



ASp
la revue du GERAS

51-52 | 2007
Varia

Prémodification et coordination : quelques problèmes de traduction des groupes nominaux complexes en anglais médical

François Maniez



Édition électronique

URL : <http://journals.openedition.org/asp/500>
DOI : 10.4000/asp.500
ISBN : 978-2-8218-0404-3
ISSN : 2108-6354

Éditeur

Groupe d'étude et de recherche en anglais de spécialité

Édition imprimée

Date de publication : 1 décembre 2007
Pagination : 71-94
ISSN : 1246-8185

Référence électronique

François Maniez, « Prémodification et coordination : quelques problèmes de traduction des groupes nominaux complexes en anglais médical », *ASp* [En ligne], 51-52 | 2007, mis en ligne le 01 décembre 2010, consulté le 22 mars 2021. URL : <http://journals.openedition.org/asp/500> ; DOI : <https://doi.org/10.4000/asp.500>

Ce document a été généré automatiquement le 22 mars 2021.

Tous droits réservés

Prémodification et coordination : quelques problèmes de traduction des groupes nominaux complexes en anglais médical

François Maniez

Introduction

- 1 Les langues de spécialité abondent en groupes nominaux complexes, et le domaine médical ne fait pas exception à cette règle (Maniez 2000). La conjugaison de la prémodification et de la coordination à l'intérieur de ces groupes nominaux complexes demeure encore à l'heure actuelle un obstacle majeur pour les systèmes de traduction automatique confrontés à la traduction de textes scientifiques (Maniez 2001). Quant à l'humain apprenant de langue étrangère de spécialité, on sait qu'il évite volontiers le recours à l'outil de décodage classique qu'est le dictionnaire bilingue lorsqu'il a le sentiment de maîtriser les constituants des groupes complexes (Cormier 1990, Thoiron 2000). Il est donc utile de sensibiliser l'apprenant aux décalages entre langue-source et langue-cible (et en particulier à la présence de faux-amis en terminologie) et de réfléchir à l'intérêt d'un enseignement dont le but serait la maîtrise des lexies complexes et des collocations en langue de spécialité, dont le décodage est souvent susceptible de devenir problématique.
- 2 Le développement suivant étudie les problèmes de compréhension et de traduction posés par les différents patrons syntaxiques des groupes nominaux de l'anglais. Les structures étudiées le seront dans un ordre de complexité croissante.

1. La prémodification nominale simple

- 3 On sait que la majorité des termes employés en anglais de spécialité sont de longueur 2 (Frantzi *et al.* 1999), et sont généralement formés selon un modèle comprenant un élément modificateur (adjectif ou nom) qui précède le nom constituant le « nœud » du terme.
- 4 Les groupes nominaux correspondant à ce schéma syntaxique sont à l'origine d'un certain nombre de difficultés pour l'apprenant. La première concerne la compréhension du lien exact entre les deux noms qui sont ainsi accolés. Rouleau (2003) résume ainsi le problème posé par la prémodification du nom par un autre nom dans son commentaire à propos de la traduction de la séquence *resin uptake* dans le groupe nominal coordonné *T3 resin uptake or direct thyroxine assays*.
- Poser la question, c'est mettre le doigt sur une des difficultés majeures en terminologie (anglais-français) : savoir débusquer le rapport qui existe entre les éléments de la chaîne nominale du syntagme. Ce manque de précision concernant la relation entre les termes, causé par l'absence d'articulateur(s), est caractéristique de l'anglais, qui n'hésite pas un seul instant à accoler deux mots, et parfois plus, et à sous-entendre les rapports qui les lient. (Rouleau 2003 : 144)
- 5 Le passage d'une langue utilisant la prémodification (comme les langues germaniques) à une langue utilisant la post-modification (langues romanes) par le biais d'un groupe prépositionnel nécessite en effet l'explicitation de cette relation. Le problème est toutefois compliqué par l'utilisation des adjectifs dits « relationnels », qui peuvent également traduire cette prémodification.
- 6 Les exemples tirés du corpus CLASS¹ (Tableau 1) montrent en effet que l'allemand suit presque systématiquement la syntaxe du groupe nominal anglais (à l'exception du sixième exemple, où *night sweats* est traduit par une expression contenant un adjectif, *nächtliche Schweißausbrüche*). Certaines expressions sont traduites à l'aide d'un complément prépositionnel dans la totalité ou la majorité des langues romanes (*stomach content*, *weight loss*), mais le recours à l'adjectif relationnel semble être la norme dans tous les autres cas, le recours à la complémentation restant relativement rare (l'italien *dolori al torace* pour traduire *chest pain* ou le portugais *folículos do cabelo* pour traduire *hair follicles*).
- 7 Les relations entre les deux noms concernés par la prémodification peuvent être de natures multiples. Le prémodificateur peut signifier la localisation anatomique (*back pain* : douleur dorsale ou dorsalgie, *brain stem* : tronc cérébral), la fonction (*taste buds* : papilles gustatives, *sweat glands* : glandes sudoripares), la cause (*heat rash* : érythème calorique) ou la forme (*sickle cell* : cellule falciforme²). La compréhension de cette relation de prémodification nécessite donc une connaissance plus précise de la réalité extra-linguistique en anglais qu'en français, où l'adjectif relationnel qui traduit la prémodification contient parfois d'autres éléments lexicaux que le simple suffixe adjectival, comme c'est le cas pour les adjectifs « sudoripares » et « falciforme » dans les exemples cités ci-dessus.

Tableau 1. Traduction de quelques groupes nominaux de type N-N dans le corpus CLASS

Anglais	Allemand	Français	Espagnol	Italien	Portugais
---------	----------	----------	----------	---------	-----------

cell culture	Zellkultur	culture cellulaire	cultivos celulares	coltura cellulare	culturas celulares
chest pain	Brustschmerzen	douleurs thoraciques	dolor torácico	dolori al torace	dor torácica
hair follicles	Haarfollikeln	follicules pileux	folículos pilosos	follicoli piliferi	folículos do cabelo
lung tissue	Lungengewebe	tissus pulmonaires	material pulmonar	tessuto polmonare	tecido pulmonar
muscle biopsy	Muskelbiopsie	biopsie musculaire	biopsia muscular	biopsia muscolare	biópsia muscular
night sweats	nächtliche Schweißausbrüche	sueurs nocturnes	sudores nocturnos	sudori notturni	suores nocturnos
skin lesion	Hautläsion	lésion cutanée	lesión cutánea	lesione cutanea	lesão cutânea
stomach content	Mageninhalt	contenu de l'estomac	contenido del estómago	contenuto dello stomaco	conteúdo do estômago
weight loss	Gewichtsverlust	perte de poids	adelgazamiento	perdita di peso	perda de peso

- 8 Une autre source de difficulté de compréhension pour l'apprenant réside dans la grande souplesse des mécanismes de conversion (Tournier 1985), qui se traduisent diachroniquement par la possibilité pour une unité lexicale d'appartenir à plusieurs parties du discours. Ce phénomène est connu pour concerner en particulier le verbe et le nom en anglais, mais on le voit également à l'œuvre dans le cas de l'adjectif et du nom. Ainsi, l'adjectif prémodificateur *twin* dans les lexies *twin town* ou *twin beds* peut-il être traduit par l'adjectif *jumeau* (ville jumelle, lits jumeaux) car il présuppose effectivement l'existence de deux entités correspondant au nom prémodifié. Ce n'est toutefois pas le cas du terme *twin pregnancy*, pour lequel le prémodificateur est le nom *twin*, qui donnera lieu à une traduction française utilisant l'adjectif relationnel (grossesse gémellaire), signifiant une caractéristique de la grossesse qui détermine l'appartenance à une classe spécifique.
- 9 Les difficultés posées par la prémodification ne se limitent pas au seul décodage de la relation sémantique entre prémodifiant et modifié. Le processus de traduction des unités lexicales correspondant à ce patron syntaxique peut en effet donner lieu à au moins deux structures distinctes : la post-modification adjectivale et la post-modification par un groupe prépositionnel. Rouleau (2003) explique que dans un certain nombre de cas, le choix entre ces deux structures n'est que théorique, car l'usage consacre en général le recours à l'une des deux solutions, comme c'est le cas pour la traduction de l'adjectif *myocardial* dans le terme *myocardial infarction* :

[Le traducteur] devra par contre utiliser une seule des deux formes dans les deux autres cas : « infarctus pulmonaire » (adjectif) et « infarctus du myocarde » (groupe prépositionnel). Il ne faudrait pas conclure, d'après ce dernier cas, que

« myocardique » est banni du discours médical. Il est tout à fait correct de parler de contraction ou d'atteinte myocardique, mais jamais d'infarctus myocardique. (Rouleau 2003 : 145)

- 10 Dans le commentaire qui suit cette assertion, Rouleau, sans s'attarder sur les raisons qui ont pu faire que l'usage a consacré l'une des deux solutions possibles grammaticalement, souligne que le traducteur n'a pas d'autre choix que de se conformer à la manière de parler des spécialistes. On pourrait objecter que s'il est vrai que « infarctus du myocarde » est beaucoup plus employé que « infarctus myocardique³ », celui-ci n'est pas totalement absent de la littérature spécialisée. Par ailleurs, il convient de distinguer les termes des collocations quand on examine les phénomènes de variation entre l'utilisation de l'adjectif relationnel et celle du complément prépositionnel⁴. Ainsi, « contraction myocardique » et « atteinte myocardique » sont des collocations plutôt que des termes, et leur utilisation est d'une fréquence comparable à celle de leurs équivalents à complément prépositionnel (« contraction du myocarde » et « atteinte du myocarde »). Toutefois, dans le cas d'autres termes et collocations d'usage courant (*myocardial depression*, *myocardial ischemia*, *myocardial function*, *myocardial necrosis*, *myocardial perfusion*), la consultation des moteurs de recherche fait apparaître une fréquence en moyenne dix fois supérieure des équivalents de traduction français utilisant l'adjectif relationnel par rapport à la complémentation.
- 11 D'un point de vue lexicographique, il est cependant intéressant que des apprenants aient accès à une information concernant le degré de variation observé dans la traduction des termes aussi bien que des collocations. La disponibilité de cette information prend tout son sens dans le cas d'un dictionnaire anglais-français regroupant l'ensemble des unités polylexicales sous leur forme dégroupée, les termes pouvant être distingués des collocations par un enrichissement typographique (Tableau 2).

Tableau 2. Traduction de quelques termes et collocations contenant l'adjectif « *myocardial* »

Anglais	Français
myocardial contraction	contraction myocardique, contraction du myocarde
myocardial contractility	contractilité myocardique, contractilité du myocarde
myocardial disorder	atteinte myocardique, atteinte du myocarde ; trouble myocardique, trouble du myocarde (rares).
myocardial infarction	infarctus du myocarde, infarctus myocardique (rare)
myocardial ischemia	ischémie myocardique, ischémie du myocarde (rare)

- 12 La mention de la fréquence, par opposition à une simple prise de position de nature binaire du lexicographe (inclusion ou omission), pose fatalement le problème du degré de raffinement qu'il est nécessaire d'apporter à ce type d'information. Là encore, le

lexicographe doit opérer des choix. Si nous prenons l'exemple de l'entrée d'un dictionnaire bilingue anglais-français consacrée au terme *bone marrow*, la première question est celle de la présence des deux traductions possibles (« moelle osseuse » et « moelle des os »). La deuxième formulation est nettement moins employée que la première⁵, mais on en relève néanmoins de très nombreux emplois. La même question pourrait être posée à propos des deux traductions possibles du terme *bone marrow transplant* : là encore, « greffe de moelle osseuse » est nettement plus fréquent que « greffe de la moelle osseuse », ce qui confirme la propension bien connue du français à préférer dans les termes polylexicaux l'ellipse de l'article devant l'élément postmodifiant (Kocourek 1991 : 139⁶).

- 13 L'importance du dégroupement complet de toutes les unités polylexicales doit ici être soulignée : la présence contiguë des termes et collocations de l'anglais résultant de l'utilisation d'une unité lexicale pour en prémodifier une autre a une valeur pédagogique indéniable. Dans le cas de l'exemple cité plus haut, la faible fréquence d'emploi de « moelle des os » peut être mise en parallèle avec l'impossibilité de la traduction de *bone marrow transplant* par « greffe de la moelle des os ». De la même manière, la présence contiguë de *sickle cell* (cellule falciforme) et de *sickle cell disease* (drépanocytose)⁷ fait apparaître clairement le passage d'une étymologie latine à son équivalent grec, la formation du mot français (drépano-cyt-ose) constituant un calque exact de l'ordre syntaxique du composé anglais.
- 14 On a vu que l'examen des traductions françaises de divers composés anglais de type N1 N2 dans la langue médicale spécialisée fait apparaître un usage intensif de l'adjectif relationnel, qui peut co-exister avec l'utilisation d'une post-modification par complément. Il est intéressant de se pencher sur la nature des relations sémantiques entre les deux noms afin de tenter d'expliquer certaines tendances récurrentes dans ce domaine. Nous examinerons à cet effet quelques traductions d'expressions anglaises dont le prémodifiant est le nom *platelet*. Le Tableau 3 décrit les fréquences relatives observées pour la traduction de ces expressions ; l'ordre adopté étant celui de la fréquence relative dégressive de l'emploi de l'adjectif relationnel.
- 15 Les deux premières lignes du tableau font apparaître le phénomène déjà observé à propos de *bone marrow transplant* : la fréquence d'emploi de l'adjectif relationnel augmente quand l'ensemble du composé N1 N2 (*platelet function*) prémodifie un N3 (*disorders*). On trouve également plus souvent l'adjectif relationnel quand le nom prémodifié est un déverbal faisant référence à une relation subjective (*function, aggregation*), alors que les emplois du complément prépositionnel deviennent aussi fréquents dès lors que la relation est objective (*count, activation, et concentration* désignent des processus subis par les plaquettes). Enfin, on observe une fréquence supérieure de la complémentation lorsque le verbe qui sous-tend la relation implique la présence de deux autres actants (verbes de transfert comme *transfuse* ou *donate*).

Tableau 3. Traductions des groupes nominaux contenant le nom « *platelet* » comme prémodificateur

Anglais	Français
platelet function disorders	troubles de la fonction plaquettaire, troubles de la fonction des plaquettes (beaucoup plus rare)

platelet function	fonction plaquettaire, fonction des plaquettes (plus rare)
platelet aggregation	agrégation plaquettaire, agrégation des plaquettes (plus rare)
platelet concentration	concentration plaquettaire, concentration des plaquettes
platelet activation	activation plaquettaire, activation des plaquettes
platelet count	numération plaquettaire, numération des plaquettes
platelet transfusion	transfusion plaquettaire, transfusion de plaquettes (plus fréquent)
platelet donor	donneur de plaquettes, donneur plaquettaire (terme inusité)

2. La double prémodification nominale (N1 N2 N3)

- 16 Les groupes nominaux constitués de suites de trois noms posent au lexicographe les mêmes problèmes que les suites de deux noms quant à leur statut lexicologique et leurs critères de dictionnarisation. Il convient d'une part de distinguer les collocations (*platelet function disorders*) des termes (*bone marrow transplant*). Ces derniers devront idéalement pouvoir être repérés, comme on l'a dit plus haut, à l'aide d'une présentation adéquate. La question de l'inclusion de collocations telles que *platelet function disorders* (troubles de la fonction plaquettaire) dans un dictionnaire pour apprenants peut également se poser en termes de prévisibilité de la traduction, qui ne pose pas de problème majeur pour qui connaît le sens de *platelet function*. Ce critère doit nécessairement être pris en compte dans le cadre d'un dictionnaire partiellement combinatoire, car même dans le cas des dictionnaires électroniques, pour lesquels le critère de la taille de l'ouvrage de référence devient secondaire, on ne peut se permettre de noyer l'information pertinente sous des masses de données partiellement répétitives.
- 17 L'imbrication des prémodifications est au cœur de cette problématique. On sait que les termes de longueur 2 sont fréquemment imbriqués à l'intérieur de termes ou de collocations de longueur supérieure, certains programmes d'extraction terminologique se fondant d'ailleurs sur cette caractéristique (Frantsi *et al.* 2001). L'analyse du schéma de prémodification par l'apprenant présuppose un choix entre deux découpages possibles de cette relation de prémodification : N1 prémodifie N2-N3 (*placebo control group*) ou N1-N2 prémodifie N3 (*bone marrow transplant*). Ce choix dépend de l'identification du lien privilégié entre les deux noms qui forment la séquence pré-construite (qu'il s'agisse d'un terme ou d'une collocation)⁸. Si un dictionnaire bilingue contient par exemple la traduction de *platelet function* mentionnée plus haut (fonction plaquettaire), on peut se demander s'il est nécessaire de faire figurer dans ce même dictionnaire les traductions de séquences comme *platelet function abnormality*, *platelet function analyzer*, *platelet function defects*, *platelet function inhibitor* ou *platelet function testing*, les traductions de ces collocations contenant toutes une post-modification prépositionnelle sous la forme de la séquence « de la fonction plaquettaire ». Même lorsque la prémodification porte sur un groupe nominal complexe (*platelet function effects of acetylsalicylic acid* – effets de l'acide acétylsalicylique sur la fonction

plaquettaire – la traduction semble aller de soi dès lors que *platelet function* est identifié en tant qu'élément linguistiquement préconstruit.

Tableau 4. Quelques exemples de traduction de groupes nominaux de type [N1-N2] N3

Anglais	Français
birth control pills	contraceptifs oraux
blood pressure changes	modifications de la pression artérielle
body fat distribution	répartition de la graisse corporelle
body mass index	indice de masse corporelle
calcium channel blockers	inhibiteurs calciques
clotting factor production	production des facteurs de coagulation
emergency department patients	patients des urgences
emergency department physicians	médecins des urgences, médecins urgentistes
health care professionals	professionnels de (la) santé
plasma exchange program	programme de plasmaphérèse
skin biopsy specimen	prélèvement cutané
weight loss group	groupe dans lequel une réduction pondérale a été obtenue

- 18 Penchons-nous dans un premier temps sur les groupes nominaux de structure [N1-N2] N3⁹. L'examen du Tableau 4 fait apparaître une absence fréquente de traduction littérale respectant strictement le nombre de composants du groupe nominal original. C'est notamment le cas pour des expressions d'usage très fréquent pour lesquelles l'ellipse correspond à la réalisation la plus courante dans les expressions polylexicales du français (*health care* – [soins de] santé, *emergency department* – [service des] urgences). Un autre exemple d'ellipse (*skin biopsy specimen* – prélèvement cutané) voit l'effacement de l'un des termes en français, le mot « prélèvement » suffisant en contexte à traduire la notion d'un « échantillon biopsique ». La traduction employée dans le corpus pour *birth control pills* est bien la plus courante, mais elle est concurrencée par une traduction plus littérale (pilules contraceptives). Cette absence de traduction littérale peut constituer un critère de sélection dans le cas de la lexicographie bilingue : la présence d'une entrée consacrée à *plasma exchange* permet de prédire de manière fiable la traduction qui sera adoptée quand ce groupe nominal est suivi d'un mot d'emploi aussi fréquent que *program*. L'inclusion de *plasma exchange program* manque donc d'intérêt dans ce cas précis. Autrement plus intéressant est le cas de *calcium channel blockers*, initialement traduit par « inhibiteurs des canaux calciques », forme qui tend à être supplantée par un équivalent elliptique, « inhibiteurs calciques¹⁰ ». On a ici affaire à un double glissement de sens, dans la mesure où le terme

« canal calcique » constitue une espèce d'hypallage, l'adjectif ne désignant pas ici la nature du canal, orifice situé dans la membrane cellulaire qui permet les mouvements de l'ion calcium de part et d'autre de celle-ci. En ce qui concerne les termes *body fat distribution* et *body mass index*, la présence d'entrées consacrées à leurs « sous-ensembles » *body fat* et *body mass* ne dispense pas le lexicographe de les inclure dans un dictionnaire bilingue, car la variation de l'usage entre « répartition » et « distribution » d'une part et « indice » et « index » d'autre part s'opère encore nettement en faveur d'une traduction non littérale, qui vaut donc la peine d'être répertoriée.

- 19 L'examen des traductions utilisées pour les groupes nominaux de structure N1 [N2-N3] (Tableau 5) fait quant à lui apparaître un certain nombre de tendances dont certaines plaident en faveur de leur recensement dans un dictionnaire bilingue. Seules deux des traductions employées utilisent trois noms en français (*defibrillation energy requirements* – niveau d'énergie utile à la défibrillation, *thallium stress test* – test d'effort au thallium), et la traduction du N1 par un adjectif relationnel (*erythrocyte* – érythrocytaire, *plasma* – plasmatique, *serum* – sérique, *tissue* – tissulaire) est fréquente. On observe également la tendance au regroupement en une seule lexie par la suffixation, déjà noté à propos de *sickle cell disease* (*blood cholesterol level* – cholestérolémie, *urine sodium excretion* – natriurie). Notons au passage que l'ellipse à la traduction n'est pas nécessairement la preuve qu'un élément de l'expression de la langue de départ est de moindre importance terminologique (*prevention study arm* – bras « prévention », *placebo control group* – groupe placebo).
- 20 Lesquelles de ces expressions sont-elles dictionnarisables ? La séquence *defibrillation energy requirements*, même si elle est d'usage fréquent, semble susceptible d'être traduite d'autres manières qu'elle ne l'est ici, et ne pose pas de problème de compréhension si l'on connaît le sens des composants isolés. Les expressions dont le N1 est *plasma* ou *serum* ne semblent pas non plus devoir être incluses, dans la mesure où la traduction par l'adjectif relationnel correspondant est systématique dans le cadre d'expressions dont le N3 indique la mesure du N2 dans N1 (*concentration*, *level*). La dictionnarisat[i]on nous paraît souhaitable dans tous les autres cas, soit parce que l'adoption de l'adjectif dans la traduction française n'est pas forcément prévisible (*tissue plasminogen activator* – activateur tissulaire du plasminogène)¹¹, soit parce que deux traductions coexistent et méritent donc d'être signalées (*erythrocyte sedimentation rate* se traduit également par « vitesse de sédimentation des érythrocytes »).
- 21 Les suites de quatre noms et plus, relativement fréquentes dans certaines spécialités comme la physique (Richalot 2003), sont plutôt rares en anglais médical. Le découpage des séquences de longueur 4 suit généralement l'un de ces trois schémas : [N1-N2] [N3-N4], N1 [[N2-N3] N4] ou [[N1-N2] N3] N4. Ce dernier schéma semble d'ailleurs être le plus fréquent. Dans de nombreux cas de figure, l'explicitation des liens entre modifiant et modifié nécessite l'usage d'autres moyens que la simple complémentation en « de ». Ainsi, dans la traduction de *LDL receptor gene locus* par « site chromosomique du gène codant les récepteurs aux LDL¹² », on retrouve le schéma [[N1-N2] N3] N4, mais le lien entre N1 et N2 est traduit par la préposition « à », et celui entre N3 et l'expression N1 - N2 l'est ici par la locution « gène codant (pour) N » qui est fréquemment utilisée pour le patron *N gene*.

Tableau 5. Quelques exemples de traduction de groupes nominaux de type N1 [N2-N3]

Anglais	Français
blood cholesterol level	cholestérolémie
blood glucose level	glycémie
defibrillation energy requirements	niveau d'énergie utile à la défibrillation
erythrocyte sedimentation rate	vitesse de sédimentation érythrocytaire
placebo control group	groupe placebo, groupe témoin sous placebo
plasma cholesterol concentration	cholestérolémie
plasma coagulation factors	facteurs plasmatiques de coagulation
plasma lipoprotein concentrations	concentrations plasmatiques de lipoprotéines
platelet survival times	durée de vie plaquettaire
prevention study arm	bras « prévention »
serum testosterone levels	taux sériques de testostérone
thallium stress test	test d'effort au thallium
tissue plasminogen activator	activateur tissulaire du plasminogène
urine sodium excretion	natriurie

- 22 La présence d'autres expressions de longueur 4 formées à partir de la même expression de type [N1-N2] N3 (*LDL receptor gene mutation* – mutation du gène codant pour les récepteurs à LDL), plaide en faveur de la dictionnarisaiton des suites *LDL receptor* et *LDL receptor gene*¹³. Comme on observe également la présence d'une forme elliptique de *LDL receptor gene mutation* (*LDL receptor mutation*), elle-même utilisée comme prémodificateur au sein de diverses expressions (*LDL receptor mutation genotype*, *LDL-receptor mutation screening*), l'inclusion des deux versions de ce terme semble également souhaitable. Mais qu'en est-il des autres expressions de longueur 3 dont *LDL receptor* est le prémodificateur ? La traduction littérale de *LDL receptor dysfunction* (dysfonctionnement des récepteurs aux LDL) employée dans le corpus bilingue semble aller de soi. Quant aux autres expressions relevées, elles ont donné lieu à des reformulations qui semblent indiquer qu'elles n'ont pas de traduction clairement établie (*inability to induce [...] LDL receptor expression* – absence de récepteurs aux LDL, *one of the two LDL receptor alleles that lead to dysfunctional gene products* – l'un des deux allèles responsables d'un codage défectueux des récepteurs aux LDL)¹⁴. Remarquons au passage que la limite de l'imbrication des termes semble bien se situer au niveau d'une prémodification faisant intervenir cinq noms : on rencontre bien *the overexpression of the LDL receptor gene in the liver* mais pas *LDL receptor gene liver overexpression*, alors que le schéma *liver overexpression of N* est attesté.

3. La prémodification adjectivale (ADJ N1 N2)

- 23 Le patron syntaxique ADJ N1 N2 est d'usage extrêmement fréquent en anglais médical¹⁵. Il pose intrinsèquement des problèmes de décodage similaires à ceux des groupes nominaux formés par la concaténation de trois noms. Le découpage syntaxique, de nature binaire, s'effectue en fonction de la reconnaissance éventuelle d'un lien de prémodification entre l'adjectif et le premier nom de la chaîne. Ainsi, on reconnaît la suite *coronary artery* dans le groupe nominal [*coronary artery*] *disease*, alors que l'absence de reconnaissance d'un lien de prémodification dans la suite *coronary heart* (qui peut se conjuguer à la connaissance préalable de l'unité terminologique *heart disease*) impose un découpage inverse pour *coronary* [*heart disease*]¹⁶.
- 24 La dictionnairisation des suites ADJ N1 et N1 N2 constituera pour l'apprenant un indice fiable pour arriver au découpage syntaxique correct. Inversement, l'absence de dictionnairisation de la suite ADJ N2 peut également être un indice favorisant le découpage faisant intervenir le préconstruit [ADJ N1]. La statistique lexicale peut de ce point de vue constituer une aide précieuse. En effet, dans les cas où l'adjectif qualifie le premier nom, on peut s'attendre à ce que la suite ainsi formée (si elle est un constituant terminologique de l'ensemble de la séquence) soit d'usage fréquent. Si l'adjectif qualifie le deuxième nom (ou l'ensemble N1-N2), on peut supposer qu'il en va de même pour la séquence de deux noms qu'il précède.
- 25 Cette hypothèse a été testée sur les séquences ADJ N1 N2 les plus fréquentes du corpus JAMA. Nous avons pour cela utilisé l'indice de probabilité de co-occurrence que constitue le z-score du logiciel Tact (les valeurs ont été arrondies à l'entier le plus proche). La valeur du z-score peut aller jusqu'à 200 dans le cas de deux mots qui sont présents exclusivement en co-occurrence dans un corpus donné¹⁷, et elle est proche de 0 pour les mots dont la co-occurrence est extrêmement rare. Par exemple, dans le Tableau 6, les valeurs très élevées observées pour *mitral valve* (193) et *valve prolapse* (172) sont dues au fait que *mitral* apparaît cinq fois dans le corpus, toujours suivi de *valve*, qui n'est lui-même employé que sept fois (*prolapse* n'apparaît que quatre fois, toujours dans la séquence *mitral valve prolapse* – prolapsus de la valve mitrale). Le codage N1 a été utilisé pour les groupes nominaux de structure syntaxique [ADJ N1] N2, et le codage N2 pour les groupes nominaux de structure syntaxique ADJ [N1 N2].
- 26 L'examen du Tableau 6 fait apparaître que dans la quasi-totalité des cas (18/19, soit 90 %) cette probabilité de co-occurrence est un indicateur fiable de la portée de l'adjectif. Le seul cas de figure dans lequel le z-score ne confirme pas la portée de la modification adjectivale (*conjugate eye deviation* – déviation conjuguée du regard) est d'ailleurs dû à un manque de finesse de l'étalon que nous avons choisi. En effet, *conjugate* et *deviation* n'étant employés qu'une seule fois dans un autre environnement, les z-scores sont identiques. Il faut également signaler que le codage que nous avons choisi opère une dichotomie dans certains cas où la portée de la modification adjectivale n'est pas clairement définie. Ainsi, dans *coronary risk factors*, il semble bien que l'expression *risk factors* tout entière soit qualifiée par l'adjectif *coronary*, mais *coronary risk* fait également partie d'un découpage syntaxique plausible. La traduction française emploie d'ailleurs systématiquement « facteurs de risque coronaire ».

Tableau 6. Probabilité de co-occurrence des suites ADJ-N1 et N1-N2 des séquences du type Adjectif-Nom-Nom les plus fréquentes

Séquence ADJ N1 N2	Codage	z-score ADJ-N1	z-score N1-N2
coronary artery disease	N1	115	88
ischemic stroke subtype	N1	67	61
mitral valve prolapse	N1	193	172
visceral fat accumulation	N1	118	63
sexual function questionnaire	N1	40	17
giant cell arteritis	N1	93	25
transvenous lead systems	N1	38	26
coronary artery lesions	N1	115	13
coronary risk factors	N2	11	33
overt heart failure	N2	56	104
normal ejection fraction	N2	21	210
low ejection fraction	N2	32	210
diastolic blood pressure	N2	16	130
systolic blood pressure	N2	17	130
valvular heart disease	N2	15	41
congestive heart failure	N2	24	105
conjugate eye deviation	N2	118	118
atherogenic lipoprotein particles	N2	17	30
prosthetic heart valves	N2	24	26

- 27 L'indice de probabilité de co-occurrence des suites ADJ-N1 et N1-N2 semble donc un indicateur fiable d'interprétation de la structure syntaxique des groupes nominaux de type ADJ N1 N2¹⁸. La dictionnairisation des groupes bilinguistiques entrant dans la composition de ces groupes nominaux est donc un élément essentiel à leur compréhension par l'apprenant. La question se pose toutefois de savoir quelles séquences ADJ N1 N2 sont elles-mêmes sujettes à une dictionnairisation. La fréquence d'emploi de ces séquences constitue bien sûr un premier critère, mais l'appartenance à la catégorie N1 semble également favoriser statistiquement la possibilité d'interpréter

ces séquences comme des termes susceptibles d'être inclus dans un ouvrage de référence ou une base de données. La comparaison des groupes distingués par les deux types de découpage fait en effet apparaître un certain nombre de distinctions d'ordre morphologique et sémantique entre les adjectifs utilisés. Ainsi, lorsque l'adjectif qualifie le deuxième nom, les adjectifs utilisés appartiennent presque toujours à l'une des catégories suivantes :

- adjectifs courts d'utilisation fréquente (*deep, high, low, new, rare*) : lorsqu'un tel adjectif est utilisé dans une séquence de type ADJ-N1-N2, il ne qualifie le premier nom que dans 6 % des cas ;
 - participes passés employés comme adjectifs (*diminished, improved, increased, shortened*). Un grand nombre d'entre eux expriment un changement d'état, et leur traduction fait très souvent apparaître une transposition vers le nom (*diminished blood flow* – diminution du débit sanguin, *shortened platelet survival* – raccourcissement de la durée de vie plaquettaire). Dans cette catégorie, l'adjectif ne qualifie le premier nom que dans 8 % des cas ;
 - comparatif et superlatif des adjectifs : dans notre corpus, l'adjectif qualifiait le deuxième nom pour toutes ces formes (13 occurrences) ;
 - adjectifs à sens temporel, exprimant la fréquence, la vitesse, l'ordre chronologique, etc. (*current, daily, frequent, original, previous, progressive, prompt, recent, rare, subsequent*) ;
 - adjectifs exprimant le degré (*absolute, complete, considerable, extensive, important, significant*) ;
 - adjectifs exprimant la quantité (*additional, cumulative, numerous, various*) ;
 - adjectifs exprimant un jugement de valeur (*adequate, appropriate, defective, effective, negative, positive*) ;
 - adjectifs exprimant la modalité assertive (*eventual, potential*).
- 28 Les adjectifs débutant de telles séquences sont statistiquement peu représentés dans la terminologie médicale¹⁹. Par comparaison, les adjectifs qualifiant le premier nom font presque toujours partie du vocabulaire spécialisé (*atherogenic, atrial, cardiac, coronary, ischemic, mitral, systolic, temporal, visceral, etc.*).
- 29 Un deuxième critère de dictionnairisation doit à nouveau être évoqué, celui de la non-littéralité de la traduction française. Les groupes nominaux étudiés précédemment donnent lieu à des traductions françaises dont le patron syntaxique dépend du découpage adopté. Le schéma [ADJ N1] N2 donnera généralement lieu à une complémentation pouvant faire intervenir diverses prépositions, « de » et « à » étant les plus employées (*giant cell arteritis* – artérite à cellules géantes, *visceral fat accumulation* – accumulation de graisse viscérale). On note que lorsque le N1 anglais est dénombrable, le choix du nombre pose problème : la traduction correcte de *giant cell arteritis* implique des connaissances de type encyclopédique (plusieurs cellules sont impliquées). Par ailleurs, lorsque N1 et N2 sont au singulier (« accumulation de graisse viscérale »), la traduction employée reproduit l'ambiguïté syntaxique de l'original, l'adjectif pouvant théoriquement qualifier le N1 ou le N2. Le schéma ADJ [N1 N2], quant à lui, est susceptible de donner lieu à deux patrons syntaxiques distincts : celui qui fait intervenir une complémentation (*normal ejection fraction* – fraction d'éjection normale) et celui pour lequel le N1 est traduit par un adjectif relationnel (*diastolic blood pressure* – pression artérielle diastolique). L'arbre décisionnel résultant pourrait être schématisé ainsi : choix entre adjectif relationnel et complémentation d'une part, puis choix de la préposition correcte dans le cas de la complémentation (« de » dans la plupart des cas). La multiplicité des choix, combinée au fait que plusieurs traductions peuvent avoir

cours, justifie la dictionnairisation au niveau de l'encodage en français, même dans les cas où le décodage de la structure syntaxique de l'anglais ne constitue pas une source d'ambiguïté.

- 30 Nous avons vu que le schéma [ADJ N1] N2 donne généralement lieu à une complémentation. Il convient cependant de relever quelques cas de variation des équivalents de traduction concernant ce patron syntaxique. Ainsi, *smooth muscle cell*, qui signifie littéralement « cellule du muscle lisse » (par opposition aux cellules du muscle strié) est le plus souvent traduit par « cellule musculaire lisse », l'équivalent à complément prépositionnel (cellule du muscle lisse) n'étant que rarement utilisé. Cette adjectivisation de l'ensemble ADJ-N1 semble encore plus courante en anglais, où l'on relève de nombreux exemples : ainsi les deux adjectifs du terme *atrial septal defect* (communication inter-auriculaire) ne sont pas à analyser sur le même plan syntaxique. Ce qui est signifié est bien *a defect of the atrial septum*, c'est-à-dire littéralement un défaut de la cloison auriculaire (*atrium* signifiant « oreillette », *atrial* se traduit par l'adjectif relationnel « auriculaire »). On assiste donc dans ce processus de prémodification complexe à la création d'un pseudo-adjectif relationnel que l'on peut analyser sous la forme (*atrial sept-*) -AL. Le mot « cloison » n'ayant pas donné lieu en français à la formation d'un adjectif relationnel, l'équivalent de traduction habituellement employé désigne le résultat de cette anomalie de la cloison (communication inter-auriculaire). D'autres formulations de structure identique ont donné naissance en français à des équivalents de traduction calqués sur l'original. C'est notamment le cas de *jugular venous pressure* (signifiant *pressure of the jugular veins*) pour lequel la formulation « pression veineuse jugulaire » est beaucoup plus employée que la structure faisant appel à la complémentation (pression des veines jugulaires), ainsi que de *right ventricular failure* (signifiant *failure of the right ventricle*), auquel on fait référence presque exclusivement en français sous la forme « insuffisance ventriculaire droite²⁰ ».

4. La coordination des groupes nominaux complexes

- 31 L'étude des différents patrons syntaxiques menée plus haut a démontré l'étendue des problèmes potentiels de décodage des groupes nominaux complexes. Les sources d'erreur de découpage peuvent de surcroît être multipliées dans le cas des structures coordonnées. Un bref exemple suffira à illustrer le nombre d'ambiguïtés que génère la combinaison de la prémodification nominale et de la coordination en anglais de spécialité :
- (1) The ability of PET to detect cancer is based on the altered substrate requirements of malignant cells, which result from increased nucleic acid and protein synthesis and glycolysis.
- 32 Au décodage, le traducteur de la phrase (1) est amené à se poser plusieurs questions concernant la structure syntaxique de la dernière partie de la phrase, *increased nucleic acid and protein synthesis and glycolysis*. Certaines de ces questions trouveront chez le traducteur humain spécialiste du domaine une réponse immédiate dans l'examen du contexte et des relations sémantiques qui lient les mots entre eux, mais poseront problème à un apprenant tout comme à un analyseur syntaxique automatique. En voici quelques-unes :
- *nucleic* qualifie-t-il *acid*, l'ensemble *acid and protein* ou bien *synthesis* ?

- *protein* est-il un prémodificateur du seul nom *synthesis* ou de l'ensemble *synthesis and glycolysis* ?
 - *increased* qualifie-t-il *acid, synthesis* ou bien l'ensemble *synthesis and glycolysis* ?
- 33 Les sources possibles d'erreurs se conjuguant, les chances d'arriver au découpage correct sans l'apport de connaissances lexicales sont réduites. Si l'on symbolise la portée des prémodifications à l'aide de crochets, le découpage correct est le suivant : *increased [[[[nucleic acid] and [protein]] synthesis] and glycolysis]*, et ce segment peut donc se traduire par « augmentation de la glycolyse et de la synthèse des protéines et de l'acide nucléique ». Les mécanismes de désambiguïsation du traducteur humain dépendent partiellement de sa connaissance de la réalité extralinguistique (la médecine) mais aussi d'une connaissance lexicale transmissible à la machine sous forme d'une base de données contenant les termes et les collocations de la langue de spécialité. Ainsi, si l'analyseur a accès à une telle base, les découpages supposant l'existence de *nucleic protein* ou de *nucleic synthesis* seront invalidés puisque ces séquences n'y figureront pas, et *nucleic acid synthesis* sera inversement validé. Quant à la résolution de l'ambiguïté concernant la portée de la prémodification par *increased*, elle repose fortement sur la connaissance scientifique (extralinguistique) du domaine de spécialité, même si la fréquence des structures coordonnées à la suite de participes passés comme *decreased* et *increased* peut donner lieu à une analyse statistique de la probabilité de distributivité de la prémodification. Nous nous bornerons ici à suggérer la manière dont les données lexicales peuvent aider l'apprenant à résoudre ce type d'ambiguïté²¹.
- 34 L'étude des structures de type ADJ N1 AND N2, ADJ N1 OR N2, N1 N2 AND N3, et N1 N2 OR N3 tirées de notre corpus bilingue fait apparaître que la résolution des problèmes provoqués par l'ambiguïté syntaxique passe par l'intériorisation de certains éléments de combinatoire lexicale, qu'il s'agisse d'une association de type paradigmatique ou syntagmatique. Dans les cas où la distributivité potentielle de la coordination s'applique (ADJ [N1 AND N2] = ADJ N1 AND ADJ N2), le lien entre les deux noms employés peut être de nature syntagmatique. Ainsi, dans les exemples (2) et (3), *aggregate sensitivity and specificity* (sensibilité et spécificité cumulées) et *extra time and expense* (temps et dépenses supplémentaires) seront interprétés correctement par l'apprenant s'il perçoit les suites de deux noms coordonnés comme des collocations. Il est intéressant de noter que dans le premier cas, la reconnaissance de la collocation fait appel à des connaissances lexicales relevant de la langue spécialisée (la sensibilité et la spécificité sont les deux modes de mesures de la fiabilité d'un test de dépistage) alors que la collocation *time and expense* appartient à la langue générale.
- (2) The aggregate sensitivity and specificity of ultrasonography and IPG in the medical literature are shown in Table 3.
- (3) Most dentists do not administer parenteral antibiotics in their offices, causing patients the extra time and expense of physicians' office visits.
- 35 Dans les exemples (4) et (5), ce n'est pas la fréquence d'occurrence de groupes coordonnés comme *studies and reviews* ou *occlusion or infection* qui induira le découpage correct (ADJ [N1 AND / OR N2]), et l'interprétation du lien unissant deux noms appartenant au même champ sémantique et partageant le même hyperonyme (document portant sur des recherches scientifiques dans un cas, complication iatrogène dans l'autre). Il s'agit donc plutôt d'une relation de nature paradigmatique entre les deux noms coordonnés²².

(4) English-language studies and reviews pertaining to LDL composition and size were identified through a MEDLINE search and reference citations.

(5) Catheter drainage was occasionally complicated by catheter occlusion or infection.

36 L'aide qu'un apprenant peut attendre d'un dictionnaire pour arriver à l'interprétation correcte de ces structures dépendra du type de lien qui unit les deux noms. La dictionnarisaiton des collocations fréquentes de la langue générale (*time and expense*) et de la langue spécialisée (*sensitivity and specificity*) semble souhaitable et réalisable²³. Celle d'une collocation comme *studies and reviews* peut paraître plus discutable, car même si sa fréquence est loin d'être négligeable, la proximité sémantique des deux co-hyponymes impliqués semble devoir assurer un décodage correct. Quant au découpage syntaxique de l'exemple (5), il sera facilité par le recensement de collocations telles que *catheter infection* et *catheter occlusion*, certainement utile dans le cadre d'une langue de spécialité.

37 Qu'en est-il des cas dans lesquels le découpage correct correspond à l'absence de distributivité de la prémodification ? L'interprétation correcte d'une structure de type [ADJ N1] AND N2 repose sur l'élimination logique du lien de prémodification entre ADJ et N2. Cette élimination dépend de connaissances lexicales de nature complexe qui peuvent intervenir à différentes étapes de l'analyse.

(6) Beyond the first few months, other factors such as social support and depression strongly influence patient function.

38 Dans l'exemple (6), l'angliciste reconnaîtra immédiatement le terme *social support* (soutien social), mais ce mécanisme de reconnaissance d'une entité lexicale préconstruite ne s'enclenchera pas forcément pour la reconnaissance de *social depression*, collocation d'usage beaucoup moins fréquent²⁴. Mais le mécanisme d'élimination du découpage distributionnel repose ici sur une autre relation sémantique : même si le lecteur connaît (et peut donc être susceptible de reconnaître) l'entité désignée par l'expression *social depression*, c'est l'improbabilité de la coordination entre les deux noms *support* et *depression*, perçus comme liés par une relation proche de l'antonymie, qui élimine logiquement la possibilité d'une distributivité de la prémodification. Cette antonymie rendrait par conséquent la répétition de l'adjectif *social* nécessaire si l'on voulait réellement évoquer ici la dépression sociale. Nous atteignons cependant dans ce cas précis les limites de l'aide que peut apporter un dictionnaire classique, les relations de synonymie et d'antonymie appartenant plutôt au domaine des thésaurus ou des bases lexicales informatisées comme Wordnet.

5. Un exemple de GN « hypercomplexe » : le patron syntaxique <ADJ ADJ N (AND/OR) N N>

39 Comme on l'a vu lors de l'étude de l'exemple (1), l'apprenant est souvent confronté à une multiplicité de découpages potentiels par le biais des phénomènes de coordination et de prémodification. Afin de tenter de décrire les mécanismes de compréhension qui favorisent l'élimination des découpages incorrects et le choix du découpage attendu, et compte tenu de la faible fréquence d'emploi des structures les plus complexes, nous avons utilisé pour la détection du patron syntaxique <ADJ ADJ N (AND/OR) N N> un corpus de plus grande taille que les précédents, constitué à partir du CD-ROM *Annals of*

Internal Medicine, et que nous désignerons désormais sous le nom de corpus AIM93. Ce corpus totalise 4,5 millions d'occurrences (*tokens*) et contient l'intégralité des articles publiés dans les revues suivantes pendant l'année 1993 : *New England Journal of Medicine*, *Journal of the American Medical Association*, *Annals of Internal Medicine*, *Lancet*, *British Medical Journal*.

40 Une fois l'ensemble des textes concaténés, on a appliqué au corpus un prétraitement automatique entraînant la séparation des mots et des signes de ponctuation, afin de lui faire subir une catégorisation en parties du discours (*part-of-speech tagging*) à l'aide du logiciel Winbrill, fourni par l'Institut National de la Langue Française (INaLF)²⁵. La consultation des contextes contenant les patrons syntaxiques étudiés a été effectuée à l'aide du concordancier Monoconc Pro²⁶.

41 Les catégoriseurs étant « entraînés » à partir de corpus de référence non spécialisés tels que le *British National Corpus*, on relève toutefois un certain nombre d'erreurs de catégorisation. Ainsi, dans l'exemple (7), l'adjectif *novel* est catégorisé par Winbrill comme un nom²⁷, car l'emploi nominal de *novel* est beaucoup plus fréquent que son emploi adjectival dans les corpus de référence.

(7) Increasing insight into the cellular mechanisms of hepatic fibrosis is providing novel prospects for more specific, effective control of the fibrogenic cascade.

42 Outre ces ambiguïtés catégorielles, les étiqueteurs automatiques peuvent également trébucher devant certaines ambiguïtés syntaxiques. Dans l'exemple (8), *mixed* est étiqueté comme un participe passé car le programme suppose une structure passive, alors qu'il s'agit d'un emploi de l'adjectif, utilisé dans une collocation disjointe avec le nom *results* :

(8) Results of selective gut decontamination trials have been mixed.

43 En raison de ces erreurs d'étiquetage, le filtre employé pour sélectionner le patron syntaxique évoqué plus haut a nécessité un tri manuel *a posteriori*. Parmi les erreurs les plus courantes, on note l'absence de reconnaissance des adjectifs substantivés (étiquetage de *contraceptive* en tant qu'adjectif dans *oral contraceptive use and cigarette smoking*) et plus généralement l'étiquetage des noms comme adjectifs si la forme adjectivale est la plus fréquente dans les corpus de référence (étiquetage de *fat* en tant qu'adjectif dans *dietary fat intake and breast cancer*).

44 L'étude des résultats de ce premier tri fait apparaître une fréquence élevée des structures dans lesquelles un lien sémantique fort unit les deux noms coordonnés, structures qui donnent généralement lieu au découpage ADJ (ADJ [N and N]) N lorsqu'elles sont intégrées dans le patron syntaxique étudié. Les structures de ce type de fréquence supérieure à 9 dans le corpus AIM93 sont regroupées dans le Tableau 7.

45 L'étude détaillée des 338 formes correspondant au patron syntaxique <ADJ ADJ N (AND/OR) N N> a permis d'identifier six découpages possibles :

46 Dans le premier cas de figure, que l'on peut symboliser par [ADJ ADJ N] and [N N], les deux groupes nominaux coordonnés sont clairement séparés par la conjonction, ce qui exclut toute « distributivité » de la prémodification. Dans l'exemple (9), on a affaire à deux termes complexes généralement connus des apprenants et facilement délimitables. Ce type de phrase posera peu de problèmes à l'apprenant muni d'un bagage lexical minimal en langue de spécialité.

(9) Four other patients had relapses affecting both the [central nervous system] and [bone marrow].

Tableau 7. Structures N AND N dans le corpus AIM93

Séquences N AND N les plus fréquentes	Fréquence
morbidity and mortality	434
age and sex	268
sensitivity and specificity	250
diagnosis and treatment	195
research and development	166
mortality and morbidity	126
dilatation and curettage	87
efficacy and safety	75
height and weight	49
cleaning and disinfection	29
detection and treatment	27
evaluation and management	26
illness and injury	10
intubation and ventilation	10

- 47 Le deuxième découpage possible consiste en la coordination de deux termes ou collocations de longueur 2, respectivement de patron ADJ N et N N, l'ensemble de cette structure coordonnée étant prémodifié par l'adjectif initial. On peut symboliser ce découpage ainsi : ADJ ([ADJ N] and [N N]).
- 48 Ce type de structure pose parfois problème à l'apprenant quant l'adjectif initial appartient au vocabulaire spécialisé, car la question de la distributivité est alors plus difficile à élucider. Dans l'exemple (10), la quasi-synonymie de *practices* et *policies* rend le découpage transparent. Dans l'exemple (11), le fait que le lien entre *data* et *rates* relève plutôt d'une relation hyperonymique rend le découpage un peu plus opaque, ce à quoi s'ajoute la difficulté de compréhension de la collocation *background rates*.
- (10) Control of epidemics relies essentially on good ([hygienic practices] and [hospital policies]) to restrict the use of antibiotics.
- (11) That system is vulnerable to biased reporting and it cannot measure the incidence of adverse reactions or provide reliable ([comparative data] or [background rates]) for everyday events in clinical practice.
- 49 Le troisième découpage possible, que l'on peut symboliser par ([ADJ ADJ N] and [N]) N²⁸, consiste en une prémodification du dernier nom par la coordination d'un nom qualifié par deux adjectifs avec un autre nom. Cette structure, relativement rare,

provoquera une erreur de découpage si l'apprenant n'exclut pas d'emblée une interprétation semblable à celle du premier découpage. Dans l'exemple (12), il s'agira donc d'éliminer comme improbable la collocation *complete an event*. Quant à l'exemple (13), une interprétation semblable à celle du premier découpage n'y serait guère dommageable, *tomography* et *ultrasonography* étant tous deux des hyponymes du nom *scan*.

(12) On completing a ([serious adverse event] or [withdrawal]) form the investigator was required to add his or her assessment of causal relation to the study drug.

(13) ([Computed abdominal tomography] and [ultrasonography]) scans were negative for biliary obstruction and liver metastases.

- 50 Le quatrième type de découpage a été évoqué plus haut et concerne les structures à l'intérieur desquelles un fort lien sémantique existe entre les deux noms qui encadrent la conjonction de coordination. On peut le symboliser par ADJ (ADJ [N AND N]). Son interprétation correcte dépend essentiellement de la perception de l'ensemble N AND N comme constituant un terme ou une collocation d'usage fréquent, comme c'est le cas dans l'exemple (14). L'autre problème de dépendance syntaxique est celui de la prémodification du dernier nom (et non pas des deux noms coordonnés) par le premier adjectif. Dans les exemples (14) et (15), cette interprétation est la plus probable dans la mesure où les collocations formées par le premier adjectif et le dernier nom du patron syntaxique sont avérées (*successful culture* et *persistent behaviors*).

(14) Little in the alternatives suggested here or in our accompanying article will strike anyone whose background lies within a successful (commercial [research and development]) culture as revolutionary.

(15) By contrast, the same treatments are less effective when administered only minutes later, once the persistent (central [excitability and pain]) behaviours have been established.

- 51 Une variante plus rare de ce dernier découpage est celle dans laquelle le premier adjectif précise le sens du deuxième de manière hyponymique. En anglais médical, le premier adjectif est fréquemment à la forme comparative dans les structures de ce type (*lower, upper, lesser, greater*). Le découpage, présenté dans l'exemple (16), peut se symboliser de la manière suivante : ([ADJ [ADJ]] [N AND N]) N.

(16) A different group did ([upper [gastrointestinal] [endoscopy and biopsy]) procedures in patients with AIDS and found HIV on 7 of 20 unwashed endoscopes.

- 52 Le sixième et dernier découpage possible est celui dans lequel les deux adjectifs initiaux modifient un groupe coordonné reliant un nom et un groupe nominal de forme N N (*fluids and tissue cultures* dans l'exemple (17) et *smoking and alcohol use* dans l'exemple (18)). Cette structure peut se symboliser par ADJ (ADJ [N AND [N N]]). L'erreur la plus commune des apprenants devant ce type de structure est l'adoption d'un découpage semblable à celui du premier patron syntaxique étudié, c'est-à-dire celui de l'exemple (9), dans lequel la prémodification est strictement interne aux deux groupes nominaux coordonnés. L'exemple (18) présente en réalité une structure un peu plus complexe, puisqu'il comporte un nom supplémentaire à la suite du patron ADJ (ADJ [N AND [N N]]). Le fait que l'adjectif *national* y modifie l'ensemble de ce qui le suit ou uniquement la partie entre parenthèses (à l'exclusion du nom *rates*) n'a en tout cas qu'une incidence marginale sur le décodage de ce groupe nominal complexe. Il est intéressant de noter que ce type de découpage fait intervenir des adjectifs dont les caractéristiques sémantiques ne les placent ni en relation de synonymie ni en relation d'hyponymie, comme c'était le cas dans l'exemple (16). En effet, *synovial* signifie une localisation

anatomique alors que *rheumatoid* indique un type de pathologie. De même, *national* a un caractère localisant d'un point de vue statistique selon une gradation des subdivisions géographiques d'un pays, alors que *occupational* indique une catégorie d'utilisation des substances nocives qui s'oppose à l'utilisation domestique (*home use*) selon une division binaire.

(17) GM-CSF and M-CSF have been found in rheumatoid (synovial [fluids and [tissue cultures]]).

(18) Singleton and Beaumont then attempted to control for smoking, alcohol, and socioeconomic status indirectly by adjusting for national (occupational [smoking and [alcohol use]]) rates.

Conclusion

- 53 L'étude et les analyses précédentes nous semblent suffisamment convaincantes pour autoriser la mise en œuvre de méthodes de recherche systématique de groupes nominaux complexes dans les corpus spécialisés. Leur traitement, dans lequel les critères statistiques habituels (fréquence et z-score notamment) seront associés aux analyses linguistiques, conduira à l'établissement d'une liste des groupes nominaux potentiellement problématiques pour l'apprenant francophone.
- 54 Si l'intégration d'un grand nombre de groupes nominaux complexes dans un dictionnaire médical bilingue fondé sur le dégroupement et produit sur support informatique nous semble souhaitable et possible, il n'en reste pas moins que le problème de la consultation de cet outil par l'apprenant reste posé. On ne peut ignorer les risques créés par la fausse transparence et les fausses amitiés terminologiques évoqués au début de cet article. C'est pourquoi nous favorisons une approche où apprenant, texte et dictionnaire seraient systématiquement associés dès les premiers stades de l'apprentissage. Cette triade est immédiatement efficace si le texte est numérique et si le dictionnaire est associé au logiciel de traitement de texte, par exemple à la manière d'un correcteur orthographique classique. Le système attirera l'attention de l'apprenant sur les groupes nominaux complexes inclus dans le dictionnaire. La consultation de celui-ci fournira une traduction en français et, si le rédacteur du dictionnaire l'a prévu, des informations supplémentaires de nature linguistique ou encyclopédique dont la valeur didactique ne peut être ignorée. La mise en commun d'informations linguistiques relatives aux deux langues peut ainsi apporter à l'apprenant une aide pour une meilleure maîtrise des concepts.

BIBLIOGRAPHIE

Cormier, Monique. 1990. « Traduction de textes de vulgarisation et de textes didactiques : approche pédagogique ». *Meta* 35/ 4, 676-688.

Cormier, Monique & Rhoda P. Roberts. 2000. « Lexicographie comparée du français et de l'anglais au Canada : le Dictionnaire canadien bilingue ». In Szende, Thomas (dir.), *Approches contrastives en lexicologie bilingue*. Paris : Éditions Champion, 213-222.

Frantzi, Katerina T., Junichi Tsujii & Sophia Ananiadou. 1999. « Clustering terms using the C-value method for automatic term recognition ». In Sandrini, P. (dir.) *TKE '99, Terminology and Knowledge Engineering*, 356-366.

Kocourek, Rostislav. 1991. *La langue française de la technique et de la science. Vers une linguistique de la langue savante*. 2^e édition. Wiesbaden : Brandstetter Verlag.

Maniez, François. 2000. « La prémodification nominale en anglais médical : quelques problèmes de traduction ». In Banks, D. (dir.), *Le groupe nominal dans le texte spécialisé*. Paris : L'Harmattan, 117-136.

Maniez, François. 2001. « L'ambiguïté syntaxique due aux structures coordonnées en anglais médical : analyse de la performance d'un logiciel d'aide à la traduction. » *Actes du colloque TALN de Tours*, 2-5 juillet 2001, Tome 2, 277-286.

Maniez, François. 2002. « The use of electronic corpora and lexical frequency data in solving translation problems ». In Altenberg, Bengt & Sylviane Granger (dir.), *Lexis in Contrast, Studies in Corpus Linguistics 7*. Amsterdam : John Benjamins, 291-306.

Richalot, Jérôme. 2003. « Approche énonciative du contexte d'occurrence des candidats termes : constitution et exploitation d'un corpus d'anglais en langue de spécialité », thèse soutenue à l'Université Lyon 3.

Roberts, Rhoda P. & Monique Cormier. 2000. « L'analyse des corpus pour l'élaboration du Dictionnaire canadien bilingue ». In Szende, Thomas (dir.), *Approches contrastives en lexicologie bilingue*. Paris : Éditions Champion, 223-240.

Rouleau, Maurice. 1994. *La traduction médicale, une approche méthodique*. Montréal : Linguatex.

Rouleau, Maurice. 2001. « La facture des principaux dictionnaires médicaux français : point de vue d'un traducteur ». *Meta* 46/1, 34-55.

Rouleau, Maurice. 2003. « La terminologie médicale et ses problèmes ». *Panacea*, vol. IV, n° 12. <http://www.medtrad.org/panacea/PanaceaPDFs/Panacea12_junio2003.pdf>.

Selva, Thierry, Serge Verlinde & Jean Binon. 2003. « Vers une deuxième génération de dictionnaires électroniques ». In Zock, Michael & John Carroll (dir.), *TAL 44 - 2/2003 Les dictionnaires électroniques*, 177-198.

Thoiron, Philippe. 2000. « La traduction des termes scientifiques : jeu entre concepts et termes ». In *Le Langage scientifique, Congrès National des sociétés historiques et scientifiques, 119^e, Amiens, 120^e, Aix-en-Provence*, 329-339.

Tournier, Jean. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris : Champion-Slatkine.

NOTES

1. Corpus de Langue Appliquée en Sciences de la Santé <http://perso.univ-lyon2.fr/~maniezf/recherche_corpus.html>. Il s'agit d'un corpus multilingue aligné formé à partir des textes disponibles sur le portail « L'accès au droit de l'Union européenne » (<http://eur-lex.europa.eu/>)

fr/index.htm). Le corpus est aligné au niveau de la phrase. Les langues représentées sont l'anglais, le français, l'allemand, l'espagnol, l'italien et le portugais.

2. On trouve quelques rares emplois de « cellule en faucille ».
3. Une requête effectuée sur le moteur de recherche Google révèle que le nombre d'occurrences de « infarctus myocardique » (environ 700) ne représente que 3 % de celui relevé pour « infarctus du myocarde ».
4. Dans le cadre de cet article nous ne reviendrons pas sur les diverses définitions proposées pour « collocation ». Nous utiliserons « terme » pour désigner les unités terminologiques lexicalisées et « collocation » pour faire référence aux co-occurrences de deux noms ou d'un adjectif et d'un nom unis par un lien de modification. Nous excluons donc les suites coordonnées (du type « *doctors and nurses* ») ainsi que les co-occurrences issues de contraintes d'ordre grammatical (parfois nommées « colligations »).
5. Exactement quinze fois moins selon des sondages effectués par requête sur divers moteurs de recherche.
6. L'absence totale d'occurrences de « greffe de (la) moelle des os » lors de requêtes effectuées sur les moteurs de recherche de la Toile semble indiquer que les prémodifications nominales en cascade de l'anglais correspondent à l'emploi de l'adjectif relationnel en français.
7. On parle également d'anémie falciforme, *sickle cell anemia* étant un synonyme de *sickle cell disease*.
8. Remarquons au passage que si l'apprenant ne peut opérer ce choix faute de connaissances lexicales, le schéma de prémodification [N1-N2] N3 est beaucoup plus fréquemment employé en anglais médical (70 % des groupes nominaux sont formés d'une suite de trois noms dans notre corpus bilingue aligné).
9. Ces exemples, ainsi que ceux du Tableau 5, sont tirés d'un corpus bilingue aligné au niveau de la phrase, constitué de 58 articles tirés du *Journal of the American Medical Association* et de leur traduction parue dans la version française de cette revue.
10. Pour une discussion plus détaillée de ce cas particulier, voir Rouleau (1994 : 192).
11. Il s'agit au demeurant d'un terme ayant donné lieu à siglaison (*TPA*).
12. Le sigle LDL représente lui-même une suite de schéma syntaxique [ADJ-N] N (*low density lipoprotein*), désignant l'un des deux principaux types de cholestérol.
13. Le statut terminologique de *LDL receptor* est au demeurant attesté par l'existence de sous-classes telles que *functional LDL receptors* ou (*human*) *hepatic LDL receptors*. Par ailleurs, et il s'agit également d'un critère directement exploitable par les programmes d'extraction terminologique, la suite *receptor gene* ne se rencontre jamais seule, mais toujours précédée d'un nom désignant un substance (hormone, acide aminé) ou d'un adjectif désignant une fonction (*olfactory receptor gene*).
14. On relève quelques rares emplois de « expression des récepteurs aux LDL ».
15. Les exemples cités dans cet article sont extraits d'un ensemble de 552 co-occurrences de ce patron syntaxique obtenues à partir d'un corpus de 30 articles tirés du *Journal of the American Medical Association* ayant pour sujet la cardiologie et totalisant 134 000 mots, auquel nous ferons référence sous le nom de corpus JAMA dans la suite de cet article.
16. Le fait qu'il existe une formulation elliptique (*coronary disease*, traduite alternativement par « maladie coronaire » ou « maladie coronarienne ») n'a pas d'influence sur le décodage de cette séquence comme signifiant « disease of the coronary artery » plutôt que « coronary disease of an artery ».
17. Ce serait par exemple le cas pour les deux mots de la forme latine d'un terme complexe comme *alopecia aerata* (pelade).
18. Il convient de remarquer que sa prise en compte nécessite l'utilisation d'un corpus lemmatisé et étiqueté grammaticalement. Ainsi, dans la séquence *transvenous lead systems*, l'indice de probabilité de co-occurrence du segment ADJ-N1 *transvenous lead* (électrode endocavitaire)

diminue fortement si les occurrences du verbe *lead* sont prises en compte et si celles de la séquence contenant le nom au pluriel (*transvenous leads*) ne le sont pas.

19. Quelques-uns d'entre eux (*negative, positive, progressive*) peuvent occasionnellement entrer dans la composition de termes polylexicaux. On a également relevé un contre-exemple concernant les catégories des adjectifs courts d'utilisation fréquente (*large artery atherosclerosis* [E] athérosclérose des gros troncs).

20. Encore plus curieux est le cas de *inferior vena caval obstruction* (obstruction de la veine cave inférieure) dans lequel le suffixe adjectival de l'anglais *-al* se substitue au suffixe adjectival féminin *-a* du latin *vena cava*.

21. Pour une réflexion plus détaillée sur les problèmes que causent ces structures coordonnées aux logiciels de traduction automatique, voir Maniez (2001).

22. L'interprétation de la structure de type N1 [N2 AND N3] *LDL composition and size* de l'exemple (4) fait intervenir les deux mécanismes : la composition et la taille d'une substance sont deux des caractéristiques essentielles d'une substance du corps humain faisant l'objet de prélèvements et de mesures. Les mots *composition* et *size* appartiennent au même champ sémantique, mais leur fréquence d'emploi en collocation dans la langue scientifique sert également à l'interprétation correcte de la structure syntaxique.

23. À titre indicatif, les moteurs de recherche consultés relèvent plus de 100 000 occurrences de ces collocations.

24. Il s'agit au demeurant d'un terme employé en psychologie et en sociologie, certains auteurs faisant une distinction entre dépression clinique et dépression sociale.

25. Le téléchargement de ce logiciel de catégorisation en parties du discours est gratuit et il est disponible à l'adresse suivante : <<http://www.inalf.cnrs.fr/WinBrill/winbrill.tele.html>>. Son utilisation nécessite toutefois la signature d'une convention avec l'INaLF

26. Les renseignements concernant ce concordancier sont disponibles sur la Toile à <<http://www.athel.com/mono.html>>.

27. Cette erreur est systématiquement reproduite par le logiciel pour les 298 occurrences de l'adjectif *novel* dans notre corpus.

28. On pourrait en fait subdiviser encore cette symbolisation en transformant [ADJ ADJ N] en [ADJ <ADJ N>], car un effet secondaire grave est un genre d'effet secondaire.

RÉSUMÉS

Les articles scientifiques contiennent une proportion élevée de groupes nominaux complexes à l'intérieur desquels la présence commune de prémodificateurs et de conjonctions de coordination est souvent à l'origine de structures ambiguës, et la littérature médicale ne fait pas exception à cette règle. Cet article étudie les divers mécanismes qui entrent en jeu chez les apprenants de l'anglais dans la compréhension des groupes nominaux complexes en anglais scientifique et milite pour la création d'une base de données bilingue exhaustive visant au recensement de ses unités polylexicales les plus courantes, qu'il s'agisse de termes complexes ou de collocations. À partir de l'étude de données tirées de corpus unilingues ou bilingues anglais-français, l'auteur se penche sur les problèmes de traduction générés par divers types de groupes nominaux, partant de cas relativement simples d'utilisation de prémodificateurs pour aborder

ensuite des structures beaucoup plus complexes faisant intervenir des noms ou des adjectifs en position de prémodificateurs à l'intérieur de groupes nominaux coordonnés.

Scientific articles contain a large proportion of complex noun phrases in which the common presence of modifiers and coordination conjunctions often gives rise to ambiguous structures, and medical literature is no exception to that rule. This article studies the various mechanisms at play in English learners' understanding of complex noun phrases in scientific English and argues for the creation of a comprehensive bilingual database designed for the inclusion of its most frequently used multi-word units, whether they are complex terms or collocations. Based on the study of monolingual as well as English-to-French bilingual corpora, the author examines the translation problems that are caused by various types of noun phrase structures, from relatively simple cases involving the use of modifiers to much more complex structures involving the use of both nouns and adjectives as modifiers within coordinated noun phrases.

INDEX

Mots-clés : anglais médical, coordination, corpus, groupe nominal, prémodificateur, traduction

Keywords : coordination, corpus, medical English, modifier, noun phrase, translation

AUTEUR

FRANÇOIS MANIEZ

François Maniez enseigne au département de LEA de l'Université Louis Lumière-Lyon 2. Il est directeur du CRTT (Centre de Recherche en Terminologie et Traduction) et membre de l'ATALA (Association pour le Traitement Automatique des Langues). Il s'intéresse principalement à la linguistique de corpus appliquée aux langues de spécialité, en particulier dans le domaine médical. francois.maniez@univ-lyon2.fr