>purulent, pyogène</i>
<i>sang</i>	<i>hémorragique, homogène</i>
<i>système lymphatique</i>	<i>lymphatique, lymphoïde</i>
<i>tuberculose</i>	<i>tuberculeux, antituberculeux</i>
<i>virus</i>	<i>viral, antiviral</i>

4.1.3 Acceptions

Les 75 adjectifs dérivés de noms que nous avons analysés se partagent quelque 350 acceptions. Nous avons observé que le nombre d'acceptions varient entre 1 et 11

et que les adjectifs possèdent en moyenne 5,10 acceptions (tableau 4.6). Dans la liste présentée en annexe, nous donnons également le nombre d'acceptions par adjectif, mais les valeurs diffèrent légèrement, car dans l'annexe, les valeurs présentes dans le tableau ne reflète pas la présence de plus d'une base nominale. Ici, il s'agit du nombre d'acceptions par fiche, soit par adjectif pour une base nominale. Par exemple, en annexe, l'adjectif *digestif* possède un total 11 acceptions, mais, en réalité, *digestif 1* (base : *digestion*) possède 5 acceptions et *digestif 2* (base : *appareil digestif*), 6.

Tableau 4.6 Acceptions des adjectifs analysés

Nombre total d'acceptions	382
Nombre moyen d'acceptions par adjectif	5,10
Répartition des adjectifs selon le nombre d'acceptions	
Nombre d'acceptions	Nombre d'adjectifs
1	2
2	14
3	11
4	7
5	10
6	11
7	5
8	5
9	4
10	4
11	2

Parmi les 75 adjectifs seulement deux ont donné lieu à une seule acception, soit les adjectifs *protéolytique* (« NomT qui provoque la protéolyse ») et *pyogène* (« NomT qui provoque la production de pus »). Contrairement à ces adjectifs qui ne possèdent qu'une acception, certains autres en possèdent un plus grand nombre. Nous avons observé deux adjectifs qui comptent 11 acceptions : *digestif* et *urinaire*. Voici les 11 acceptions de l'adjectif *digestif* :

- 1a) NomT qui affecte la digestion (ex. *trouble digestif*)
- 1b) NomT où se produit la digestion (ex. *voie digestive*)
- 1c) NomT qui se manifeste lors de la digestion (ex. *signes digestifs*)
- 1d) NomT qui se produit au moment de la digestion (ex. *hypersécrétion digestive*)
- 1e) NomT qui participe à la digestion (ex. *enzyme digestive*)
- 2a) NomT qui est une partie de l'appareil digestif (ex. *muqueuse digestive*)
- 2b) NomT qui se situe dans l'appareil digestif (ex. *flore digestive*)
- 2c) NomT qui est effectué sur l'appareil digestif (ex. *biopsie digestive*)
- 2d) NomT qui se produit dans l'appareil digestif (ex. *tropisme digestif*)
- 2e) NomT de (1) l'appareil digestif (ex. *hyperinfestation digestive*)
- 2f) NomT de (2) l'appareil digestif (ex. *motricité digestive*)

4.1.4 Paraphrases

Comme le montre le tableau 4.7, malgré leur nombre important (74 paraphrases rédigées lors de l'analyse), les paraphrases définitives sont souvent répétées pour un même groupe d'adjectifs. Par exemple, la paraphrase simple *de(2)* explique une acception pour 40 adjectifs alors que la paraphrase complexe *NomT qui affecte NomB* est observée dans le cas de 33 adjectifs. Cette observation nous permet de croire que les paraphrases semblent être réutilisées lorsque les NomT sont apparentés sémantiquement.

Tableau 4.7 Paraphrases définitives produites lors de l'analyse

<u>Total de paraphrases</u>		74
Adjectifs par paraphrase	Minimum	1
	Maximum	40
	Moyenne	5,14

Le tableau 4.8 présente un échantillon du système de paraphrases (voir l'annexe C). La première colonne énumère les NomT qui ont été associés à la paraphrase, la deuxième colonne présente la paraphrase neutralisée, la troisième colonne énumère les adjectifs qui possèdent cette acception et la quatrième colonne donne les bases nominales associées aux adjectifs énumérés précédemment.

Tableau 4.8 Échantillon du système de paraphrases

Nom T	Paraphrase	Adjectif	Nom B
maladie (tuberculose, etc.), atteinte, infection, pathologie, forme, ascariose, agression, complication, affection, colique, cirrhose, infarctus, fibrose, insuffisance, syndrome, problème, protozoose, atrophie (villositaire), atonie, helminthiase, métraplasié, biharziose, pneumopathie (interstitielle), phénomène, phase, empyème, trouble, lésion, défaillance, sécheresse, réaction inflammatoire, néphrite, pneumopathie, œdème, vibrance, néoplasie, (neuro-)immunologie, immunité, vaccin, stratégie, virus, maladie, épidémie, viroses, agent infectieux (bactérie), épidémiologie [des maladies], microbiologie, contamination, bactériologie, problème, entité, pathologie, ligand, parasitisme, effecteur, phénomène, grippe, souche, botulisme, détresse, virus, syndrome, trouble, défaillance, difficulté, parasite	Nom T qui affecte Nom B	pulmonaire, abdominal, cellulaire, cutané, urinaire, pancréatique, encéphalique, hépatique, osseux, méningé, rénal, intestinal, lymphoïde, ganglionnaire, biliaire, vasculaire, pleural, neurologique, épithélial, parenchymateux, muqueux, rachidien, épithélioïde, interstitiel, épidual, extrapulmonaire, péricardique, viral, animal, clinique, aviaire, respiratoire, digestif	poumon, abdomen, cellule, peau, système urinaire, pancréas, encéphale, foie, os, méninges, rein, intestin, système lymphatique, ganglion, voie biliaire, système vasculaire, plèvre, système nerveux, épithélium, parenchyme, muqueuse, colonne vertébrale (rachis), cellule épithéliale, tissu interstitiel, (surface externe de la) dure-mère, péricarde, virus, animal, malade, oiseau, respiration, digestion
radiographie, prélèvement, examen, suivi, investigation, surveillance, échographie, poussée, biopsie, test, chirurgie, endoscopie, bilan, scintigraphie, exploration (isotopique), ponction, tomographie par émission de positrons (TEP), drainage, coupe, scanner, imagerie, drainage, injection, traitement, urographie, antibiothérapie, ensemencement, ponction, collection, sérologie, PCR (réaction en chaîne de la polymérase), examen	Nom T qui est effectué sur/dans Nom B	pulmonaire, abdominal, clinique, respiratoire, cutané, digestif, hépatique, osseux, méningé, rénal, intestinal, ganglionnaire, biliaire, pleural, neurologique, tissulaire, thoracique, intraveineux, parenchymateux, lombaire, épidual, extrapulmonaire, péricardique, viral, clinique	poumon, abdomen, malade, voies respiratoires, peau, tractus digestif, foie, os, méninges, rein, intestin, ganglion, voie biliaire, plèvre, système nerveux, tissu, thoracique, veineux, parenchyme, lombes, (surface externe de la) dure-mère, péricarde, virus, clinique
protéine, génome, gène, peptide, acide nucléique, ARN, glycoprotéine, molécule, élément, enveloppe, déterminant génique, composant, spore, structure membranaire, corps, paroi, membrane, constituant, élément, débris, protéine, récepteur, molécule, secteur, glycoprotéine, fraction, effecteur, gamète, souche, préparation, parenchyme, alvéole, muscle, cellule, muqueuses, épithélium, capillaire, capillaire, surface, segment, (sous-) muqueuse, lumière, glande, canal, acini, structure, moëlle, tissu, fenêtre, matrice, glomérule, artère, arbre artériel, pôle, lithiase, sel, pigment, paroi, cavité, canal, fragment	Nom T qui est une partie de Nom B	viral, bactérien, cellulaire, antigénique, lymphocytaire, parasitaire, vaccinal, pulmonaire, respiratoire, cutané, digestif, pancréatique, hépatique, osseux, rénal, intestinal, ganglionnaire, biliaire, vasculaire, tissulaire, épithélial, rachidien	virus, bactérie, cellule, antéigène, lymphocyte, parasite, vaccin, poumon, voies respiratoires, peau, tractus digestif, pancréas, foie, os, rein, intestin, ganglion, bile, vaisseau, système vasculaire, tissu, épithélium, colonne vertébrale (rachis)

En observant les adjectifs dérivés de noms, nous constatons que la relation sémantique qui unit le nom de tête d'un syntagme (NomT) et la base nominale de l'adjectif (NomB) varie selon les propriétés des noms de tête des syntagmes nominaux. Toutefois, nous avons constaté que les relations sémantiques qui traduisent l'ajout de sens qu'apporte l'adjectif se répètent chez les adjectifs qui partagent des caractéristiques semblables. Par exemple, la relation qui unit un *agent infectieux* et une *structure anatomique* est la même (*qui affecte*) que le syntagme soit *virus pulmonaire* (« virus qui affecte les poumons ») ou *bactérie gastrique* (« bactérie qui affecte l'estomac »).

4.1.5 Noms de tête

Au moment d'incorporer les paraphrases dans la base de données, nous avons tenu compte des noms situés en tête des syntagmes nominaux que nous avons analysés. Lors de l'analyse des 1300 syntagmes nominaux, nous avons observé plus de 600 noms différents qui se combinent aux 75 adjectifs dérivés de noms énumérés à la section 4.1.1. Nous avons également observé que certains de ces noms se combinent à plus d'un adjectif. Le nom de tête qui a produit le plus grand nombre de syntagmes est *infection*. Il se combine à 24 adjectifs dérivés de noms dont la liste est donnée au tableau 4.9.

Tableau 4.9 Adjectifs qui se combinent au nom *infection*

Adjectif	Base	Adjectif	Base
abdominal	abdomen	muqueux	muqueuse
asymptomatique	symptôme	osseux	os
bactérien	bactérie	pancréatique	pancréas
biliaire	voie biliaire	parasitaire	parasite
cellulaire	cellule	parenchymateux	parenchyme
cutané	peau	polymicrobien	microbe
digestif	digestion	pulmonaire	poumon
épidural	dure-mère	rachidien	rachis
grippal	grippe	respiratoire	respiration
hématogène	sang	tuberculeux	tuberculose
intestinal	intestin	urinaire	appareil urinaire
méningé	méninges	viral	virus

Cette liste, nous permet d'observer que les bases nominales des adjectifs peuvent être regroupées selon 3 classes sémantiques soit : agent pathogène (ex. *bactérie*), structure anatomique (ex. *abdomen*) et fonction physiologique (ex. *digestion*).

4.2 Modèle d'encodage

L'encodage des données consiste à inscrire « dans un article prévu à cette fin, les renseignements qui résultent de l'analyse qui [est faite] des termes : une définition, un contexte, des équivalents bilingues ou multilingues » (L'Homme 2004b : 46). Dans le cadre de notre travail, nous avons d'abord projeté d'élaborer un modèle d'encodage sous forme de dictionnaire papier. Toutefois, l'avancement du projet nous a mené vers la création d'une base de données en format électronique. En optant pour un tel modèle, il nous est possible d'enrichir et de mettre la base de données en ligne pour permettre à un grand nombre d'utilisateurs d'y accéder.

Dans cette section, nous présentons les champs de la base de données d'analyse qui ont été choisis pour le modèle d'encodage ainsi que ceux que nous avons jugés essentiel d'ajouter pour assurer une bonne compréhension et un usage efficace par le traducteur. Par la suite, nous effectuons un survol de la base de données pour voir les différents types de recherches qu'il est possible d'y effectuer.

4.2.1 *MédiTerm*

La base de données, que nous avons nommée *MédiTerm*, sert de point de départ à la création d'un *Dictionnaire multilingue des adjectifs médicaux* destiné à un usage Internet. L'intention d'intégrer d'autres langues à la base de données nous a incitée à adopter un modèle d'encodage simple pouvant renvoyer à la fiche jumelle des autres langues. Pour gérer la base de données, nous avons utilisé le logiciel Access, comme nous l'avons fait pour la base de données d'analyse.

4.2.1.1 Champs de la fiche terminologique

Lors de l'élaboration du modèle d'encodage, nous avons gardé en tête que le lectorat visé est constitué de traducteurs et de rédacteurs qui ont besoin de certaines données bien précises pour comprendre le texte de départ, effectuer une recherche terminologique ponctuelle ou produire un texte idiomatique. Au moment de créer la fiche de l'adjectif, nous avons décidé que chaque acception posséderait sa propre fiche au lieu de regrouper les acceptions sous leur base nominale. En procédant de la sorte, le traducteur peut voir toutes les données dont il a besoin sur une même page, selon l'acception qui convient à ses recherches (voir la section 4.2.1.2 pour la présentation des résultats générés par les recherches). Cela dit, notre modèle n'avait pas besoin de contenir tous les champs de la base d'analyse (fréquence de l'adjectif en corpus, indice de spécificité). Par contre, nous avons dû enrichir les fiches de quelques données, comme l'équivalent anglais et la définition de la base nominale. La figure 4.1 présente la fiche terminologique type d'un adjectif dérivé de nom étudié dans le cadre de ce travail.

Microsoft Access - [Fiche2]	
Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre 2 Arial 9 G J S [formatting icons]	
Acception viral 1a	Type Dérivé de nom Base Virus Définition
Anglais [empty field]	Paraphrase NomT qui est causé par un virus
Coréen [empty field]	NomT possible(s) infection, maladie (hépatite, méningite, etc.), fièvre (hémorragique), inflammation, affection, atteinte, lésion
Contexte(s)	
Dans l'hépatite virale, il existe des lésions dégénératives vacuolaires (aspect de clarification des cytoplasmes) et granulaires (apparition de granulations rouges acido-philés dans les cytoplasmes, liées à la précipitation des protéines).	
Les patients ont présenté des symptômes de fièvre, de gorge irritée, de toux et, pour les cas mortels, de troubles respiratoires sévères dus à l'infection pulmonaire virale.	
[empty text box]	
Env : 14 4 3 3 sur 3	

Figure 4.1 Fiche de l'acception *viral 1a* dans *MédiTerm*

Acception : Adjectif numéroté selon son acception, sous sa forme canonique.

Anglais : Équivalent anglais de l'adjectif pour cette acception. Dans le cadre du projet futur de dictionnaire multilingue, nous souhaitons développer le volet anglais de la base de données. Nous avons donc prévu un champ qui accueillera l'équivalent anglais correspondant à notre acception.

Coréen : Équivalent coréen pour cette acception. Dans le cadre du projet de dictionnaire multilingue, Bae (voir la section 2.2.4) a accepté de fournir des équivalents coréens à nos acceptions. Nous avons donc prévu le champ qui permettra d'accueillir les données et d'effectuer le lien vers la fiche du terme adjectival coréen correspondant à notre acception.

Type : Type de l'adjectif selon la typologie lexico-sémantique.

Base : Base sémantique des adjectifs dérivés de noms et de verbes. Les adjectifs qualificatifs et déterminatifs n'ayant pas de base sémantique nous avons mis la mention « aucune base » dans ce champ.

Définition : Bouton qui renvoie à une fenêtre (figure 4.2) donnant la définition de la base nominale, ainsi que sa source. En général, les définitions sont tirées du *Dictionnaire des termes de médecine* (Garnier, Delamare et al. 2002) et du *Dictionnaire médical* (Manuila, Manuila et al. 2001). Pour certaines unités lexicales plus générales (ex. *air*), nous avons fait appel à la version électronique du *Petit Robert* (Rey et Rey-Debove 2004).

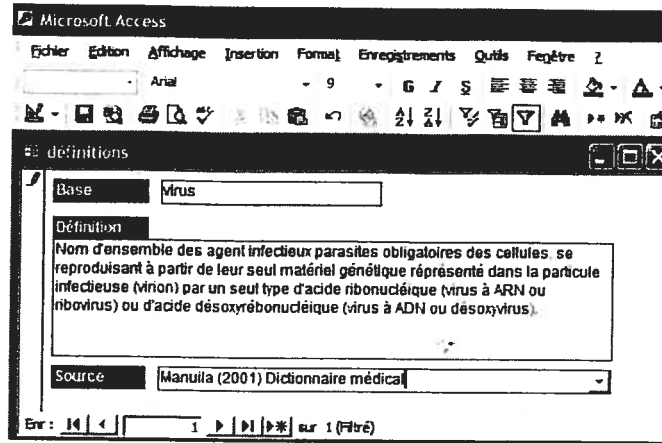


Figure 4.2 Fenêtre de définition de la base nominale *virus* dans *MédiTerm*

Paraphrase : Paraphrase définitoire expliquant l'acception dont il est question lorsque l'adjectif est combiné avec les noms de têtes de syntagmes (NomT).

NomT possible(s) : Liste des noms de tête de syntagmes auxquels peut se combiner un adjectif pris dans l'acception spécifique qui fait l'objet de la description. La liste est dressée à partir des occurrences de l'acception dans notre corpus et d'autres noms peuvent s'ajouter à la liste.

Contexte : Contextes tirés du corpus médical dans lesquels l'acception paraît et qui permettent de visualiser l'utilisation de l'adjectif dans un texte.

La fiche est conçue de façon à être traduite en anglais, en coréen et éventuellement dans d'autres langues.

4.2.1.2 Présentation des résultats générés par les recherches

Au moment de consulter la base de données, l'utilisateur choisit parmi trois critères de recherche : l'adjectif, la base nominale ou le nom de tête (figure 4.3). Cette section présente les résultats obtenus lors de chaque type de requête. Nous verrons un exemple des résultats obtenus pour chaque critère et nous exposerons les raisons qui peuvent motiver une telle recherche.

Microsoft Access

Fichier Edition Affichage Insertion Format Enregistrements Outils Fenêtre ?

Tahoma 10 G I S

Recherche : Formulaire

MédiTerm

Dictionnaire multilingue des adjectifs médicaux

Adjectif * Recherche

Base de l'adjectif * Recherche Ex. Écrire « virus » pour trouver l'adjectif « viral »

Nom de tête * Recherche Ex. Écrire « souche » pour trouver « souche virale » et d'autres adjectifs utilisés avec ce nom

Conseil : Conserver « * » à la fin de votre requête facilite vos recherches.

Entr : 1 sur 1

Figure 4.3 Formulaire de recherche de *MédiTerm*

Pour lancer une recherche, il suffit d'entrer le critère dans le champ correspondant. Nous suggérons d'encadrer la chaîne de caractères d'astérisques pour que la liste de résultats soit la plus vaste possible.

a) Adjectif

Effectuer une requête selon l'adjectif permet de voir les différentes acceptions de celui-ci. La liste de résultats générée par la requête présente la ou les bases nominales de l'adjectif et toutes les acceptions. À la figure 4.4, il est possible de voir les acceptions de l'adjectif, mais nous observons également que l'adjectif *muqueux* possède deux bases nominales, soit *mucus* et *muqueuse*.

The screenshot shows a Microsoft Access window with a table titled "Résultats Adj". The table has three columns: "Acception", "Base", and "Paraphrase". Each row represents a different sense of the adjective "muqueux", with a "Fiche" button next to the paraphrase.

Acception	Base	Paraphrase	Fiche
muqueux 1a	mucus	NomT qui produit du mucus	Fiche
muqueux 1b	mucus	NomT qui est caractérisé par la présence de mucus	Fiche
muqueux 1c	mucus	NomT qui est marquée par la présence de mucus	Fiche
muqueux 2a	muqueuse	NomT qui affecte la muqueuse	Fiche
muqueux 2b	muqueuse	NomT qui se situe sur/dans la muqueuse	Fiche
muqueux 2c	muqueuse	NomT qui est constitué de la muqueuse	Fiche

Figure 4.4 Liste de résultats générée par une requête selon l'adjectif *muqueux*.

b) Base de l'adjectif

La recherche à partir du nom de base de l'adjectif permet de voir les adjectifs qui sont dérivés d'un même nom. À partir de la simple liste, nous pouvons voir des ressemblances, souvent formelles, entre les dérivations. À la figure 4.5, nous voyons les adjectifs qui sont dérivés du nom *air*. Nous observons que les adjectifs possèdent la chaîne de caractère *aér-* qui renvoie au préfixe *aéro-* pour *air* (Rey et Rey-Debove 2004). Une telle information apporte un élément de sens qui pourrait donner des pistes de recherche lorsque l'adjectif qui pose problème n'est pas présent dans la base de données.

The screenshot shows the Microsoft Access interface with a table titled 'Résultats Base'. The table has three columns: 'Acception', 'Base', and 'Paraphrase'. Each row represents a different semantic class of adjectives related to 'air', with a 'Fiche' link on the right of each row.

Acception	Base	Paraphrase	Fiche
aérien 1a	air	NomT où circule l'air	Fiche
aérien 1b	air	NomT au moyen de l'air	Fiche
aérobie 1a	air	NomT qui nécessite la présence d'air	Fiche
anaérobie 1a	air	NomT qui nécessite l'absence d'air	Fiche
anaérobie 1b	air	NomT qui ne contient pas d'air	Fiche
antianaérobie 1a	air	NomT qui combat ce qui nécessite l'absence d'air	Fiche

Figure 4.5 Liste de résultats générée par une requête selon la base *air*.

c) Nom de tête

La requête lancée à partir du nom de tête permet de consulter la liste des adjectifs de notre corpus qui se combinent à ce nom. Ce type de recherche nous permet d'approfondir l'analyse des adjectifs en observant, entre autres, que les adjectifs d'une même classe sémantique possèdent la même paraphrase définitoire. Par exemple, la figure 4.6 permet d'observer que lorsque que *chirurgie* est combiné à un adjectif dérivé d'un nom dénotant une structure anatomique (ex. *biliaire*, *digestif*, *intestinal*, *thoracique*), la paraphrase définitoire est *NomT qui est effectué sur NomB*. Une telle information devient intéressante si l'utilisateur veut connaître le sens d'un adjectif dérivé d'une structure anatomique, comme *cardiaque* (*cœur*) ou *rénal* (*rein*), qui ne se trouve pas dans la base de données et qui est combiné à *chirurgie*.

The screenshot shows the Microsoft Access interface with a table titled 'Résultats Tête'. The table has three columns: 'Acception', 'Base', and 'Paraphrase'. Each row represents a search result with a corresponding 'Fiche' button. The status bar at the bottom indicates 'Enr: 14 | 4 | 5 sur 5'.

Acception	Base	Paraphrase	
biliaire 2b	voie biliaire	NomT qui est effectué sur les voies biliaires	Fiche
digestif 2c	appareil digestif	NomT qui est effectué sur le appareil digestif	Fiche
intestinal 1c	intestin	NomT qui est effectué sur l'intestin	Fiche
septique 1b	sepsie	NomT qui provoque une sepsie	Fiche
thoracique 1b	thorax	NomT qui est effectué sur le thorax	Fiche

Figure 4.6 Liste de résultats générée lors d'une requête d'après le nom de tête *chirurgie*.

Dans les trois cas, l'utilisateur qui souhaite obtenir de plus amples renseignements sur l'une ou l'autre des acceptions présentes dans la liste de résultats, n'a qu'à sélectionner l'acception de son choix pour accéder à la fiche présentée à la figure 4.1.

Conclusion

Notre étude visait à isoler et analyser les adjectifs dérivés de noms afin de les regrouper dans une base de données terminologiques dont l'encodage permet une désambiguïsation de leur sens.

Pour parvenir à dresser une liste d'adjectifs dérivés de noms spécifiques au domaine médical et à en désambiguïser le sens, nous avons mis en forme un corpus de 220 000 mots regroupant 67 textes médicaux portant sur l'infectiologie. À partir de ce corpus, nous avons procédé au traitement automatique du texte de façon à extraire 2597 candidats adjectifs à l'aide de TermoStat, qui leur a attribué un indice de spécificité. Nous avons ensuite effectué un nettoyage des candidats adjectifs pour obtenir une liste finale de 1645 adjectifs.

Dans un deuxième temps, nous avons trié les adjectifs d'après leur indice de spécificité de façon à obtenir les plus spécifiques en tête de liste. Nous avons alors procédé à leur classification pour isoler ceux qui étaient dérivés de noms. Les adjectifs de ce groupe possèdent deux principales caractéristiques : 1) ils possèdent une base nominale (nom dont le sens est présent dans celui de l'adjectif); et 2) ils expriment une relation avec le nom auquel ils sont combinés.

Nous avons choisi de procéder à l'analyse du sens des adjectifs au moment de les classer, car cette dernière elle était nécessaire à la vérification des critères de classification qui sont présentés à la section 1.2 du premier chapitre. Nous avons choisi d'analyser les adjectifs en commençant par les plus spécifiques et nous avons cessé lorsque l'analyse de 5 adjectifs dérivés de noms n'avait pas généré de nouvelle paraphrase définitoire. Nous avons donc analysé 86 adjectifs dont 75 dérivés de noms. Pour désambiguïser ces adjectifs, nous avons étudié la relation sémantique qui unit le

nom de base de l'adjectif au nom de tête dans près de 1300 syntagmes nominaux différents. Cet exercice nous a permis de mettre en évidence près de 382 acceptions au moyen de 74 paraphrases définitoires.

Après avoir complété l'analyse des adjectifs, nous avons procédé à l'encodage des données dans une base de données qui permet à des utilisateurs, principalement des traducteurs et des rédacteurs médicaux, d'accéder à divers renseignements sur les adjectifs : acception, base nominale, définition de cette dernière, noms avec lesquels l'adjectif est combiné dans notre corpus, équivalent anglais et coréen de l'adjectif pour l'acception désirée et contextes. Le modèle d'encodage permet de lancer une requête à partir d'un formulaire de recherche. Cette requête génère une liste d'acceptions et l'utilisateur peut choisir celle pour laquelle il souhaite consulter la fiche détaillée.

Le modèle servira à la mise en ligne d'une base de données terminologiques nommée *MédiTerm* qui regroupe, dans un premier temps, les adjectifs médicaux faisant l'objet du présent projet. Le modèle d'encodage a également été conçu en prévision d'une exportation des données pour la création d'une ressource terminologique en format papier.

Le système de paraphrases utilisé lors de la rédaction des paraphrases définitoires compte 74 paraphrases types qui présentent à la fois des relations sémantiques simples et complexes. Ce système a d'abord été élaboré pour assurer l'uniformité des paraphrases au fil de l'analyse, mais il nous a permis de constater une répétition de celles-ci. En effet, les paraphrases ont servi en moyenne à expliquer une des acceptions de 5,14 adjectifs. Soulignons toutefois que la relation sémantique simple *de(2)* a servi dans le cas de 40 acceptions et la relation complexe *qui affecte NomB* a servi pour 33 acceptions.

À la lumière de ces résultats, nous avons tiré un certain nombre de conclusions. Premièrement, la classification des adjectifs nous a permis de constater que les plus

spécifiques à notre corpus étaient principalement dérivés de noms. Cette observation confirme l'importance des adjectifs dérivés de noms dans la langue médicale et justifie la création de ressources terminologiques telle que *MédiTerm*.

Deuxièmement, nous avons observé que l'analyse des adjectifs en corpus permet de tenir compte des distinctions sémantiques et de fournir des paramètres menant à une description plus précise dans les dictionnaires médicaux. Les adjectifs étant des modificateurs, nous avons concentré notre analyse sur leur sens pour bien mettre en évidence la relation qui est véhiculée par l'adjectif lorsqu'il est combiné à un nom donné. De cette façon, les paraphrases définitives permettent d'obtenir une définition plus adaptée au contexte dans lequel l'adjectif est utilisé permettant à l'utilisateur de la base de données de faire un meilleur usage de celui-ci dans le cadre de ses travaux.

Enfin, l'analyse des paraphrases définitives nous a permis de mettre en évidence des régularités parmi les relations qui existent entre les noms de tête et de base. Nous avons observé qu'une même paraphrase décrit la relation qui existe entre deux classes sémantiques. Nous avons, par exemple, remarqué que la paraphrase *qui est effectué sur* sert à lier des NomT qui appartiennent à la classe sémantique <intervention> et des NomB qui appartiennent à la classe sémantique <structure anatomique>. Prenons l'adjectif *abdominal*. Lorsqu'il est combiné aux noms de tête *chirurgie*, *échographie* et *poussée*, noms qui appartiennent à la classe sémantique *intervention*. Son acception se décrit alors comme suit : *qui est effectué sur l'abdomen*. Par la suite, nous avons tenté d'effectuer un rapprochement sémantique semblable avec les noms de base des adjectifs partageant cette acception, soit *abdomen*, *poumon*, *peau*, *os*, *rein* et *méninges*, pour ne nommer qu'eux. Comme les NomT, les bases nominales appartiennent toutes à la même classe sémantique, soit *structure anatomique*.

Cette observation nous permet de croire qu'il est possible d'exploiter la modélisation d'un système de paraphrases pour automatiser l'enrichissement d'outils

terminologiques comme *MédiTerm*. En effet, il semble qu'une modélisation soit possible en appliquant le raisonnement qui suit au système de paraphrases :

Si *infection virale* =

infection + qui est causée par + un virus

et *inflammation bactérienne* =

inflammation + qui est causée par + une bactérie

alors la combinaison des classes sémantiques <état physiologique> et <agent pathologique> =

<état physiologique> + qui est causé par + <agent pathologique>

Par des travaux ultérieurs, nous aimerions vérifier l'hypothèse selon laquelle la modélisation d'un système de paraphrases qui repose sur les classes sémantiques permet de prédire la relation existant entre un nom de base et un nom de tête. Une telle modélisation permettrait de donner des informations sur les acceptions adjectivales même si les syntagmes recherchés ne sont pas répertoriés comme tels dans un dictionnaire. Enfin, nous espérons utiliser ce système de paraphrases modélisé pour enrichir une base de données multilingue de façon semi-automatique.

En effet, nous sommes tentée de croire que la relation sémantique qui unit deux noms (NomT-NomB) sera la même pour les différentes langues. Cela dit, nous croyons qu'il est possible, dans un premier temps, de modéliser le système de paraphrases élaboré dans le cadre de cette étude en utilisant les classes sémantiques des noms de tête et de base des adjectifs analysés. Dans un deuxième temps, nous pourrions vérifier l'applicabilité du système de paraphrases sur des adjectifs extraits de corpus français issus d'autres sous-domaines de la médecine (ex. anatomie, chirurgie). Enfin, nous procéderons au même exercice sur des corpus comparables anglais, en effectuant les modulations nécessaires au système de paraphrases modélisées.

De cette façon, nous pourrions automatiser, en partie, l'attribution des paraphrases définitoires aux paires *NomT-NomB*, en faisant référence à leurs classes sémantiques. Une telle automatisation permettrait de développer le volet anglais de la base de données terminologiques *MédiTerm* qui a été créée dans le cadre du présent projet de recherche.

À la lumière des recherches effectuées au cours du projet, nous voyons également d'autres pistes de travaux portant sur les adjectifs dérivés de noms. Il serait intéressant d'explorer le rôle des affixes gréco-latins dans l'explication des acceptions adjectivales. De plus, l'utilisation d'une segmentation automatique des adjectifs construits d'après le patron *préfixe-base-suffixe* pourrait servir à mettre en évidence les composés savants présents et permettre d'obtenir des renseignements sémantiques de façon automatique.

Nous sommes consciente que l'intérêt grandissant pour les adjectifs en terminologie se justifie, entre autre, par le rôle qu'ils jouent dans les textes spécialisés, dans le cas présent, en médecine. Notre projet de recherche nous a permis de faire le point sur les caractéristiques des groupes adjectivaux et de brosser un tableau de la place que les adjectifs dérivés de noms occupent dans notre corpus médical. Il a également permis de faire un pas de plus dans l'analyse des adjectifs dérivés de noms en corpus médical tout en soulevant différents points qui pourront être approfondis dans le cadre d'études ultérieures. Nous reconnaissons, que la modélisation proposée devra faire l'objet de vérifications supplémentaires pour assurer la stabilité du système de paraphrases. Par contre, nous croyons qu'elle apportera des pistes de solutions qui permettront une optimisation du traitement des textes spécialisés.

Bibliographie

- ARRIVÉ, Michel, GADET, Françoise et GALMICHE, Michel (1986) : *La grammaire d'aujourd'hui : guide alphabétique de linguistique française*. Paris, Flammarion.
- BAAAYEN, Harald (2001) : *Word frequency Distribution*. Dordrecht et Boston, Kluwer Academic Publishers.
- BAE, Hee Sook (2006) : «Termes adjectivaux en corpus médical coréen : repérage et désambiguïsation ». *Terminology* 12(1) pp. 19-50.
- BOURIGAULT Didier, FABRE Cécile, FRÉROT Cécile, JACQUES Marie-Paule et OZDOWSKA Sylwia (2005) : « Syntex, analyseur syntaxique de corpus ». Dans *Actes des 12èmes journées sur le Traitement Automatique des Langues Naturelles*, Dourdan, France.
- BOWKER, Lynne et PEARSON, Jennifer (2002) : *Working with Specialized Language : A practical guide to using corpora*. Londres et New York, Routledge.
- Bureau de la traduction (2006) : « Termium Plus : la base de données terminologiques et de linguistique du gouvernement du Canada », Bureau de la traduction. tiré de http://www.termium.gc.ca/site/accueil_home_f.html
- CABRÉ, Maria Teresa (1998) : *La terminologie : théorie, méthode et applications*. Ottawa, Les Presses de l'Université d'Ottawa.
- CARRENO CRUZ, Sahara Iveth (2005) : *Analyse de la variation terminologique en corpus parallèle anglais-espagnol et de son incidence sur l'extraction de termes bilingues*. Département de linguistique et de traduction, Montréal, Université de Montréal. Mémoire de maîtrise.

- CLAS, André (1994) : « Collocations et langues de spécialité ». *Meta* 39(4), pp. 576-580.
- CÔTÉ, Roger A. (1996) : *Répertoire d'anatomopathologie de la SNOMED internationale*. v3.4, Sherbrooke, Université de Sherbrooke.
- DAILLE, Béatrice (2001) : « Qualitative terminology extraction : Identifying relational adjectives ». Dans Marie-Claude L'Homme, Didier Bourigault et Christian Jacquemin. *Recent Advances in Computational Terminology*. Amsterdam, J. Benjamins Pub. Co. pp. 149-166.
- DELISLE, Jean (1993) : *La traduction raisonnée*. Ottawa, Les Presses de l'Université d'Ottawa.
- DROUIN, Partrick (2003) : « Term extraction using non-technical corpora as a point of leverage ». *Terminology* 9(1), pp. 99-117.
- DROUIN, Patrick (2004) : « Spécificités lexicales et acquisition de la terminologie ». *7^{es} journées internationales d'Analyse statistique de Données textuelles (JADT-2004)*, Louvain-la-Neuve, Belgique, p. 345-352.
- DUBOIS, Jean, GIACOMO, Mathée, GUESPIN, Louis, MARCELLEŞI, Christiane, MARCELLESI, Jean-Baptiste et MÉVEL, Jean-Pierre (2002) : *Dictionnaire de linguistique*. Paris, Larousse.
- GARNIER, Marcel, DELAMARE, Valery, DELAMARE, Jean et DELAMARE, Thérèse (2002) : *Dictionnaire des termes de médecine*. 27e, Paris, Maloine.
- GRABAR, Natalia et ZWEIGENBAUM, Pierre (2003) : « Productivité à travers domaines et genres : dérivés adjectivaux et langue médicale ». *Langue française*, 140, pp. 102-125.

- L'HOMME, Denis, L'HOMME, Marie-Claude et LEMAY, Chantal (2002) : « Benchmarking the Performance of two Part-of-speech (POS) Taggers for Terminological Purposes ». *Knowledge Organization* 29(3-4), pp. 204-216.
- L'HOMME, Marie-Claude (1998) : « Le Statut du verbe en langue de spécialité et sa description lexicographique ». *Cahiers de lexicologie* 73(2), pp. 61-64.
- L'HOMME, Marie-Claude (2003) : « Capturing the lexical structure in special subject fields with verbs and verbal derivatives. A model for specialized lexicography ». *International Journal of Lexicography* 16(4), pp. 403-422.
- L'HOMME, Marie-Claude (2004a) : « Adjectifs dérivés sémantiques (ADS) dans la structuration des terminologies ». Dans *Actes Terminologie, ontologie et représentation des connaissances*, Université Jean-Moulin Lyon 3, 22-23 janvier 2004.
- L'HOMME, Marie-Claude (2004b) : *La terminologie : principes et techniques*. Montréal, Les Presses de l'Université de Montréal.
- L'HOMME, Marie-Claude (2005) : « Conception d'un dictionnaire fondamental de l'informatique et de l'Internet : sélection des entrées ». *Le langage et L'homme* 40(1), pp. 137-154.
- L'HOMME, Marie-Claude, HEID, Ulrich et SAGER, Juan C. (2003) : « Terminology during the past decade ». *Terminology* 9(2), pp. 151-161.
- LEESA, Bruce D. (éd.)(2000) CPS - Compendium des produits et spécialités pharmaceutiques, 35^e édition, Toronto, Association des pharmaciens du Canada.

- LEMAY, Chantal (2003) : « Identification automatique du vocabulaire caractéristique du domaine de l'informatique fondée sur la comparaison de corpus ». *Département de linguistique et de traduction*, Montréal, Université de Montréal. Mémoire de maîtrise.
- LETHUILLER, Jacques (2003) : « Enseignement des langues de spécialités comme préparation à la traduction spécialisée ». *Meta*, 48(4), pp. 379-392
- MANIEZ, François (2001) : « La traduction du nom adjectival en anglais médical ». *Meta* 46(1), pp. 56-67.
- MANIEZ, François (2002) : « Distinguer les termes des collocations : étude sur corpus du patron <Adjectif - Nom> en anglais médical ». *Actes du colloque TALN de Nancy*, 24-27 juin 2002, Tome 1, pp. 345-350.
- MANIEZ, François (2004) : « Identification automatique des adjectifs relationnels : une étude sur corpus ». Dans Henri Béjoint et François Maniez. *De la mesure dans les termes*. Lyon, Presses Universitaires de Lyon.
- MANUILA, Ludmila, MANUILA, Alexandre, LEWALLE, Pierre et NICOULIN, Monique (2001) : *Dictionnaire médical*. 9^e édition, Paris, Masson.
- MÉLIS-PUCHULU, Agnès (1991) : « Les adjectifs dénominatifs, des adjectifs de "relation" ». *Lexique* 10, pp. 33-60.
- MONCEAUX, Anne (1997) : « Adjectifs de relation, complémentation et sous-classification ». In J.-S. Nam. *La description syntaxique des adjectifs pour les traitements informatiques*. Paris. 126, pp. 39-59.
- NOAILLY, Michèle (1999) : *L'adjectif en français*. Gap, Ophrys.

- NORMAND, Sylvie et BOURIGAU, Didier (2001) : « Analysing adjectives used in a histopathology corpus with NLP tools ». *Terminology* 7(2), pp. 155-165.
- OMS (1995): « Classification statistique internationale des maladies et des problèmes de santé connexes - dixième révision », Organisation mondiale de la santé, Genève.
- OQLF. (2006, 14-11-2005) : « Grand dictionnaire terminologique (GDT) ». Consulté le 21 novembre, 2005, tiré de http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp.
- POLGUÈRE, Alain (2003) : *Lexicologie et sémantique lexicale : notions fondamentales*. Montréal, Presses de l'Université de Montréal.
- RACINE, Amélie (2006) : « Extraction des néologismes dans des textes biomédicaux ». *Département de linguistique et de traduction*, Montréal, Université de Montréal. Travail dirigé pour la Maîtrise professionnelle en traduction.
- REY, Alain et REY-DEBOVE, Josette (2004) : *Le Petit Robert - Dictionnaire alphabétique et analogique de la langue française*. Paris, Société Dictionnaires Le Robert.
- RIEGEL, Martin, PELLAT, Jean-Christophe et RIOUL, René (2004) : *Grammaire méthodique du français*. 3^e éd. (Quadrige), Paris, Quadrige/PUF.
- SAGER, Juan C. (1990) : *A Practical Course in Terminology Processing*. Amsterdam/Philadelphie, John Benjamin Pub.
- SCHMID, Helmut (1994) : « Probabilistic part-of-speech tagging using decision trees ». Proceedings of *International Conference on New Methods on Language Proceeding*, septembre 1994, Manchester, Royaume-Uni, sp.

- SCHNEDECKER, Catherine (2002) : « Présentation : les adjectifs « inclassables », des adjectifs du troisième type ». *Langue française* 136, pp. 3-19.
- TAMBA-MECZ, Irène (1980) : « Sur quelques propriétés de l'adjectif de relation ». *Travaux de linguistique et de littérature* XVIII(1), pp. 119-132.
- ZWEIGENBAUM, Pierre, HADOUCHE, Fadila et GRABAR, Natalia (2003) : « Apprentissage de relations morphologiques en corpus ». Dans Béatrice Daille. *Actes de TALN 2003 (Traitement automatique des langues naturelles)*, Batz-sur-Mer, juin 2003., pp. 285-294.

Annexe A : Liste des textes du Corpus

Liste des textes qui constituent le corpus

Fichier	Référence	Consulté le	Nbre mots
anapath1.txt	FACULTE DE MEDECINE NECKER ENFANTS-MALADES (éd.) « RÉPONSE IMMUNITAIRE SPÉCIFIQUE ET PATHOLOGIE AUTO-IMMUNE », chapitre 5 du polycopié 1998-1999 d'Anatomie Pathologique de la Faculté Necker-Enfants Malades. En ligne. Adresse URL : http://www.anapath.necker.fr/enseign/poly/chap5_nem.html	2005-04-20	3641
anapath2.txt	FACULTE DE MEDECINE NECKER ENFANTS-MALADES (éd.) « RÉPONSE IMMUNITAIRE SPÉCIFIQUE ET PATHOLOGIE AUTO-IMMUNE », chapitre 4 du polycopié 1998-1999 d'Anatomie Pathologique de la Faculté Necker-Enfants Malades. En ligne. Adresse URL : http://www.anapath.necker.fr/enseign/poly/chap4_nem.html	2005-04-20	7501
anapath3.txt	DUYCKAERTS C., P. FOURET et J.-J. HAUW (2003) « Chapitre 3 - Inflammation », Anatomie pathologique PCEM2 2002 - 2003, Université PARIS-VI Pierre et Marie Curie, Faculté de Médecine Pitié-Salpêtrière. En ligne. Adresse URL : http://www.chups.jussieu.fr/polys/anapath/Cours/anapath.pdf	2005-04-20	7007
bacterio1.txt	NASSIF X. et A. PHILIPPON (éd.) « Relation hôte-pathogène », Cours de bactériologie générale, Faculté de Médecine Necker-Enfants-Malades, Paris V et Faculté de Médecine Cochin-Port-Royal, Paris V. En ligne. Adresse URL : http://www.microbe-edu.org/diagnostics/index.html	2005-04-20	1859
bacterio2.txt	PHILIPPON A. et L. PROTS (éd.) « Diagnostic d'une infection bactérienne I », Cours de bactériologie générale, Faculté de Médecine Cochin-Port-Royal, Paris V. En ligne. Adresse URL : http://www.microbe-edu.org/diagnostics/index.html	2005-04-20	1175
bacterio3.txt	PHILIPPON A. et L. PROTS (éd.) « Diagnostic d'une infection bactérienne II », Cours de bactériologie générale, Faculté de Médecine Cochin-Port-Royal, Paris V. En ligne. Adresse URL : http://www.microbe-edu.org/diagnostics/index.html	2005-04-20	1775
botulisme1.txt	FAURE E. « Le botulisme », Dossier santé. Réseaux et Systèmes d'Information Santé au service des professionnels, Octobre 2001. En ligne. Adresse URL : http://www.caducee.net/DossierSpecialises/infection/botulisme.asp	2005-04-17	1513
botulisme2.txt	ORGANISATION MONDIALE DE LA SANTÉ « Botulisme », Aide-mémoire no 270, révisé août 2002. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs270/fr/	2005-04-20	1267
clostri1.txt	EUZÉBY J.P. (2003) « CLOSTRIDIUM BOTULINUM, CLOSTRIDIUM ARGENTINENSE, CLOSTRIDIUM BARATII, CLOSTRIDIUM BUTYRICUM » (Partie I), Dictionnaire de Bactériologie Vétérinaire, Ressources de la Faculté de	2005-04-17	8818

	médecine vétérinaire de l'Université de Montréal. En ligne. Adresse URL : http://www.bacterio.cict.fr/bacdico/cc/botulinum.html		
clostri2.txt	EUZÉBY J.P. (2003) « CLOSTRIDIUM BOTULINUM, CLOSTRIDIUM ARGENTINENSE, CLOSTRIDIUM BARATII, CLOSTRIDIUM BUTYRICUM » (Partie II), Dictionnaire de Bactériologie Vétérinaire, Ressources de la Faculté de médecine vétérinaire de l'Université de Montréal. En ligne. Adresse URL : http://www.bacterio.cict.fr/bacdico/cc/botulinum.html	2005-04-17	4678
clostri3.txt	EUZÉBY J.P. (2003) « CLOSTRIDIUM DIFFICILE », Dictionnaire de Bactériologie Vétérinaire, Ressources de la Faculté de médecine vétérinaire de l'Université de Montréal. En ligne. Adresse URL : http://www.bacterio.cict.fr/bacdico/cc/difficile.html	2005-04-17	8121
clostri4.txt	GROLLIER G., G. LE MOAL et R. ROBERT. « Infections dues aux bactéries anaérobies de la flore endogène (Clostridium difficile et Actinomyces exclus) (Partie 1) », EMC - Maladies Infectieuses. Volume 1, Numéro 4, Novembre 2004, Pages 262-280. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317984%232004%23999989995%23526967%23FLA%23Volume_1,_Issue_4,_Pages_235-315_(November_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=2ecec2c35c8c4c4ae3715608a5209fe2	2005-04-17	5263
clostri5.txt	GROLLIER G., G. LE MOAL et R. ROBERT. « Infections dues aux bactéries anaérobies de la flore endogène (Clostridium difficile et Actinomyces exclus) (Partie 2) », EMC - Maladies Infectieuses. Volume 1, Numéro 4, Novembre 2004, Pages 262-280. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317984%232004%23999989995%23526967%23FLA%23Volume_1,_Issue_4,_Pages_235-315_(November_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=2ecec2c35c8c4c4ae3715608a5209fe2	2005-04-17	3720
ebola1.txt	ORGANISATION MONDIALE DE LA SANTÉ « Fièvre hémorragique à virus Ebola », Aide-mémoire no 103, mai 2004. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs103/fr/index.html	2005-04-20	1522
ebola2.txt	LE GUENNO B. « Fièvre hémorragique à virus Ebola », Bulletin épidémiologique hebdomadaire, février 1996, Ministère du travail et des affaires sociales - Direction générale de la santé. En ligne. Adresse URL : http://www.invs.sante.fr/beh/1996/9606/	2005-04-20	2348
ebola3.txt	CAMPAGNE F. « Fièvre hémorragique à virus Ebola », Dossier santé, Réseaux et Systèmes d'Information Santé au service des professionnels, nov. 2000 En ligne. Adresse URL : http://www.caducee.net/DossierSpecialises/infection/ebola.asp	2005-04-22	1384
ehrlichia1.txt	EUZÉBY J.P. (2001) « NEORICKETTSIA RISTICII », Dictionnaire de Bactériologie Vétérinaire, Ressources de la	2005-04-17	3908

	Faculté de médecine vétérinaire de l'Université de Montréal. En ligne. Adresse URL : http://www.bacterio.cict.fr/bacdico/nn/risticii.html		
ehrlichia2.txt	EUZÉBY J.P. (2001) « EHRlichia CHAFFEENSIS », Dictionnaire de Bactériologie Vétérinaire, Ressources de la Faculté de médecine vétérinaire de l'Université de Montréal. En ligne. Adresse URL : http://www.bacterio.cict.fr/bacdico/ee/chaffeensis.html	2005-04-17	4406
enterovirus1.txt	ORGANISATION MONDIALE DE LA SANTÉ « Entérovirus non poliomyélitiques », Aide-mémoire no 174, révisé mai 2000. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs174/fr	2005-04-20	727
enterovirus2.txt	COLIMON Pr. « Picomaviridae », Département de Virologie, CHU de Rennes, novembre 2001. En ligne. Adresse URL : http://www.med.univ-rennes1.fr/resped/s/viro/picorna/picorna.htm	2005-04-22	1607
fièvre1.txt	STAHL, J.-P. (éd.) « Fièvre aiguë chez l'adulte. Critères de gravité d'un syndrome infectieux », Corpus médical - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www.sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/203/lecon203.htm	2005-04-19	7363
gremip1.txt	MITTAL K.R. (2004) « Nouveautés concernant le diagnostic des infections à Haemophilus parasuis » Bulletin Info-Gremip, Numéro 1, janvier 2004, St-Hyacinthe. En ligne. Adresse URL : http://www.medvet.umontreal.ca/gremip/pdf/Bulletins/bulletin-jan2004.pdf	2005-04-14	712
gremip2.txt	MITTAL K.R. (2004) « Détection par PCR et génotypage d'isolats d' Haemophilus parasuis » Bulletin Info-Gremip, Numéro 2, juin 2004, St-Hyacinthe. En ligne. Adresse URL : http://www.medvet.umontreal.ca/gremip/pdf/Bulletins/bulletin-juin2004.pdf	2005-04-14	583
gremip3.txt	GOTTSCHALK M. (2004) « Actualités sur la pleuropneumonie porcine (partie 1) : biotypes et sérotypes » Bulletin Info-Gremip, Numéro 3, décembre 2004, St-Hyacinthe. En ligne. Adresse URL : http://www.medvet.umontreal.ca/gremip/pdf/Bulletins/bulletin-dec2004.pdf	2005-04-14	1031
grippe1.txt	STAHL, J.-P. (éd.) « Grippe », Corpus médical - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www.sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/82/lecon82.htm	2005-04-19	1682
grippe2.txt	ORGANISATION MONDIALE DE LA SANTÉ « La grippe », Aide-mémoire no 211, révisé mars 2003. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs211/fr/	2005-04-20	1697
grippe3.txt	INVS (éd.) « La grippe », Cours de virologie générale, Aide-Mémoire, octobre 2002. Institut de Veille Sanitaire. En ligne. Adresse URL :	2005-04-21	1074

	http://www.invs.sante.fr/communication/index_cp.htm		
grippeav1.txt	ORGANISATION MONDIALE DE LA SANTÉ « Grippe aviaire », Aide-mémoire grippe aviaire, janvier 2004. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/avian_influenza/fr/	2005-04-20	2314
hepatB.txt	ORGANISATION MONDIALE DE LA SANTÉ « Hépatite B », Aide-mémoire no 204, révisé octobre 2000. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs204/fr/	2005-04-20	1316
hepatB2.txt	VILLENEUVE J.P. « Hépatite B », Les hépatites virales, Le réseau d'information sur l'hépatite, révisé septembre 1998. En ligne. Adresse URL : http://www.hepnet.com/hepbfr/qag.html	2005-04-22	1605
hepatC.txt	ORGANISATION MONDIALE DE LA SANTÉ « Hépatite C », Aide-mémoire no 164, révisé octobre 2000. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs164/fr/	2005-04-20	1361
hepatC2.txt	VILLENEUVE J.P. « Hépatite C », Les hépatites virales, Le réseau d'information sur l'hépatite, révisé septembre 1998. En ligne. Adresse URL : http://www.hepnet.com/hepcfr/qag.html	2005-04-22	889
hepatE.txt	ORGANISATION MONDIALE DE LA SANTÉ « Hépatite E », Aide-mémoire no 280, août 2004. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs280/fr/	2005-04-20	996
infectio1.txt	DESENCLOS J.-C., H. DE VALK . « Les maladies infectieuses émergentes : importance en santé publique, aspects épidémiologiques, déterminants et prévention », Médecine et maladies infectieuses. Volume 35, numéro 2, février 2005, Pages 49-61. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=JournalURL&_cdi=6613&_auth=y&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=a335c59962dfdd92aacbc2af1663e69c	2005-04-17	8793
infectio2.txt	VABRET A. « Émergences et barrières d'espèces », Médecine et maladies infectieuses. Volume 34, numéro 11, Novembre 2004, Pages 506-513. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%236613%232004%23999659988%23527456%23FLA%23Volume_34,_Issue_11,_Pages_493-549_(November_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=16d00399e0cb87b371381a72561dc382	2005-04-17	5342
infectio3.txt	ADER F., J. SALOMON, C. PERRONNE AND L. BERNARD. « Origine de l'infection osseuse : endogène ou exogène ? Éléments de physiopathologie », Médecine et maladies infectieuses. Volume 34, Numéro 11, Novembre 2004, Pages 530-537. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%236613%232004%23999659988%23527456%23FLA%23Volume_34,_Issue_11,_Pages_493-549_(November_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=16d00	2005-04-17	3884

	399e0cb87b371381a72561dc382		
inftoxico1.txt	STAHL, J.-P. (éd.) « Infection chez le toxicomane », Corpus médicale - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www.sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/hp1/leconhp1.htm	2005-04-19	1251
infvirale1.txt	LE GALL P. (2004) « Inflammation virale », Cours de médecine. En ligne. Novembre 2001. Adresse URL : http://www.med.univ-rennes1.fr/resped/s/anapath/inflam/inflam.htm	2005-01-25	1601
infvirale2.txt	BRAHIC M. « Pathogenèse des infections virales neurotropes », Cours de l'Unité d'immunologie virale, Institut Pasteur, DEA Jussieu 2001-2002. En ligne. Adresse URL : http://virologie.free.fr/09-Pathogenese_infections_vir_neuro/Pathogenese_infections_vir_neuro.htm	2005-04-22	8360
inoculation1.txt	STAHL, J.-P. (éd.) « Pathologie d'inoculation », Corpus médicale - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www.sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/101/lecon101.htm	2005-04-19	1036
meningo1.txt	STAHL, J.-P. (éd.) « Méningites infectieuses et Méningo-encéphalite de l'adulte », Corpus médicale - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www.sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/96/lecon96.htm	2005-04-19	2478
meningo2.txt	ORGANISATION MONDIALE DE LA SANTÉ « Méningite à méningocoques », Aide-mémoire no 141, révisé mai 2003. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs141/fr/	2005-04-20	2063
neuroimmuno1.txt	LAFON M. « Pathogenèse des infections virales neurotropes », Cours de l'Unité d'immunologie virale, Institut Pasteur, janvier 2001. En ligne. Adresse URL : http://virologie.free.fr/10-NeuroImmunologie_virale/NeuroImmunologie_virale.htm	2005-04-22	2632
paludisme1.txt	CASALINO E. « Paludisme », EMC - médecine. Volume 1, Numéro 6, décembre 2004, Pages 580-591. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317983%232004%23999989993%23530071%23FLA%23Volume_1_Issue_6_Pages_487-597_(December_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=df5c53aa5430f70f98d26f0cc72a025c	2005-04-21	4927
paludisme2.txt	ROSIER F. « Paludisme », D. Mazier et V. Lavarde (éd.) Dossier, Fondation recherche médicale. En ligne. Adresse URL :	2005-04-22	3095

	http://www.frm.org/informez/info_ressources_dossiers_artic_e_sommaire.php?id=13&type=10&numeroChapitre=		
pancreas1.txt	DEBONNE J.M., F. klotz et P. rey. « Pancréas parasitaire », EMC - hépato-gastroentérologie. Volume 2, Numéro 1, février 2005, Pages 1-58. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2324647%232005%23999979998%23568169%23FLA%23Volume_2,_Issue_1,_Pages_1-58_(February_2005)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=e8521decd2c696ee5bf75f776dde67a0	2005-04-21	4490
patho3.txt	CORDOLIANI, Y.S. (HIA du Val-de-Grâce Paris) (Page consultée le 6 janvier 1998). Pathologie infectieuse encéphalique. En ligne. Adresse URL : http://www.med.univ-rennes1.fr/cerf/edicerf/NR/NR012.html#RTFToC1 (605268) [patho3.txt]	1998-01-06	5642
pediatrie1.txt	PLANTAZ D. (2004) « Fièvre aiguë et critères de gravité d'un syndrome infectieux chez l'enfant », Cours de médecine. En ligne. Novembre 2004, En ligne. Adresse URL : http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/pedia/malinfped/203a/econ203a.htm	2005-01-25	2434
pediatrie2.txt	PLANTAZ D. (2004) « Diarrhée aiguë du nourrisson », Cours de médecine. En ligne. Février 2004. Adresse URL : http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/pedia/malinfped/194a/econ194a.htm	2005-01-25	1073
pediatrie3.txt	BOST-BRU C. et D. PLANTAZ « Méningites infectieuses de l'enfant », Cours de médecine. En ligne. Novembre 2004. Adresse URL : http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/pedia/malinfped/96/lec096.htm	2005-01-25	2051
pediatrie4.txt	DUBOS F., b., M. Llorota, M. Souliera, F. Rozenbergc, P. Lebonc et D. Gendrela . « Production d'interféron alpha dans le sérum des très jeunes nourrissons lors d'infections virales », Médecine et maladies infectieuses. Volume 34, numéro 12, décembre 2004, Pages 561-565. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=JournalURL&_cdi=6613&_auth=y&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=a335c59962dfdd92aacbc2af1663e69c&chunk=34#34	2005-04-17	3099
pericar.txt	ALMANGE, Pr. (Page consultée le 9 janvier 1999). Affections du péricarde, En ligne. Adresse URL : http://www.med.univ-rennes1.fr/galesne/cardio/index4.htm .	1999-01-09	2502
rach.txt	MANELFE, C. et I. BERRY (Page consultée le 6 janvier 1998). Infections du rachis, En ligne. Adresse URL : http://www.med.univ-rennes1.fr/cerf/edicerf/NR/NR021.html#RTFToC1 (605268).	1998-01-06	4207
risquesmicro1.txt	BOSSI P., F. KLOTZ et P. HAUSFATER. « Risques bactériologiques (Partie I) », EMC - médecine. Volume 2, Numéro 1, février 2005, Pages 107-125. En ligne. Adresse URL :	2005-04-21	6668

	http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317983%232005%23999979998%23562019%23FLA%23Volume_2,_Issue_1,_Pages_1-125_(February_2005)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=b4d42c9a6df9896e35aada563873b37		
risquesmicro2.txt	BOSSI P., F. KLOTZ et P. HAUSFATER. « Risques bactériologiques (Partie II) », EMC - médecine. Volume 2, Numéro 1, février 2005, Pages 107-125. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317983%232005%23999979998%23562019%23FLA%23Volume_2,_Issue_1,_Pages_1-125_(February_2005)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=b4d42c9a6df9896e35aada563873b37	2005-04-21	5488
toxoplas1.txt	BRENIER-PINCHART, M.-P. et H. PELLOUX « La toxoplasmose », Corpus médicale - Faculté de médecine de Grenoble, mai 2003. Cours universitaire. En ligne. Adresse URL : http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/parasitomyco/parasito/hp1/leconhp1.htm	2005-04-20	2112
tubercul1.txt	BLANIEA, M., J.L. PELLEGRINB et J. MAUGEIN. « Apport de la PCR dans le diagnostic des tuberculoses extrapulmonaires », Médecine et maladies infectieuses. Volume 35, numéro 1, janvier 2005, Pages 17-22. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=JournalURL&_cdi=6613&_auth=y&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=a335c59962dfdd92aacbc2af1663e69c	2005-04-17	3650
tubercul2.txt	BILLY C. et C. PERRONNE. « Aspects cliniques et thérapeutiques de la tuberculose chez l'enfant et l'adulte (Partie 1) », EMC - Maladies Infectieuses. Volume 1, numéro 2, Mai 2004, Pages 81-98. En ligne. Adresse URL : Ref. : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317984%232004%23999989997%23504265%23FLA%23Volume_1,_Issue_2,_Pages_67-127_(May_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=f5514bf978f662ca64bbf94c80a98c43	2005-04-17	4197
tubercul3.txt	BILLY C. et C. PERRONNE. « Aspects cliniques et thérapeutiques de la tuberculose chez l'enfant et l'adulte (Partie 2) », EMC - Maladies Infectieuses. Volume 1, numéro 2, Mai 2004, Pages 81-98. En ligne. Adresse URL : Ref. : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317984%232004%23999989997%23504265%23FLA%23Volume_1,_Issue_2,_Pages_67-127_(May_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=f5514bf978f662ca64bbf94c80a98c43	2005-04-17	5334
tubercul4.txt	ORGANISATION MONDIALE DE LA SANTÉ « Tuberculose », Aide-mémoire no 104, révisé mars 2004. En ligne. Adresse URL : http://www.who.int/mediacentre/factsheets/fs104/fr/	2005-04-20	1761

tubercul5.txt	DOMBRET M.-C. « Tuberculose pulmonaire chez l'adulte », EMC - médecine. Volume 1, Issue 5, Octobre 2004, Pages 406-416. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317983%232004%23999989994%23520480%23FLA%23Volume_1,_Issue_5,_Pages_365-485_(October_2004)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=c9ef219c64331a0486a015ebed06766b	2005-04-21	5064
tularemie.txt	GUIHOTA A., F. BRICAIREA, P. BOSSIA. « Tularémie », EMC - Maladies Infectieuses. Volume 2, Numéro 1, Mars 2005, Pages 1-10. En ligne. Adresse URL : http://www.sciencedirect.com/science?_ob=IssueURL&_tockey=%23TOC%2317984%232005%23999979998%23577661%23FLA%23Volume_2,_Issue_1,_Pages_1-58_(March_2005)&_auth=y&view=c&_acct=C000043357&_version=1&_urlVersion=0&_userid=789722&md5=7132be0b132a5f6d214525273a2ce6c2	2005-04-17	5336
tularemie2.txt	FAURE E. « La tularémie », Jeannin J. (éd.) Dossier santé. Réseaux et Systèmes d'Information Santé au service des professionnels, novembre 2001. En ligne. Adresse URL : http://www.caducee.net/DossierSpecialises/infection/tularemie.asp	2005-04-22	2125
urin.txt	PRACROS, J.P. Infection urinaire de l'enfant, janvier 1998. En ligne. Adresse URL : http://www.med.univ-rennes1.fr/cerf/edicerf/PEDIATRIE/6_INFECTIION_URINAIRE.html . (605268)	2005-01-15	3989
vaccin1.txt	STAHL, J.-P. (éd.) « Vaccinations : bases immunologiques, indications, efficacité, complications », Corpus médical - Faculté de médecine de Grenoble, Association des Professeurs de Pathologie Infectieuse et Tropicale, juin 2003. Cours universitaire. En ligne. Adresse URL : http://www-sante.ujf-grenoble.fr/SANTE/corpus/disciplines/malinf/malinf/76/lecon76.htm	2005-04-19	2102
virologie1.txt	LAFON M.E. (éd.) « Méthodologie de diagnostic virologique », Cours de virologie générale, Laboratoire de Virologie, Faculté de Médecine, Université Victor Segalen, Bordeaux. En ligne. Adresse URL : http://www.microbe-edu.org/glossaire/glossaire.html	2005-04-19	3155
virologie2.txt	VIRÉLIZIER J.-L. « Stratégie de survie des virus et défense de l'hôte », Cours de l'Unité d'immunologie virale, Institut Pasteur, DEA Jussieu 2001-2002. En ligne. Adresse URL : http://virologie.free.fr/07-Strategie_survie_virus/Strategie_survie_virus.htm	2005-04-22	7779

Total de mots

224583

Annexe B : Rapport de lemmatisation manuelle

Rapport de lemmatisation¹⁶

Adjectif lemmatisé	fréq. tot.	Fréquences par flexion				indice. sup.
		m.s.	m.p.	f.s.	f.p.	
polymicrobien	9		1	3	5	21
toxicogène	46	11	35			62
épithélioïde	28	7	21			48
extrapulmonaire	33	14	19			45
immunodéprimé	30	10	17 [†]	1	2	40
subaigu	17	2		13	2	32
mononucléé	12	2			10	30
asymptomatique	34	24	10			50
multinucléé	11		2	1	8	28
ostéoarticulaire	17	9	8			30
aérosolisé	10	1		2	7	26
microglial	8	1	7			26
multirésistant	17		1	8	8	26
bronchopulmonaire	7	1	6			24
axonal	8	2		6		22
membranaire	19	13	6			37
caliciel	11	2		5	4	21
tuberculoïde	8	3	5			21
nosocomial	22		1	10	11	29
autoimmun	7			3	4	18
fungique	5	1	4			18
multifocal	1				1	18
oropharyngé	7	2		4	1	18
séquellaire	6	2	4			18
abdominopelvien	4		1		3	15
aérotolérant	4		1		3	15
bacillifère	10	7	3			26
biliopancréatique	6	3	3			même
extraréspiratoire	4	1	3			15
hyperintense	4	1	3			15
immunodominant	9	5	3	1		21
interfolliculaire	4	1	3			15
périportal	7		3	2	2	15
typable	4	1	3			15
délectable	12	7	5			19
antibotulique	4	2	2			même
cytopathogène	9	7	2			26
fibrinolytique	3	1	2			11
gangréneux	3			2	1	11
gigantocellulaire	6	4	2			18

¹⁶ Puisqu'il s'agit de lemmatiser la forme des unités lexicales dans le but d'éliminer les doublons pouvant se trouver dans la liste de candidats adjectifs, les formes épécènes ont été traitées comme des flexions masculines.

immunoenzymatique	3	1	2			11
intracytoplasmique	3	2	1			11
locorégional	4	2	2			même
macrophagique	4	2	2			même
malpighien	4	1	2		1	11
médicochirurgical	4	1		2	1	11
monomicrobien	3	1		2		11
multifactoriel	4		1	1	2	11
neuroméningé	3			1	2	11
monoclonal	4	1	3			10
invasif	5	3			2	15
bronchoalvéolaire	5	4	1			18
chondral	3	2		1		11
cytolytique	4	3	1			15
entérotoxigène	3	2	1			11
extradigestif	3	1	1	1		même
histopathologique	2	1	1			même
hypersécrétoire	2	1	1			même
immunodéprimant	2		1		1	même
immunopathologique	2	1	1			même
immunorégulateur	3	1		1	1	même
intracanaulaire	2	1	1			5
intraparenchymateux	4	2		1	1	11
lymphomateux	2			1	1	même
lymphotrope	2	1	1			même
micronodulaire	2	1	1			même
mucopurulent	2	1		1		même
oculoganglionnaire	4	3	1			15
paracentral	2			1	1	même
paracortical	3		1	1	1	même
phlycténulaire	4	3	1			13
pleuropulmonaire	2	1	1			même
pléiomorphe	2	1	1			même
péribronchique	2	1	1			même
périodontal	3	1		1	1	même
périprothétique	5	4	1			18
recombinant	2		1		1	même
rétropéritonéal	3	2		1		11
rétrosternal	2			1	1	même
rétroviral	2	1			1	même
saccharolytique	2	1	1			même
sarcoïdique	2	1	1			même
trachéobronchique	2	1	1			même
transbronchique	2	1	1			même
transcriptionnel	2		1	1		même
ulcéroganglionnaire	8	7	1			26
épidermoïde	3	2	1			11
nucléotidique	7	6	1			22

Annexe C : Système de paraphrases

Système de paraphrases

NomT	Paraphrase	Adjectif	NomB
<p>(toxi-)infection (résultat), maladie, fièvre, inflammation, affection, atteinte, lésion, complication, syndrome, problème, phénomène, pathologie, anévrysme, hémolyse, infiltrat, œdème, granulome, exsudat, tumeur, fibrose, abcès, cicatrice, choc, ostéoartrite, pleurésie, cholangite, choc, accident vasculaire cérébral (AVC), intoxication, arthrite, dermatohypodermite, thrombose, embolie, thrombophlébite, ostéophlébite, pseudoarthrose, péricardite, complication, duodénite, épidémie, diffusion</p>	<p>NomT qui est causé par NomB</p>	<p>viral, bactérien, inflammatoire, tuberculeux, infectieux, hémorragique, botulique, polymicrobien, septique, parasitaire, neurologique</p>	<p>virus, bactérie, infection, inflammation, tuberculose, hémorragie, <i>Clostridium botulinum</i>, microbe, sepsie, parasite, système nerveux</p>
<p>agent, dose, particule (virale), vecteur, virion, spore, syndrome, lésion, maladie, bacille, mycobactérie, bactérie, virus, espèce, forme, prion, toxine, infection, souche, complexe, locus, chirurgie, manœuvre, injection, organisme, activité, colonisation, germe, streptocoque, méningite, myalgie</p>	<p>NomT qui provoque NomB</p>	<p>infectieux, inflammatoire, tuberculeux, pathogène, grippal, botulique, protéolytique, asymptomatique, pyogène, épidémique</p>	<p>infection, inflammation, tuberculose, maladie, grippe, botulisme, sepsie, protéolyse, symptôme (nég.), (production de) pus, épidémie</p>
<p>maladie (tuberculose, etc.), atteinte, infection, pathologie, forme, ascariose, agression, complication, affection, colique, cirrhose, infarctus, fibrose, insuffisance, syndrome, problème, protozoose, atrophie (villositaire), atonie, helminthiase, métaplasie, biharziose, pneumopathie (interstitielle), phénomène, phase, empyème, trouble, lésion, défaillance, sécheresse, réaction inflammation, néphrite, pneumopathie, œdème, vibrance, néoplasie, (neuro-)immunologie, immunité, vaccin, stratégie, virus, maladie épidémie, viroses, agent infectieux (bactérie), épidémiologie [des maladies], microbiologie, contamination, bactériologie, problème, entité, pathologie, ligand, parasitisme, effecteur, phénomène, grippe, souche, botulisme, détresse, virus, syndrome, trouble, défaillance, difficulté, parasite</p>	<p>NomT qui affecte NomB</p>	<p>pulmonaire, abdominal, cellulaire, cutané, urinaire, pancréatique, encéphalique, hépatique, osseux, méningé, rénal, intestinal, lymphoïde, ganglionnaire, biliaire, vasculaire, pleural, neurologique, épithélial, parenchymateux, muqueux, rachidien, épithélioïde, interstitiel, épidual, extrapulmonaire, péricardique, viral, animal, clinique, aviaire, respiratoire, digestif</p>	<p>poumon, abdomen, cellule, peau, système urinaire, pancréas, encéphale, foie, os, méninges, rein, intestin, système lymphatique, ganglion, voie biliaire, système vasculaire, plèvre, système nerveux, épithélium, parenchyme, muqueuse, colonne vertébrale (rachis), cellule épithéliale, tissu interstitiel, (surface externe de la) dure-mère, péricarde, virus, animal, malade, oiseau, respiration, digestion</p>
<p>sujet, personne, hôte</p>	<p>NomT qui subit NomB</p>	<p>infectieux</p>	<p>infection</p>
<p>sujet, personne, hôte, transplanté</p>	<p>NomT qui a reçu NomB</p>	<p>immun, rénal</p>	<p>immunisation, rein</p>

protéine, génome, gène, peptide, acide nucléique, ARN, glycoprotéine, molécule, élément, enveloppe, déterminant génique, composant, spore, structure membranaire, corps, paroi, membrane, constituant, élément, débris, protéine, récepteur, molécule, secteur, glycoprotéine, fraction, effecteur, gamète, souche, préparation, parenchyme, alvéole, muscle, cellule, muqueuses, épithélium, capillaire, capillaire, surface, segment, (sous-) muqueuse, lumière, glande, canal, acini, structure, moëlle, tissu, fenêtré, matrice, glomérule, artère, arbre artériel, pôle, lithiase, sel, pigment, paroi, cavité, canal, fragment	NomT qui est une partie de NomB	viral, bactérien, cellulaire, antigénique, lymphocytaire, parasitaire, vaccinal, pulmonaire, respiratoire, cutané, digestif, pancréatique, hépatique, osseux, rénal, intestinal, ganglionnaire, biliaire, vasculaire, vasculaire, tissulaire, épithélial, rachidien	virus, bactérie, cellule, antigène, lymphocyte, parasite, vaccin, poumon, voies respiratoires, peau, tractus digestif, pancréas, foie, os, rein, intestin, ganglion, bile, vaisseau, système vasculaire, tissu, épithélium, colonne vertébrale (rachis)
antigène	NomT qui est présent dans NomB	urinaire	urine
suspension, charge, bofilim, produit, aérosol, chambre, atmosphère, milieu, jarre, flacon, vaccin, rappel (de vaccin), liquide (céphalo-rachidien (LCR), infiltrat, pus, prélèvement, abcès, charge, diarrhée, matière	NomT qui contient NomB	viral, bactérien, anaérobie, cellulaire, antigénique, hémorragique, lymphocytaire, polymicrobien, parasitaire, purulent	virus, bactérie, air(anti), cellule, antigène, sang, lymphocyte, microbe, parasite, pus
caverne, séquelleaire	NomT qui a la forme de NomB	tuberculeux	tubercule
spondylite, spondylite, poumon, réaction inflammatoire, chorioméningite, fièvre, nécrose, maladie, diarrhée, infarctus, diathèse, stigmata, desquamation, inflammation, méningite, pleurésie, caractère, conjonctivite, ulcération, diarrhée	NomT qui est caractérisé par la présence de NomB	tuberculeux, lymphocytaire, hémorragique, purulent, muqueux, asymptomatique	tubercule, lymphocyte, hémorragie, pus, mucus, symptôme (nég.)
visée, contrainte, aspect	NomT dont le caractère est observé en NomB	microbiologique, histologique	microbiologie, histologie
tropisme, phase, développement, réplication, inclusion, persistance, hémorragie	NomT qui se produit dans/sur NomB	digestif, intracellulaire, interstitiel	digestion, cellule, interstice
transmission, contamination, contagion	NomT qui se produit de NomB	interhumain	humain
antigène, hypersécrétion, intolérance, manifestation, colonisation	NomT qui est produit lors de NomB	tuberculeux, digestif	tuberculose, digestion

voie, tractus, appareil, sphère, tube, appareil, follicule, organe, tissu	NomT où se produit NomB	respiratoire, digestif, urinaire, lymphoïde	respiration, digestion, lymphocyte
souche, statut, agent pathogène (C. Difficile, etc), corps	NomT qui produit NomB	toxigène, muqueux	toxine, mucus
mécanisme, voie, arme	NomT qui utilise NomB	hématogène, bactériologique	sang, bactériologie
technique, prélèvement, méthode, démarche, moyen, test, critère, outil, geste, stratégie, élément, situation, apport, recours, bandelette réactive	NomT qui est utilisé en/dans NomB	bactériologique, sérologique, microbiologique, diagnostique, urinaire	bactériologie, sérologie, microbiologie, diagnostic, urine
diagnostic, traitement, orientation, recherche	NomT qui vise NomB	étiologique	étiologie
diagnostic, examen, étude, analyse, enquête, test, suivi, dépistage, différenciation, identification, investigation	NomT qui fait appel à NomB	bactériologique, sérologique, microbiologique, histologique	bactériologie, sérologie, microbiologie, histologie
voie, artéfact, lamelle, infiltrat, réseau, paroi, tissu, barrière, réservoir, origine, milieu, modèle, particule, matériel, agent, souche, population, famille, sous-type, espèce, colonie, communauté, genre, groupe, lignée, type, sous-population, répertoire, microorganisme, amas, écoulement	NomT qui est constitué de NomB	viral, pulmonaire, cutané, osseux, lymphoïde, vasculaire, endothélial, épithélial, intraveineux, muqueux, animal, purulent	virus, poumon, peau, os, lymphé, vaisseau, endothélium, épithélium, veine, muqueuse, animal, pus
origine, source, suc, isoamylase, chymotrypsine, enzyme, ostéomyélite, infection, mucus, suppuration, épanchement, liquide, sécrétion, toxine, anatoxine, neurotoxine, clone, antigène	NomT qui provient de NomB	cutané, pancréatique, hépatique, hématogène, intestinal, ganglionnaire, pleural, botulique, parasitaire	peau, pancréas, foie, sang, intestin, ganglion, plèvre, <i>Clostridium botulinum</i> , parasite
produit, farine	NomT fait à partir de NomB	animal	animal
foyer, site, effraction, poche, plage, épisode, nodule, vésicule, localisation, compartiment, cavité, zone	NomT où se situe NomB	infectieux, inflammatoire, hémorragique, lymphoïde, biliaire, septique, vasculaire, pleural, purulent, épithélial	infection, inflammation, hémorragie, lymphé, bile, sepsie, vaisseau, plèvre, pus, épithélium
voie, canalicule, canal, vaisseau, système	NomT où circule NomB	biliaire, lymphatique	bile, lymphé

abcès, lésion, carcinome, sécrétion, source, fibrose, cavité, foyer (purulent), hôte, gouttelette, éruption, zone, plaie, verrue, effraction, manifestation, porte d'entrée, fistule, cystadénocarcinome, tumeur, masse, kyste, dommage, brèche, contenu, flore, cellule, noyau, oxygène, dépôt, débris, étage, membre, rachis (syn. colonne vertébrale), épanchement, brûlure, traumatisme, nécrose, infiltrat, noyau, plage, anomalie, colonie, rachis, uretère, région, bactérie, bacille, parasite, organisme (pathogène), cible, tissu, espace, substance, milieu, graisse, site, localisation	NomT qui se situe dans/ sur NomB	pulmonaire, respiratoire, urinaire, encéphalique, hépatique, rénal, intestinal, endothélial, tissulaire, thoracique, parenchymateux, muqueux, intracellulaire, lymphatique, interstitiel, extrapulmonaire, péricardique	poumon, abdomen, voies respiratoires, cutané, tractus digestif, système urinaire, pancréas, encéphale, foie, rein, intestin, endothélium, tissu, thorax, parenchyme, muqueuse, lombes, cellule, (en périphérie d'un) vaisseau lymphatique, interstice, (surface externe de la) dure-mère, péricarde
signe, embolie, râle, oedème, forme, présentation, infarctus, diffusion, symptomatologie, douleur, algie, réaction, symptôme, séquelle, déficit, nécrose, infarctus, lésion, aideur, pathologie, agression, immunologie, installation	NomT qui se manifeste dans/sur/chez NomB	pulmonaire, abdominal, osseux, méningé, pleural, neurologique, tissulaire, épithélial, thoracique, rachidien, lombaire, clinique	poumon, abdomen, os, méninges, plèvre, système nerveux, tissu, épithélium, thorax, colonne vertébrale, lombes, malade
réaction, réponse, état, infiltration, phénomène, privilège, fonction	NomT qui se manifeste par NomB	inflammatoire, immun	inflammation, immunisation
symptôme, signe	NomT qui se manifeste lors de NomB	respiratoire, digestif	respiration, digestion
porteur, cas, sujet, patient, nourrisson	NomT qui manifeste NomB	asymptomatique	symptôme (nég.)
symptôme, activité	NomT qui est observé lors de NomB	grippal	grippe
variation, glissement, dérive, changement, signe, présentation, caractéristique, aspect, symptôme, expression, contexte, symptomatologie, paramètre, lésion, critère, arguments, élément, manifestation, situation, cas, amélioration, phase, évolution, état, réponse	NomT qui est observé sur/chez NomB	antigénique, clinique	antigène, malade
zone	NomT où est observé NomB	épidémique	épidémie
protéase, enzyme, antigène, protéine, niche	NomT qui participe à NomB	digestif, structural	digestion, structure
défense	NomT qui est possible au moyen de NomB	immunitaire	immunité

	NomT au moyen de NomB	vaccinal	vaccination
prévention, prophylaxie		clinique	malade
diagnostic, orientation	NomT qui est possible au moyen d'observations effectuées chez NomB		
radiographie, prélèvement, examen, suivi, investigation, surveillance, échographie, poussée, biopsie, test, chirurgie, endoscopie, bilan, scintigraphie, exploration (isotopique), ponction, tomographie par émission de positrons (TEP), drainage, coupe, scanner, imagerie, drainage, injection, traitement, urographie, antibiothérapie, ensemençement, ponction, collection, sérologie, PCR (réaction en chaîne de la polymérase), examen	NomT qui est effectué sur/dans NomB	pulmonaire, abdominal, clinique, respiratoire, cutané, digestif, hépatique, osseux, méningé, rénal, intestinal, ganglionnaire, biliaire, pleural, neurologique, tissulaire, thoracique, intraveineux, parenchymateux, lombaire, épidual, extrapulmonaire, péricardique, viral, clinique	poumon, abdomen, malade, voies respiratoires, peau, tractus digestif, foie, os, méninges, rein, intestin, ganglion, voie biliaire, plèvre, système nerveux, tissu, thoracique, parenchyme, lombes, (surface externe de la) dure-mère, péricarde, virus, clinique
échantillon	NomT qui est prélevé sur NomB	clinique	malade
microorganisme	NomT qui est un/une NomB	bactérien	bactérie
filtre, barrière	NomT que procure NomB	pulmonaire, hépatique, intestinal	poumon, foie, intestin
cellule, lymphocyte, complexe	NomT qui procure NomB	immun	immunisation
facteur	NomT qui annonce la présence d'un/une NomB	bactérien	bactérie
griffure, agression, déjection, morsure	NomT qui est fait par un NomB	animal	animal
vaccin, (médicament)	NomT qui est destiné à le/la NomB	animal	animal
données, renseignement	NomT qui est obtenu auprès de NomB	clinique	malade
résultat, preuve, documentation, diagnostic, données	NomT qui est obtenu au moyen de NomB	batériologique, histologique	bactériologie, sérologie, histologie
caractère, risque, orientation, réaction, forme, négativité, différenciation, caractéristique, lésion	NomT qui est étudié en NomB	batériologique, histologique	bactériologie, sérologie, microbiologie, histologie

maladie, infection, atteinte, milliaire, complication, pneumopathie atypique	NomT qui est secondaire à NomB	tuberculeux, grippal	tuberculose, grippe
cellule, granulome, nécrose, mode, allure	NomT qui ressemble à NomB	épithélioïde, épidémiologique éosinophile,	cellule épithélioïde, éosine, épidémie
cytoplasme, sécrétion, exsudat, granulome, contenu, substance	NomT qui possède une affinité pour NomB	éosinophile	éosine
agent	NomT qui est NomB	étiologique	étiologie
cellule, tissu, élément, système, réponse, réaction, effecteur	NomT qui participe à NomB	inflammatoire, immunitaire	inflammation, immunité
mère	NomT qui souffre de NomB	tuberculeux	tuberculose
bactérie, culture, agent, germe	NomT qui nécessite NomB	anaérobie	air (nég.)
protéine, agent, anticorps, lymphocyte, molécule, médicament, antibiotique, substance	NomT qui combat NomB	antituberculeux, antiviral	tuberculose (anti-), virus
traitement	NomT qui vise NomB	urinaire	système urinaire
sondage	NomT qui permet de prélever NomB	urinaire	urine
argument	NomT qui permet d'identifier NomB	étiologique	étiologie
immunité, phase, médiation	NomT auquel participe NomB	cellulaire, lymphatique	cellule, lymphocyte, lympho
phase, contexte, période	NomT au cours duquel/de laquelle se déroule NomB	épidémique	épidémie
substance (quinine, quinolone), drogue	NomT qui est administré dans NomB	intraveineux	veine

<p>motricité, hyperinfection, stase, transplantation, invasion, hypertrophie, altération, greffe, destruction, réparation, réponse, remaniement, hypervascularité, contamination, inflammation, épaississement, détérioration, croissance, péristaltisme, tractus, déplétion, hyperplasie, gonflement, ulcération, remplissage, obstruction, piqure, connexion, néogénèse, réaction, perturbation, frottement, turgescence, hydratation, vieillissement, oxygénation, collection, prélèvement, déformation, extension, décollement, calcification, processus, métabolisme, clone, adhésion, réplication, éradication, transmission, prolifération, flambée, diffusion, menace, poussée, prélèvement, rétention, excrétion, circulation, soin, protection, neutralisation, dissémination, multiplication, maturation, tropisme, transcription, réactivation, émergence, survie, culture, colonisation, inoculum, identification, lyse, développement, manifestation, coloration, mort, infiltration, recoloration, abrasion, colonisation (fongique), contact, surveillance, effondrement, restitution, élimination, dépression, persistance, système, alimentation, épanchement, infiltration</p>	<p>NomT de(1) NomB</p>	<p>viral, bactérien, infectieux, cellulaire, cutané, immunitaire, infectieux, digestif, urinaire, hémorragique, respiratoire, antiparasitaire, hépatique, pancréatique, osseux, rénal, intestinal, ganglionnaire, vasculaire, endothélial, parasitaire, tissulaire, purulent, lymphomateux, rachidien, épipharyngé, épidermique, épidual, péricardique</p>	<p>virus, bactérie, infection, cellule, peau, immunité, appareil digestif, urine, voie urinaire, sang, appareil respiratoire, antigène, pancréas, foie, lymphocyte, os, rein, intestin, ganglion, vaisseau, système vasculaire, endothélium, parasite, tissu, pus, parenchyme, colonne vertébrale, lymphome, épidémie, surface de la dure-mère, péricarde</p>
<p>motilité, sémiologie (radiologique), hyperperméabilité, morphologie, stade, histologie, toxicité, contour, homéostasie, extrémité, physiologie, image, tableau, fonction, perméabilité, amas, localisation, territoire, aire, richesse, continuité, spécificité, andulation, instabilité, opacité, écho, plan, efficacité, performance, intérêt, valeur, certitude, orientation, supériorité, difficulté, rentabilité, bilan, particularité, rythme, couverture, stratégie, schéma, statut, obligation, composition, activité métabolique, anatomie, tableau, symptomatologie, état, excédent, structure, forme, cycle, stade du développement, débit, origine, sérotype, étiologie, antigénicité, épidémiologie, variabilité, nomenclature, résistance, caractère, cause, signe, nature, lignée, type, cytotoxicité, facteur, mécanisme, matrice, variante, analyse, propriété, plasticité, caractéristique, hétérogénéité, déterminant, similitude, endémie, incidence, pandémie, saison, déficit, capacité, base (fondement), risque, coalescence, diversité, tableau [des observations], observation, description, bilan, plan, gravité, sévérité, complication</p>	<p>NomR de (2) NomB</p>	<p>viral, bactérien, infectieux, cellulaire, tuberculeux, grippal, immunitaire, hémorragique, bactérien, respiratoire, pulmonaire, antiparasitaire, hépatique, clinique, inflammatoire, pathogène, digestif, urinaire, pancréatique, hépatique, lymphocytaire, osseux, bactériologique, rénal, intestinal, ganglionnaire, biliaire, vasculaire, endothélial, parasitaire, tissulaire, rachidien, diagnostique, vaccinal, structural, épidémiologique, péricardique</p>	<p>virus, bactérie, infection, cellule, tuberculose, grippe, immunité, hémorragie, infection, bactérie, respiration, poumon, antigène, malade, inflammation, pathologie, appareil digestif, urine, pancréas, foie, lymphocyte, os, bactériologie, rein, intestin, ganglion, voie biliaire, vaisseau, système vasculaire, endothélium, parasite, tissu, colonne vertébrale, diagnostic, vaccin, vaccination, structure, épidémie, étiologie, péricarde</p>

carcasse, restes, série	NomR de (3) NomB	animal, vaccinal	animal, vaccin
tolérance, échappement	NomT à NomB	immunitaire, vaccinal	immunité, vaccination
assistance, isolement	NomT au moment de NomB	respiratoire	respiration
infestation, inoculation, contamination, inclusion, infection (action), réinfestation, diffusion, dissémination, essaimage	NomT par NomB	viral, parasitaire, osseux, hématogène	virus, parasite, os, sang
localisation, niveau?, champ, manifestation, hémorragie (action), diffusion, duplication, hypersécrétion, intégration, dissémination, analogie, similitude, délai, retard, pivot, différence, effet, risque, suspicion, présomption, argument, doute, migration, pli, manifestation, localisation, dissémination, échec, efficacité	NomT dans/sur NomB	pulmonaire, pancréatique, hépatique, méningé, rénal, intestinal, tissulaire, lymphatique, clinique, diagnostique, structural, cutané, neurologique, épidermal	poumon, pancréas, foie, méninges, rein, intestin, tissu, lymph, malade, diagnostic, structure, peau, système nerveux, (surface externe de la) dure-mère
interaction, jonction, message, contact	NomT entre les NomB	cellulaire, interhumain	cellule, humain
vaccin, traitement, lutte, immunité, défense, thérapie, réponse	NomT contre NomB	grippal, antituberculeux, antiviral	grippe, tuberculose (anti-), virus
concentration, déficit, composition, prédominance, expertise	NomT en NomB	bactérien, lymphocytaire, microbiologique, cellulaire, osseux,	bactérie, cellule, lymphocyte, os, microbiologie

Annexe D : Liste des adjectifs analysés

Liste des adjectifs analysés

#	Forme canonique	fréquence	Indice de spécificité	Base	nbre acceptations
1	viral	420	197.113224082514	virus	10
2	infectieux	237	143.224116900643	infection	7
3	clinique	279	138.109692767387	malade	10
4	bactérien	180	131.41799843362	bactérie	10
5	inflammatoire	136	116.405140718082	inflammation	6
6	pulmonaire	140	113.088471450809	poumon	8
7	anaérobie	104	108.858617719297	air	2
8	cellulaire	173	103.016404286331	cellule	8
9	immunitaire	133	102.40307453405	immunité	5
10	tuberculeux	98	101.513146070349	tuberculose, tubercule	8
11	pathogène	128	99.5541128438415	maladie	2
12	aigu	178	92.3655320292529	Qualificatif	
13	respiratoire	123	87.0004447171984	respiration, voies respiratoires	9
14	cutané	86	86.0691184681168	peau	8
15	grippal	69	85.7724146296361	grippe	5
16	antituberculeux	64	84.8789590944516	tuberculose	2
17	digestif	98	84.5502937362455	digestion, tractus digestif	11
18	antigénique	61	83.4892110183993	antigène	6
19	hémorragique	66	80.3369515330299	hémorragie, sang	6
20	botulique	61	80.207853953276	<i>Clostridium botulinum</i> , botulisme	6
21	immun	55	78.4821629421348	immunisation	3
22	urinaire	91	77.6187431834214	urine, appareil urinaire	11
23	pancréatique	53	76.2731913765522	pancréas	7
24	hépatique	63	75.6236320127037	foie	9
25	aviaire	50	73.9582736009471	oiseaux	2
26	lymphocytaire	49	73.9115903238401	lymphocyte	7
27	spécifique	240	73.7607941954727	Qualificatif	
28	antiviral	49	72.450766099773	virus	2
29	osseux	71	70.1669082130634	os	9
30	hématogène	42	69.0159063258528	sang	3
31	méningé	42	69.0159063258528	méninges	6
32	polymicrobien	9	21.6713337602186	microbe	2
33	bactériologique	74	66.6861111269973	bactériologie	6
34	chronique	138	66.1369496422677	Qualificatif	
35	rénal	66	64.2688866303181	rein	8
36	intestinal	50	63.5461255372159	intestin	9
37	toxinogène	46	62.8494826796184	toxine	2
38	lymphoïde	38	62.3063093990416	lymphe, système lymphatique, lymphocyte	6
39	ganglionnaire	40	62.0032256003039	ganglion	6
40	biliaire	34	61.0119840887086	bile, voie biliaire	6
41	sérologique	39	60.4175578615615	sérologie	4
42	cytotoxique	33	60.054167459793	Qualificatif	
43	septique	37	59.9057483521733	septie	3
44	vasculaire	47	59.6482002378728	vaisseau, système vasculaire	10
45	pleural	34	59.3098605614959	plèvre	6

46	fréquent	148	58.3066801002566	Qualificatif	
47	endothélial	29	57.0362736583504	endothélium	3
48	interhumain	30	55.2962168811255	humain	2
49	parasitaire	39	55.2394322458615	parasite	7
50	neurologique	48	55.1593105451357	système nerveux	5
51	tissulaire	35	54.6730595555293	tissu	7
52	purulent	30	54.4611931276549	pus	5
53	épithélial	30	52.8961046334368	épithélium	4
54	thoracique	36	52.8531685845102	thorax	3
55	intraveineux	37	52.8201153094207	veine	3
56	parenchymateux	25	52.8069053939305	parenchyme	4
57	muqueux	30	52.1612740074196	mucus, muqueuse	5
58	secondaire	136	52.150499007508	Numéral	
59	rachidien	25	51.7633911192003	colonne vertébral (rachis)	5
60	protéolytique	24	51.6956086265784	protéolyse	1
61	asymptomatique	34	50.6327480846761	symptôme	3
62	résistant	78	50.3569603056309	Résister	
63	lombaire	31	49.9590690810137	lombe	3
64	diagnostique	44	49.5462028039547	diagnostic	4
65	pyogène	22	49.398206345453	pus	1
66	intracellulaire	21	48.2085491793009	cellule	2
67	épithélioïde	28	48.2085491793009	cellule épithéliale	2
68	neutralisant	25	48.1219178420461	Neutraliser	
69	vaccinal	34	47.9760789032762	vaccin, vaccination	6
70	lymphatique	28	47.9284941962741	lymphe, système lymphatique	5
71	structural	32	47.2928409029494	structure	3
72	structural	32	47.2928409029494	structurer	
73	épidémique	29	47.2097163137163	épidémie	6
74	interstitiel	20	46.988850165102	interstice	2
75	épidural	21	46.0265028502713	surface externe de la dure-mère	5
76	étiologique	20	45.836349224004	étiologie	4
77	extrapulmonaires	33	45.7367169507404	poumon	3
78	abdominal	35	45.4171969722725	abdomen	4
79	microbiologique	22	45.3675810879689	microbiologie	5
80	précoce	81	45.1943782784131	Qualificatif	
81	histologique	20	44.7628876073049	histologie	4
82	éosinophile	18	44.4494225026337	éosinophile	2
83	péricardique	18	44.4494225026337	péricarde	5
84	septicémique	18	44.4494225026337	septicémie	2
85	encéphalique	5	43.7597872349182	encéphale	2

autre type