

Extraction d'une phraséologie bilingue en langue de spécialité : corpus parallèles et corpus comparables

François Maniez

Volume 46, Number 3, septembre 2001

URI: <https://id.erudit.org/iderudit/003549ar>

DOI: <https://doi.org/10.7202/003549ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Maniez, F. (2001). Extraction d'une phraséologie bilingue en langue de spécialité : corpus parallèles et corpus comparables. *Meta*, 46(3), 552–563. <https://doi.org/10.7202/003549ar>

Article abstract

The extraction of terminology and phraseology is a task that is currently performed by tools that have proved their efficacy in monolingual environments. However, the task of extracting collocations or multi-word units and their translation equivalences in another language remains difficult, whether parallel or comparable corpora are used. This article examines a semi-automatic approach to such an extraction process using both types of corpora.

ÉTUDES TERMINOLOGIQUES ET LINGUISTIQUES

Extraction d'une phraséologie bilingue en langue de spécialité: corpus parallèles et corpus comparables

FRANÇOIS MANIEZ
Université Lyon-II, Lyon, France

RÉSUMÉ

L'extraction terminologique et phraséologique repose à l'heure actuelle sur des outils éprouvés en environnement unilingue. Toutefois, le repérage des collocations ou lexies composées (*multi-word units*) et de leur équivalent de traduction dans une autre langue, que ce soit à l'aide de corpus parallèles ou de corpus comparables, reste une tâche ardue. Cet article expose une approche semiautomatique de ce processus d'extraction grâce à l'utilisation de ces deux types de corpus.

ABSTRACT

The extraction of terminology and phraseology is a task that is currently performed by tools that have proved their efficacy in monolingual environments. However, the task of extracting collocations or multi-word units and their translation equivalences in another language remains difficult, whether parallel or comparable corpora are used. This article examines a semi-automatic approach to such an extraction process using both types of corpora.

MOTS-CLÉS/KEYWORDS

anglais médical, collocations, corpus bilingues, extraction automatique, terminologie

1. Introduction

Un grand nombre de chercheurs ont souligné la popularité grandissante de l'utilisation des corpus dans le domaines des études touchant à la traduction. Certaines expériences portant sur les outils d'aide à la traduction en langue de spécialité (Bowker 1998) ont montré que les corpus parallèles, outre leur utilité dans le domaine de la recherche, pouvaient avoir d'intéressantes applications en pédagogie. Le numéro de *Meta* récemment consacré aux diverses utilisations des corpus bilingues témoigne de l'intérêt croissant pour ce genre d'outils dans le domaine de la traduction.

La question de la représentativité de la langue traduite, par opposition à la langue produite par l'expression spontanée (Baker 1998), a amené un certain nombre de chercheurs à souhaiter l'utilisation conjointe des corpus parallèles et de corpus dits « comparables », constitués de textes originaux partageant une similitude de style et de contenu. L'utilisation de ces textes permet en effet d'avoir un accès à une langue dont on peut supposer que sa phraséologie sera plus riche que celle d'une langue traduite que d'aucuns nomment « *translationese* » pour en souligner le caractère non

idiomatique. Certains lexicographes (Teubert 1996) se sont penchés sur la pertinence de cette double approche, et la jugent favorablement dans le cadre de la lexicographie. Nous nous proposons dans cette étude d'étudier son utilité dans le cadre de la recherche de phrasèmes en langue de spécialité.

L'extraction terminologique est une discipline déjà ancienne qui a vu la mise au point d'un certain nombre d'outils au cours des dernières années, dont *Lexter* (Bourigault *et al.*, à paraître). L'extraction d'une phraséologie semble toutefois d'une automatisation plus difficile, en particulier dans le cas de phrasèmes non figés contenant des verbes, qui nécessite d'une part un processus préalable de lemmatisation et d'étiquetage grammatical, et d'autre part un repérage des éléments constitutifs du phrasème en contexte, ce dernier point étant sans doute la difficulté principale de la démarche d'extraction automatique. Ainsi, dans l'énoncé « *Results of selective gut decontamination have been mixed* », que nous avons déjà commenté ailleurs (Maniez, à paraître), la détection automatique du phrasème *mixed results* nécessite un étiquetage grammatical correct de *mixed* (sous la forme d'un adjectif et non pas d'un participe passé, forme adoptée par la majorité des programmes d'étiquetage dans ce cas précis), mais également le repérage d'une collocation dont les deux membres sont ici séparés par six autres mots. L'automatisation d'un tel repérage est incontestablement liée à la possibilité d'effectuer une désambiguïsation sémantique des mots polysémiques en fonction du contexte, domaine qui mobilise actuellement un nombre important de chercheurs (voir notamment Michiels, 1996 et Dini *et al.*, 1998).

Certains travaux actuellement en cours visent à élaborer un programme de repérage automatique de collocations équivalentes à partir de bitextes. Un logiciel de ce type a été mis au point par Langlois (1998) pour l'exploitation d'un corpus bilingue aligné constitué de huit années du *Hansard*. Les auteurs ont utilisé pour l'extraction une méthode d'aller-retour entre les deux parties du corpus visant à déterminer à l'aide de dictionnaires probabilistes la probabilité qu'un mot soit traduit par l'un des mots de la phrase alignée de l'autre partie du corpus, la démarche inverse permettant d'obtenir une liste de collocations potentielles contenant le mot de départ, ses synonymes ou d'autres mots appartenant au même champ sémantique. Ainsi, une liste obtenue à partir du mot *erreur* contient entre autres les collocations lui associant les verbes *commettre*, *corriger* et *rectifier*, mais également d'autres collocations contenant *bévue*, *échec*, ainsi que *correction*, *fautif* et *erroné*. L'application de la même technique à partir de la partie anglaise du corpus permet d'obtenir une liste de collocations dont les auteurs tirent une liste de celles qui contiennent les substantifs anglais de sens voisins (*blunder*, *error*, *flaw*, *goof*, *mistake*, *oversight*). La mise en correspondance des deux listes permet d'établir une liste d'équivalences possibles. Les auteurs soulignent toutefois que cette méthode ne produit pas une liste exhaustive dans la mesure où seuls les 200 mots les plus fréquemment employés dans les traductions des phrases contenant le mot de départ sont utilisés pour la deuxième phase du processus d'aller-retour.

2. Corpus

La recherche effectuée dans le cadre de cet article utilise deux types de corpus de langue de spécialité dans le domaine médical, le but étant de repérer certains équivalents de traductions entre le français et l'anglais.

On a tout d'abord eu recours à un corpus parallèle constitué de 58 articles du *Journal of the American Medical Association* et de leur traduction. Les articles originaux ont été extraits du CD-ROM *Internal Medicine 1993* sous la forme de fichiers de type texte. Leurs traductions, qui n'étaient pas disponibles sous forme électronique, ont été encodées à l'aide d'un logiciel de reconnaissance optique de caractères (les traductions sont celles de la version française du *Journal of the American Medical Association*). L'alignement du corpus a ensuite été effectué par les logiciels *Minimark* et *Multiconcord*. Nous ferons référence à ce corpus sous l'appellation MED_TRAD, et les équivalents de traduction des collocations concernées y seront soulignés pour en faciliter le repérage visuel. Le tableau 1 donne un exemple de présentation du corpus aligné.

TABLEAU 1

Présentation des traductions dans *Multiconcord*

Once <u>the diagnosis of NF2 is made</u> , relatives who are at risk should be screened for the disease.	Une fois <u>le diagnostic de NF2 posé</u> , la maladie doit être recherchée chez les membres de la famille à risque.
A negative gadolinium-enhanced MRI is accepted in current practice as effectively <u>excluding the diagnosis</u> of vestibular schwannoma.	Dans la pratique courante, on admet que la négativité de l'IRM au gadolinium permet raisonnablement <u>d'exclure le diagnostic</u> de schwannome vestibulaire.

On a par ailleurs utilisé deux corpus unilingues, l'un en anglais, l'autre en français. Le corpus anglais est composé de 100 articles tirés du CD-ROM cité plus haut et nous y ferons référence sous le nom MED_EN dans la suite de cet article. Le corpus français est celui des 26 articles décrivant les diverses spécialités de la médecine sur le site de l'Université Joseph-Fourier ([www-sante.ujf-grenoble.fr/SANTE/corpmmed/Corpus/corpus]), et nous y ferons référence sous le nom de MED_FR.

3. Utilisation du corpus bilingue

En nous concentrant sur toutes les collocations du type VERBE + NOM dans lesquelles le nom anglais était *diagnosis*, nous avons tout d'abord cherché à établir une liste d'équivalents de traduction en examinant les occurrences de ce mot dans la partie anglaise du corpus traduit (MED_TRAD) et en relevant les équivalents de traduction observés dans la partie française du corpus correspondant. Le nom *diagnosis* a été choisi en raison de sa fréquence d'emploi élevée dans la langue médicale. Par ailleurs, sa phraséologie semble peu étudiée dans les ouvrages consacrés aux collocations de la langue générale. À titre d'exemple, Benson, dans son ouvrage sur les collocations de l'anglais, ne mentionne pour l'entrée qui lui est consacrée que deux collocants verbaux, *make* et *confirm*.

Les résultats de cette première étape font apparaître qu'il est important de disposer d'un corpus bilingue de taille conséquente si l'on souhaite étudier la phraséologie, et tout particulièrement si l'on restreint l'étude à certaines constructions grammaticales. Alors que le nom sélectionné (*diagnosis*) est d'une utilisation très fréquente en langue médicale (196 utilisations dans la partie anglaise de MED_TRAD),

on ne relève que 43 occurrences d'emploi de celui-ci dans une collocation de type VERBE + NOM, soit moins du quart de la totalité des emplois. Enfin, si l'on examine les traductions françaises des énoncés contenant ces collocations, on s'aperçoit que 29 d'entre elles seulement contiennent un équivalent utilisant une forme verbale. Le tableau 2 résume ces résultats.

TABLEAU 2

Synthèse des équivalents de traduction dans MED_TRAD
(N.T. = non traduit par une structure verbale)

acquire	N.T.
agree on	N.T.
arrive at	arriver à
confirm	confirmer (3)
consider	N.T.
delay	retarder
describe	décrire
discuss	discuter
establish	N.T. (2) affirmer, établir, poser
exclude	exclure
expedite	préciser
facilitate	faciliter
indicate	faire évoquer
make	N.T. (2), effectuer, faire, parvenir à, poser (5), retenir
manifest	N.T.
mimimize	mimimiser l'éventualité de
miss	N.T.
overlook	N.T. (2)
receive	faire l'objet de
review	être une synthèse sur
secure	poser
suggest	N.T., suggérer
support	militer en faveur de
suspect	N.T.
tell	N.T.

Il est intéressant de se pencher sur les exemples d'absences d'équivalents de traduction de la forme VERBE+NOM en français. Plusieurs facteurs interviennent en effet.

Tout d'abord, on trouve le cas des collocations pour lesquelles une traduction littérale paraît impropre. Dans l'exemple (1), la collocation *acquire a diagnosis*, qui est d'occurrence rare, et qui présente de surcroît la particularité d'avoir pour sujet grammatical le mot *patient* (cette particularité étant gommée de la traduction, puisque le sujet grammatical implicite de la version française est le praticien), n'a pas été traduite littéralement. Ce réflexe d'évitement est fréquent chez le traducteur lorsque la traduction littérale ne semble pas possible et qu'il ne dispose pas d'un synonyme

pouvant convenir (« *obtenir* » semble pourtant une solution acceptable ici). Le changement d'agent implicite interdit ici de comptabiliser la collocation française « *faire un diagnostic* » comme équivalent de traduction pour *acquire a diagnosis*.

Exemple (1)

The 1991 Consensus Development Conference on Vestibular Schwannoma was convened to consider how patients can acquire an accurate diagnosis and to review the best options for management of this disease, including primary therapy, follow-up, and rehabilitation.

Une conférence de consensus sur le schwannome vestibulaire a été organisée du 11 au 13 décembre 1991 sur les points suivants : comment faire un diagnostic précis et quelles sont les meilleures options thérapeutiques, en particulier le traitement primaire, la surveillance et la réadaptation.

Un autre cas de figure est celui des verbes-supports, dont le faible contenu sémantique justifie souvent l'effacement dans la langue d'arrivée. Le verbe *establish* disparaît ainsi de la traduction française à deux reprises dans notre corpus :

Exemple (2)

The clinical condition may warrant urgent intervention to relieve the effusion and the hemodynamic effects of cardiac tamponade, and establishing the precise diagnosis may be a secondary goal.

L'état clinique justifie parfois une intervention urgente afin d'évacuer l'épanchement et de lever les conséquences hémodynamiques d'une tamponnade. Le diagnostic étiologique précis passe alors au second plan.

Cet effacement se produit également à deux reprises pour le verbe *make*. Dans l'exemple (3), on devrait plutôt parler d'une transposition, l'utilisation du nom « *methode* » servant d'équivalent à la locution verbale *make (sth) by*.

Exemple (3)

Viral RNA has been accepted as the current "gold standard" by which to make the diagnosis of hepatitis C infection.

La détection de l'ARN viral est actuellement considérée comme l'« étalon or » des méthodes de diagnostic de l'infection par le virus de l'hépatite C.

Les contraintes grammaticales contribuent également à ce phénomène d'érosion du nombre des collocations VERBE+NOM dans la traduction française. Dans l'exemple (4), l'impossibilité de traduire littéralement la structure interrogative *how well*, conjuguée à l'absence de traduction suffisamment brève pour *agree on*, amène le traducteur à une transposition vers le nom.

Exemple (4)

How well do physicians agree on the diagnosis of stroke type based on clinical data (precision), and how accurate is that diagnosis?

Quelle est la qualité de l'accord entre médecins sur le type de l'AVC en fonction des données cliniques (précision), et quelle est l'exactitude de ce diagnostic ?

Il existe également un certain nombre de verbes pour lesquels une traduction par une structure transitive est difficile (*miss, overlook, tell*), en particulier lorsque le verbe est employé à la voix passive. Les exemples (5), (6) et (7) nous en donnent l'illustration.

Exemple (5)

Another speculation pertaining to the underrecognition of granulomatous coronary arteritis is that an elderly patient who dies of a myocardial infarction may not undergo postmortem examination, and the diagnosis of coronary arteritis may be missed.

Autre explication possible de la méconnaissance diagnostique de la coronarite granulomateuse, un patient âgé décédant d'un infarctus du myocarde ne subira peut-être pas d'autopsie et le diagnostic d'artérite pourra alors passer inaperçu.

Exemple (6)

Unfortunately, because of the young age of patients with MINC, the correct diagnosis may be overlooked.

Malheureusement le jeune âge de ces patients conduit parfois à des erreurs diagnostiques.

Exemple (7)

The guidelines recommend that patients be told the diagnosis, prognosis, and treatment options, including costs, duration, and potential side effects.

Les patients devraient connaître le diagnostic, le pronostic et les options thérapeutiques, y compris leurs coûts, la durée et les effets indésirables potentiels des traitements.

La transposition de l'exemple (6) nous a par ailleurs amené à rechercher les occurrences de l'adjectif *diagnostique* dans la partie française du corpus bilingue et à vérifier les énoncés correspondants de la langue source afin de vérifier si l'adjectif est d'occurrence plus fréquente en français qu'en anglais. Cette hypothèse s'est trouvée partiellement confirmée par les faits : la transposition de *diagnosis* vers *diagnostique* correspond à 20% des cas d'emploi de *diagnosis* dans notre corpus. Toutefois, une vérification à partir de la partie anglaise du corpus nous montre que la transposition inverse (utilisation du nom français *diagnostic* pour traduire les emplois de l'adjectif anglais *diagnostic*) est encore plus fréquente (elle se produit dans 30% des cas). Dans de nombreux cas, la transposition ne semble pas résulter d'une difficulté à traduire littéralement le verbe, mais plutôt d'une affinité entre nom et adjectif qui n'a pas d'équivalent dans la langue cible. Ainsi, alors que des noms comme *approche*, *classification*, *critère*, *méthode*, *outil*, *problème*, *règle*, *technique*, *valeur* sont généralement traduits littéralement lorsqu'ils sont employés avec l'adjectif, un nombre important de collocations semblent ne pas supporter une traduction littérale. Parmi celles-ci, on compte les collocations anglaises utilisant les noms *accuracy*, *modalities*, *possibility*, *reevaluation*, *study* et *yield*. Inversement, un certain nombre de traductions font apparaître l'adjectif diagnostique là où l'anglais n'employait qu'un nom (*valeur diagnostique de l'examen clinique* pour *accuracy of the clinical examination*; *second avis diagnostique* pour *second opinion*; *test diagnostique individuel* pour *individual case testing*). Enfin, il est intéressant de noter que l'emploi de *diagnostic* comme adjectif attribut ne donne pas lieu à une traduction littérale, comme le fait apparaître l'exemple suivant :

Exemple (8)

It is interesting that skeletal biopsy was diagnostic in only 8% of the cases, though confirmatory in an additional 28% of cases.

Il est intéressant de souligner que la biopsie osseuse n'a permis le diagnostic que dans 8 % des cas mais qu'elle l'a confirmé dans 28 % de cas supplémentaires.

Dans l'exemple (9), la réactivation d'une structure passive se double d'une réduction qui occulte l'opposition des verbes *suspect* et *identify* dans l'énoncé original et empêche là aussi la comptabilisation d'équivalents de traduction :

Exemple (9)

The diagnosis of MINC can be suspected but not identified with any certainty at the time of presentation to the physician.

Le diagnostic d'IDM à coronaires saines n'est pas toujours évident d'emblée lorsque le patient consulte.

Enfin, il arrive tout simplement que le traducteur, pour des raisons stylistiques, choisisse une autre solution que la traduction littérale alors que celle-ci est possible. Il est intéressant de noter que là encore, c'est une structure passive de l'anglais qui donne lieu à cette transformation.

Exemple (10)

On the other hand, these rules must be broad enough to point out those presentations in which elements are missing but the diagnosis should be considered.

Les critères diagnostiques doivent permettre d'identifier non seulement les cas où les trois éléments sont présents mais aussi ceux qui peuvent correspondre à une anaphylaxie malgré l'absence de certains éléments.

On peut considérer l'information regroupée dans le tableau 2 comme peu probante : outre les nombreux cas d'absence de traduction par une collocation de type V+N en français, on relève un grand nombre de traductions littérales et deux verbes seulement (*establish* et *make*) pour lesquels il y a variation des équivalents. Quant aux autres traductions, il faudrait qu'elles soient confirmées par un nombre d'occurrences dépassant l'unité pour que leur recensement puisse comporter un intérêt lexicographique. Les équivalents proposés pour *indicate*, *minimize*, *receive*, *review* et *support* paraissent plausibles et pourraient être reproduits dans un corpus de taille supérieure, mais la traduction du verbe *expedite* par « préciser » semble plus atypique.

4. Utilisation des corpus comparables

Une recherche similaire a ensuite été effectuée en utilisant les deux corpus comparables, MED_EN et MED_FR. Le tableau 3 fournit les données statistiques concernant les deux corpus. Il est intéressant de noter qu'en dépit des tailles comparables du nombre total de mots (*tokens*) et du vocabulaire (*types*), *diagnostic* est 3 fois plus utilisé dans MED_FR que *diagnosis* ne l'est dans MED_EN. La composition du corpus est probablement à la source de ce déséquilibre, puisque le MED_FR est constitué uniquement de descriptions des diverses pathologies, alors que MED_EN, plus varié, contient un certain nombre d'articles décrivant les résultats d'études cliniques.

TABLEAU 3

Données statistiques et occurrences de DIAGNOSTIC / DIAGNOSIS dans les corpus monolingues

corpus	tokens	types	occurrences de <i>diagnostic</i> / <i>diagnosis</i>	collocations (tokens)	collocations (types)	collocations employées plus d'une fois
MED_FR	302802	29100	1023	151	47	21
MED_EN	306222	25701	367	71	35	13

On remarquera que le nombre d'occurrences trois fois plus élevé dans le corpus français ne produit pas un nombre de collocations beaucoup plus élevé dans le corpus français (21 collocations utilisés plus d'une fois contre 13 en anglais), ce qui peut être interprété comme le signe que le corpus français est de taille suffisante pour «épuiser» la phraséologie de *diagnostic*, les nouvelles occurrences du mot n'apportant pas de nouveaux exemples de phrasèmes au-delà d'une certaine taille critique. La différence peut également être due au fait qu'un plus grand nombre d'occurrences de MED_FR correspondent à une utilisation de *diagnostic* comme sujet grammatical.

Le tableau 4 dresse la liste des collocants de *diagnostic* dans le corpus MED_FR. Il est intéressant de noter que plus de la moitié des collocants (26 sur 47) n'ont qu'une seule occurrence sur l'ensemble du corpus, ce qui jette un doute sur la représentativité statistique de l'emploi des phrasèmes associés. Il faut également souligner que certaines formes verbales complexes (*faire évoquer, faire soupçonner*) n'ont pas été comptabilisées.

TABLEAU 4

Collocants de *diagnostic* dans MED_FR après lemmatisation

Verbes	F1	F2
confirmer	19	19
permettre	16	33
faire	16	154
évoquer	13	31
affirmer	11	13
poser	7	16
redresser	6	5
conduire à	4	28
éliminer	4	34
orienter	4	5
apporter	3	9
exclure	3	3
préciser	3	55
amener à	2	3
assurer	2	10
contribuer à	2	2
entraîner	2	49
rectifier	2	2
résumer	2	4
signer	2	109
suivre	2	11
aboutir à	1	1
affiner	1	1
aider à	1	4

Verbes	F1	F2
arriver à	1	2
associer	1	17
conforter	1	1
constituer	1	31
correspondre à	1	26
diffuser	1	4
discuter	1	18
équivaloir à	1	3
établir	1	6
justifier	1	21
méconnaître	1	1
nécessiter	1	9
obtenir	1	9
parvenir à	1	1
porter	1	12
précéder	1	6
reconsidérer	1	1
représenter	1	31
rétablir	1	3
retenir	1	19
s'assurer de	1	14
servir à	1	2
suffire à	1	11
TOTAL	151	849

F1 = Fréquence d'emploi en collocation

F2 = Fréquence totale d'emploi dans le corpus.

Le tableau 5 donne la liste des 71 collocations représentées pour les 367 occurrences de *diagnosis* répertoriées dans MED_EN. On remarque ici une répartition des fréquences très différente de celle du corpus français, même si là aussi les occurrences uniques constituent la majorité. Seuls trois verbes sont employés plus de trois fois, contre dix verbes dans le corpus MED_FR. Ici, la nécessité d'un corpus de taille supérieure se fait clairement sentir, ne serait-ce que pour recueillir d'autres occurrences de l'emploi des collocations à occurrence unique, qui constituent près des deux tiers de l'ensemble.

TABLEAU 5

Nombre d'occurrences des collocants de *diagnosis* dans MED_EN après lemmatisation

Verbe	Nombre
make	12
confirm	10
establish	7
facilitate	3
favor	3
overlook	3
qualify	3
agree on	2
arrive at	2
exclude	2
focus on	2
suggest	2
support	2
achieve	1
acquire	1
assist in	1
carry	1
delay	1

Verbe	Nombre
describe	1
discuss	1
disfavor	1
expedite	1
give	1
indicate	1
minimize	1
precede	1
prohibit	1
prove	1
provide	1
receive	1
relate to	1
resolve	1
revolutionize	1
tell	1
use	1

Afin de procéder à une synthèse des résultats obtenus dans les trois corpus, nous avons établi la liste des verbes anglais et de leurs équivalents de traduction tels qu'ils ont été répertoriés dans le corpus bilingue MED_TRAD. À l'aide des *synsets* de WordNet, nous avons effectué des regroupements de verbes aux sens apparentés, puis nous avons tenté d'affecter les verbes français présents dans MED_FR à l'une des 13 différentes catégories ainsi créées. Le tableau 6 résume le résultat de ces regroupements. Il va de soi que cette classification pourrait être réduite à un nombre inférieur de composants de base. Ainsi, certains des emplois de *support* et *confirm* pourraient être considérés comme proches de la synonymie. Par ailleurs, le degré de synonymie de verbes considérés comme équivalents mériterait d'être étudié plus en détail. Les fonctions lexicales d'Igor Mel'čuk permettraient sans doute dans certains cas une description plus précise des relations sémantiques entre ces différents synonymes. C'est sans doute le cas pour le verbe *signer*, proposé parmi les traductions de *confirm*, mais qui correspond au degré maximal de la confirmation dans un énoncé tel que :

« La palpation de la fosse iliaque droite réveille la douleur et provoque une défense abdominale qui signe le diagnostic. »

L'avantage de notre approche est de permettre dans certains cas la mise à la disposition du traducteur d'un ensemble d'équivalents de traduction là où une traduction littérale est souvent la règle (par exemple pour le verbe *confirm*), ou encore d'associer entre eux deux verbes qui ne sont pas donnés comme équivalents de traduction dans la plupart des dictionnaires bilingues non-spécialisés (*overlook* et *méconnaître*, par exemple). Il n'en reste pas moins que le nombre relativement faible d'équivalents de traduction dont l'usage a été effectivement observé (ceux qui figurent en gras dans la dernière colonne) font apparaître le besoin d'un corpus bilingue d'une taille supérieure.

TABLEAU 6

Collocants de *diagnosis*, sens et synonymes de WordNet et équivalents de traduction (les verbes en gras ont été utilisés dans le corpus bilingue pour traduire les verbes des première ou quatrième colonnes)

VERBE	SENS	SYNONYMES DE WORDNET	SYNONYMES DU CORPUS	ÉQUIVALENTS DE TRADUCTION
MAKE	6.	produce, make, create	provide	faire, évoquer, poser,
	7.	draw, make (make an estimate)		
	16.	make (perform or carry out, "make a decision") etc.		
CONFIRM	1.	corroborate, sustain, substantiate, support, affirm		confirmer, conforter, assurer, signer, s'assurer de
ESTABLISH	3.	prove, demonstrate, establish, show	prove, resolve	affirmer, préciser, établir
FACILITATE	1.	facilitate, ease, alleviate	assist in, expedite	faciliter, orienter, aider à
OVERLOOK	1.	overlook (look past, fail to notice)		méconnaître
	3.	neglect, omit, drop, miss, leave out, overlook, overleap		
FAVOR	1.	prefer, favor, favour		conduire à, amener à, contribuer à, entraîner
QUALIFY		qualify, restrict, modify		redresser, rectifier, préciser, affiner
AGREE ON	1.	agree, concur		
	7.	achieve harmony of opinion		
ARRIVE AT	1.	reach, attain, make, hit, arrive at, gain	achieve	arriver à, aboutir à, parvenir à
FOCUS ON	1.	center on, revolve around, revolve about, concentrate on, center		
EXCLUDE	1.	except, leave out, leave off, omit, take out	disfavor, prohibit	éliminer, exclure
SUGGEST	3.	suggest, intimate	indicate	évoquer, suggérer,
	5.	suggest, evoke, paint a picture		permettre, faire évoquer, conduire à
SUPPORT	5.	confirm, corroborate, sustain, substantiate, support, affirm		justifier

5. Conclusion

L'utilisation d'un corpus bilingue de traduction est une approche initiale utile qui permet d'effectuer certains appariements d'équivalents de traduction. Toutefois, dans la mesure où les traducteurs choisissent souvent la traduction littérale quand elle est possible et évitent de traduire certains verbes par de stricts équivalents syntaxiques afin de contourner certaines difficultés inhérentes à leur traduction, la partie du corpus bilingue qui correspond à la langue d'arrivée est parfois d'une moins grande richesse lexicale que celle correspondant à la langue de départ. La description doit donc être complétée par l'étude de corpus comparables, ou éventuellement par celles de corpus bilingues dans lesquels l'ordre langue de départ-langue d'arrivée serait inversé.

Par ailleurs, on peut se poser la question de savoir si l'approche adoptée ici, la recherche des collocations de type V+N en faisant porter la recherche initiale sur un nom, se justifie en langue de spécialité. Dans la mesure où la plus grande partie de la terminologie scientifique est constituée de noms, l'approche consistant à repérer les collocations V+N en faisant porter la recherche initiale sur les verbes paraît plus économique, l'établissement d'une liste finie de verbes utilisés en langue de spécialité semblant plus facile à réaliser de prime abord.

REFERENCES

- ATKINS, S., J. CLEAR and N. OLSTER (1992): "Corpus Design Criteria", *Literary and Linguistic Computing*, 7-1, pp. 1-16.
- BAKER, M. (1998): « Réexplorer la langue de traduction: une approche par corpus », *Meta*, 43, pp. 480-485
- BOWKER, L. (1998): "Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot study", *Meta*, 43, pp. 631-651.
- BOURIGAULT, D., C. CHODKIEWICZ & J. HUMBLEY (à paraître): "Making a workable glossary of a specialised corpus: term extraction and expert knowledge" In Bengt, Altenberg & Sylviane Granger (eds), *Lexis in Contrast*, Amsterdam, John Benjamins.
- CHURCH, KENNETH W. and William A. GALE (1991): "Concordances for Parallel Text", Paper presented at the Seventh Annual Conference of the UW Centre for the New OED and Text Research, *Using Corpora*, September 29-October 1, 1991, St. Catherine's College, Oxford, England
- CORMIER, M. (1991): « Traduction de textes destinés à des spécialistes: Approche pédagogique », *Meta*, 36-2/3, p. 440-447.
- DANIELSSON, P. and D. RIDINGS (1996): "Corpus and Terminology: Software for the Translation Program at Göteborgs Universitet or Getting Students to do the Work", *Proceedings of Teaching and Language Corpora 1996 (TALC'96)*, Lancaster, University of Lancaster (UCREL), pp. 57-67.
- DINI, L., V. DI TOMASO & SEGOND F. SEGOND (1998): "Word Sense Disambiguation with Functional Relations", Language Resource and Evaluation Conference, Granada, May 98.
- LANGLOIS, L. and P. PLAMONDON (1998): « Le repérage automatique de collocations équivalentes à partir de bitextes » In Fontenelle, Thierry, Philippe Hiligsmann, Archibald Michiels, André Moulin, Siegfried Theissen (eds), *Euralex'98: Proceedings of the Eighth Euralex International Congress Liège*, Université de Liège, pp. 175-186
- MANIEZ, F. (à paraître): "The Use of Electronic Corpora and Lexical Frequency Data in Solving Translation Problems." In Bengt, Altenberg & Sylviane Granger (eds), *Lexis in Contrast*, Amsterdam, John Benjamins.

- MCENERY, T. and A. WILSON (1993): "Corpora and Translation: Uses and Future Prospects", Technical Report from the Unit for Computer Research on the English Language (UCREL), University of Lancaster.
- MICHIELS, A. (1996) "An Experiment in Translation Selection and Word Sense Discrimination Using the Metalinguistic Apparatus of Two Computerized Dictionaries", DEFI Technical Report, University of Liège, 24 p. ([<http://engdep1.philo.ulg.ac.be/michiels/defi.htm>]).
- PETERS, C. and E. PICCHI (1998): "Bilingual Reference Corpora for Translators and Translation Studies", L. BOWKER, M. CRONIN, D. KENNY and J. PEARSON (Eds), *Unity in Diversity? Current Trends in Translation Studies*, Manchester, St. Jerome Publishing.
- SINCLAIR, J. (1991): *Corpus, Concordance, Collocation*, Oxford, Oxford University Press.
- TEUBERT, W. (1996): "Comparable of Parallel Corpora?", *International Journal of Lexicography* 9-3, pp. 238-264.
- VAN DOORSLAER, L. (1995): "Quantitative and Qualitative Aspects of Corpus Selection in Translation Studies", *Target*, 7-2, pp. 245-260.