HAL
open science

La phraséologie transdisciplinaire des écrits scientifiques : quelles unités enseigner à un public d'étudiants non francophones ?

Catherine Carras

▶ **To cite this version:**

Catherine Carras. La phraséologie transdisciplinaire des écrits scientifiques : quelles unités enseigner à un public d'étudiants non francophones ?. 9° journées scientifiques Lexicologie, Terminologie, Traduction, La notion d'unité en Sciences du Langage, Université Paris 13, Sep 2011, Paris, France. hal-01985747

HAL Id: hal-01985747

https://hal.univ-grenoble-alpes.fr/hal-01985747v1

Submitted on 2 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Catherine Carras
Université Stendhal Grenoble 3
LIDILEM
Grenoble
France
Catherine.Carras@u-grenoble3.fr

La phraséologie transdisciplinaire des écrits scientifiques : quelles unités enseigner à un public d'étudiants non francophones ?

Résumé

Cet article se situe dans le cadre de la didactique du lexique en Français Langue Etrangère. Nous nous intéressons à la phraséologie scientifique transdisciplinaire, caractéristique des écrits scientifiques. Cette recherche a été effectuée dans le cadre d'un projet de description linguistique et terminologique de corpus d'écrits scientifiques, dans l'objectif de création d'outils d'aide à la rédaction d'écrits universitaires pour publics non francophones. A partir d'un corpus d'articles scientifiques du domaine de la médecine, nous avons extrait les collocations V+N ayant un caractère transversal. Nous nous interrogeons ensuite sur les propriétés de combinatoire et la délimitation des unités observées.

Mots clés : Didactique du lexique ; phraséologie ; écrits scientifiques.

Introduction

La question de l'unité terminologique intéresse de nombreux domaines de la linguistique. La présente recherche s'inscrit dans la perspective de la didactique du lexique en Français Langue Etrangère. Nous nous intéressons particulièrement aux propriétés de combinatoire des unités lexicales, propriétés qui limitent la capacité de ces unités à se combiner avec d'autres unités et qui ne peuvent être déduites ni de leur signifié ni de leur signifiant (Polguère 2003 : 37).

Cette combinatoire, pourtant indispensable à la maîtrise d'une langue, fait partie des aspects lexicaux que le public d'étudiants non francophones avec lequel nous travaillons a le plus de mal à acquérir. Dans le cadre d'un projet d'élaboration d'outils d'aide à la rédaction d'écrits universitaires pour les publics non francophones, nous nous intéressons à la phraséologie des écrits scientifiques. Nous ne nous attacherons pas à la terminologie propre à un domaine, mais aux collocations spécifiques aux écrits scientifiques qui traversent l'ensemble des disciplines. Nous parlerons donc de phraséologie transdisciplinaire (Coxhead *et al.*, 2000).

Cette phraséologie transdisciplinaire nous intéresse pour deux raisons. Tout d'abord pour des raisons linguistiques, en ce que cette phraséologie transdisciplinaire est une phraséologie de genre, qui caractérise les écrits scientifiques. Ensuite, pour des raisons didactiques, car il a été constaté que les étudiants non francophones avaient des difficultés à utiliser cette phraséologie spécifique aux écrits universitaires scientifiques. Il a ainsi été observé que ce public n'ignore pas l'existence de cette phraséologie, mais qu'il se trouve en difficulté face à la combinatoire des différentes unités lexicales (Cavalla 2008 et 2010). Or cette combinatoire reste peu enseignée. Alors que les collocations sont très présentes dans les écrits scientifiques (on dit *émettre une hypothèse* et non pas *dire une hypothèse*, par exemple), et que leur emploi correct marque la maîtrise de la langue (l'emploi des collocations revêt un caractère normatif), elles ne sont généralement pas enseignées en tant que telles, l'accent étant souvent mis sur les unités isolées plutôt que sur leur combinatoire.

La question qui va se poser est la suivante : s'il importe de ne plus travailler à partir d'unités lexicales isolées, comment déterminer la taille de l'unité phraséologique ? Comment repérer et délimiter les collocations qu'il convient d'enseigner à ce public ? Dans un premier temps, notre étude portera sur les collocations V+N, car d'une part ce type de construction présente une fréquence importante dans les corpus étudiés, et revêt d'autre part un caractère fondamental dans la production de l'écrit, dans la mesure où ces collocations participent à la cohérence structurelle du texte.

1. Cadre et objectifs de l'observation

Cette recherche se situe dans le cadre d'un projet PPF (Plan Pluri-Formation) intitulé FULS[i]. Ce projet avait deux objectifs distincts et complémentaires: une description linguistique et terminologique de corpus d'écrits scientifiques (lexique et énonciation), et une réflexion à visée didactique pour l'enseignement en classe de FLE/FOU[ii]. Cet enseignement vise notamment l'aide à l'écriture universitaire aux niveaux Master et Doctorat pour les étudiants non francophones venus étudier en France. L'accent sera mis dans un premier temps sur la combinatoire syntaxique puis sur les associations lexicales privilégiées au sein de ces constructions, et enfin sur leur fonction rhétorique dans les discours concernés. Les analyses linguistiques ont permis de dégager certaines caractéristiques du lexique transdisciplinaire présent dans ces écrits. Dans le cadre des écrits scientifiques, nous avons exploité deux corpus : Scientext[iii] et un corpus d'écrits de la médecine, CRTT-MED, élaboré au CRTT[iv]. Cet article présente une analyse à partir du corpus des écrits de la médecine uniquement.

L'objectif de la recherche présentée ici est d'extraire, dans le corpus CRTT-MED, des collocations appartenant au discours scientifique transdisciplinaire.

Ce lexique transdisciplinaire des écrits scientifiques peut être défini comme :

> *Le lexique mis en œuvre dans la description et la présentation de l'activité scientifique (…). Ce lexique peut être considéré comme un lexique de genre, n'intégrant pas la terminologie du domaine, mais renvoyant aux concepts mis en œuvre dans l'activité scientifique ; (Tutin 2005 : 248)*

Nous nous intéresserons au lexique dans ses aspects combinatoires, puisque notre étude porte sur des collocations. Les collocations extraites du corpus décrit ci-dessous seront analysées d'abord à partir de critères de fréquence, puis essentiellement à partir de critères de formes. L'observation se concentrera sur la combinatoire V + N. L'objectif final du projet FULS étant d'élaborer des outils d'aide à la rédaction pour public non francophone, une attention particulière sera accordée au différents types de co-occurrents verbaux rencontrés pour une même base lexicale, et au fonctionnement syntaxique de ces collocations.

2. Méthodologie

Notre observation a été réalisée à partir du corpus CRTT-MED. Ce corpus, partie française d'un corpus comparable élaboré par le CRTT, rassemble des articles de recherche en français dans le domaine médical, toutes disciplines confondues.

Nous avons travaillé à partir d'une sélection déjà effectuée (grâce au logiciel AntConc[v]) d'une liste de collocations Verbe + déterminant + Nom, organisée par fréquence décroissante. Afin de repérer et dégager de cette sélection les collocations transdisciplinaires objets de notre étude, l'idée de départ était d'opérer un codage à trois valeurs :

1. les collocations proprement terminologiques, caractéristiques des discours de la médecine
2. les collocations connotées médicalement
3. les collocations non connotées médicalement mais appartenant au discours scientifique, qui constitueraient des collocations transversales

Les collocations proprement terminologiques ont été préalablement éliminées de la liste sur laquelle nous avons travaillé puisque l'objectif de ce pan du projet – au plan didactique – était de s'arrêter sur les collocations transdisciplinaires et non pas sur la terminologie propre au domaine du corpus.

Dans un premier temps, et afin de définir, à partir des résultats initiaux, des pistes de travail qui nous permettront de poursuivre cette investigation, nous avons décidé de travailler à partir d'un échantillon, et de ne prendre en compte que les 500 premières collocations de cette liste. Ces collocations non lemmatisées présentent une fréquence supérieure ou égale à 12 (avant lemmatisation).

C'est donc à partir de cet échantillon que nous avons tenté d'opérer la distinction entre « collocations connotées médicalement » et « collocations non connotées médicalement ». Pour ce faire, nous nous sommes concentrée sur les noms, les verbes assumant la plupart du temps la fonction de verbe-support. Nous avons donc éliminé de la liste les collocations contenant des noms :

- appartenant clairement au domaine médical tels que *carcinome, cancer, asthme*
- connotés médicalement tels que *diagnostic, traitement, patient*

Nous nous sommes interrogée sur la pertinence, dans l'optique de notre recherche, des collocations contenant le mot *risque* ; nous avons décidé de les conserver dans un premier temps, mais nous verrons ci-dessous que les résultats obtenus confirment nos doutes initiaux.

Nous avons ensuite éliminé des collocations qui nous semblaient clairement connotées médicalement, même si le nom entrant dans leur construction n'est pas à proprement parler un terme médical, telles que *prévenir le rejet* ou *améliorer l'état*. Ont également été retirées des séquences qui ne constituent pas des collocations, telles que *est le seul*, *est le suivant*, *montre le tableau*, *a un aspect*.

Ce tri a été effectué manuellement, ce qui est évidemment chronophage et pose la question du traitement d'un corpus plus important et non plus d'un échantillon.

3. Résultats

Après le filtrage détaillé ci-dessus et lemmatisation, nous avons obtenu une liste de 75 collocations différentes. Ces premiers résultats, obtenus rappelons-le à partir de l'échantillon des 500 premières collocations non lemmatisées, nous montrent une liste de noms très « communs », alors que les verbes présentent plus de variété. Il semble donc que le plus intéressant soit la combinatoire de ces différents éléments.

Les collocations les plus fréquentes dans l'échantillon de corpus observé sont :

- *faire l'objet* (1225 occurrences)
- *être un facteur* (792)
- *jouer un rôle* (442)
- *diminuer le risque* (268)
- *avoir un effet* (256)

Plusieurs remarques peuvent être faites à ce stade de l'analyse :

Ces collocations, si elles apparaissent effectivement dans les écrits scientifiques, appartiennent également aux discours non scientifiques, et l'on peut contester leur pertinence, dans le cadre de notre étude, considérant que l'on se situe ici à la limite de la phraséologie des écrits scientifiques. Toutefois, si l'on considère une collocation comme une co-occurrence statistiquement saillante de deux éléments lexicaux, ces associations V + N correspondent à ce critère, et sont statistiquement significatives dans les écrits scientifiques. On peut ici faire la remarque que les phénomènes les plus intéressants ne se situent pas toujours dans les hautes fréquences. De même, si l'on considère que la phraséologie des écrits disciplinaires est une phraséologie de genre, ces collocations en font indéniablement partie. Et si nous gardons à l'esprit le public final visé par cette étude, à savoir les

étudiants non francophones, ces collocations, très familières pour des natifs, ne font pas obligatoirement partie du stock lexical des non natifs, surtout si l'on prend en compte la combinatoire.

3.1. Noms

Les 75 collocations observées sont construites avec 19 noms différents. Nous les présentons dans le tableau 1, classés par « productivité », c'est-à-dire par le nombre de collocations différentes dans lesquels ils entrent :

nom	Nombre de collocations différentes où il apparaît	Fréquence cumulée
Risque	24	1590
Intérêt	7	549
Rôle	6	657
Effet	5	410
Problème	5	296
Avantage	3	235
Bénéfice	3	54
Facteur	3	253
Objet	3	1293
Résultat	3	144

Argument	2	73
Lien	2	49
Objectif	2	75
Point	2	192
Elément	1	24
Préalable	1	19
Seuil	1	14
Signe	1	83
Spectre	1	13

Tableau 1 : noms classés par productivité

Nous voyons donc que *risque* figure en tête de cette liste, car il peut se combiner avec 24 verbes différents pour former des collocations. Quant à *objet*, si sa fréquence cumulée est importante, alors qu'il n'entre que dans trois collocations différentes, c'est que la collocation *faire l'objet* présente une fréquence très élevée. Le classement par le nombre de collocations dans lesquels ces éléments entrent plutôt que par leur fréquence absolue nous semble donc un critère de classement pertinent. En effet, le public d'étudiants non francophones que nous visons a besoin de connaître la combinatoire des unités lexicales, et d'avoir à sa disposition un panel d'expressions lui permettant de rédiger des écrits avec une précision lexicale satisfaisante.

Si nous observons de plus près les collocations dans lesquelles entrent les noms les plus productifs, voici ce que nous pouvons constater :

- Risque :

    a) *augmenter le risque / réduire le risque / diminuer le risque/ réduire le risque / limiter le risque* /etc.

b) prendre le risque / courir le risque

c) comporter un risque / présenter un risque / exister un risque Les collocations de la première série confirment nos doutes initiaux quant à la pertinence de conserver le mot *risque*. En effet, même si ce mot ne constitue pas un terme à proprement parler, il est clair que ces collocations sont connotées médicalement. La dernière série est par contre plus intéressante pour notre étude, car ces collocations sont plus transversales et applicables dans le contexte des écrits scientifiques.

- Intérêt :

a) avoir un intérêt

b) *démontrer l'intérêt / présenter un intérêt / souligner l'intérêt / confirmer l'intérêt*

Contrairement aux exemples précédents autour du mot *risque*, ces collocations sont éminemment transversales, peu spécialisées mais néanmoins caractéristiques des écrits scientifiques.

- Rôle

a) Jouer un rôle

b) Avoir un rôle

c) *Montrer le rôle / confirmer le rôle / souligner le rôle / évaluer le rôle*Se pose ici la question de la pertinence de ces collocations dans le cadre de cette étude, on peut considérer que l'on se situe ici à la limite de la phraséologie, le mot *rôle* et ses verbes associés étant également très fréquents dans les discours standards, hors du champ des écrits scientifiques. Toutefois, si nous prenons en compte à la fois le critère statistique, et le public non francophone auquel nous nous adressons, elles peuvent avoir leur place dans un outil d'aide à la rédaction.

*3.2. Verbes*Les collocations observées se construisent autour de 43 verbes différents. Ce nombre plus élevé par rapports aux noms montre qu'un même nom peut se combiner avec une variété de verbes qui peut être importante, en ce qui concerne les noms les plus fréquents, et comme nous le verrons, c'est l'emploi du verbe le plus approprié en fonction du contexte qui conditionne la précision lexicale.

verbes	Nbre de collocations différentes où il apparaît	Fréquence totale
Etre	7	1189
Avoir	5	909

Constituer	4	171
Evaluer	4	197
Montrer	4	123
Présenter	3	96
Atteindre	2	46
Confirmer	2	50
Démontrer	2	62
Exister	2	118
Faire	2	1359
Représenter	2	78
Rester	2	35
Souligner	2	84
Accroître	1	27
Améliorer	1	22
Apporter	1	40
Augmenter	1	220

Comporter	1	33
Compter	1	13
Courir	1	39
Diminuer	1	268
Donner	1	14
Elargir	1	13
Estimer	1	20
Etablir	1	25
Etudier	1	39
Eviter	1	38
Exercer	1	41
Exposer	1	49
Influencer	1	13
Jouer	1	442
Limiter	1	56
Majorer	1	58

Mettre	1	58
Minimiser	1	25
Obtenir	1	43
Offrir	1	12
Poser	1	183
Prendre	1	18
Prévenir	1	33
Réduire	1	174
Résoudre	1	20

Tableau 2 : verbes présentés par productivité

On peut considérer que ces résultats comportent une part de bruit, ou tout au moins font apparaître des éléments peu marquants. En effet, des verbes comme *réduire*, *augmenter* ou *diminuer* présentent une fréquence cumulée importante car ils se combinent avec *risque* ; quant à *faire* et *jouer*, leur fréquence s'explique par leur combinaison avec *objet* (*faire l'objet*, première collocation par la fréquence) et *rôle* (*jouer un rôle*, troisième par la fréquence). De même, les verbes figurant en tête de liste sont, sans grande surprise, *être* et *avoir*. Cela conforte en partie l'idée déjà énoncée selon laquelle les phénomènes les plus intéressants ne se situent pas forcément dans les fréquences les plus élevées.

Ce qui nous semble beaucoup plus intéressant à observer, c'est qu'un nombre nettement plus important de verbes (29 sur les 43 de la liste ci-dessus) que de noms (seuls 5 sur 19) n'apparaissent que dans une seule collocation. L'intérêt pour un non francophone lors de la rédaction d'écrits scientifiques réside donc dans sa connaissance de la combinatoire entre un même nom, porteur d'une signification que l'on peut qualifier de générale (comme *effet* ou *avantage*, par exemple) et une série de verbes qui lui permettront de préciser sa pensée ou d'éviter les répétitions (nous verrons des exemples ci-dessous).

Si l'on s'attarde sur les trois verbes les plus fréquents, on constate qu'ils sont utilisés dans des collocations très générales :

- *être : être un objectif / être un facteur / être l'objet / être un préalable / être un signe / être le résultat / être le risque / être un signe*
- *avoir : avoir l'avantage / avoir un effet / avoir un intérêt / avoir un risque / avoir un rôle*
- *constituer : constituer un argument / constituer un élément / constituer un facteur / constituer un problème*

Il est par contre intéressant d'observer que les verbes *être* et *avoir* peuvent avoir des substituts plus précis dans les discours scientifiques : si les collocations construites avec *être* et *avoir* sont effectivement présentes, et comme nous venons de le voir, avec des fréquences élevées, d'autres associations sont également possibles et actualisées dans le corpus :

- *être un argument / constituer un argument*
- *être un élément / constituer un élément*
- *être un facteur / constituer un facteur*
- *avoir l'avantage / présenter l' avantage / offrir l'avantage*
- *avoir un effet / exercer un effet*
- *avoir un intérêt / présenter un intérêt / démontrer l'intérêt*
- *faire un lien / établir un lien*

Les scripteurs non natifs auront tendance, dans leurs écrits, à utiliser les collocations construites avec *être* et *avoir* « par défaut », par manque de stock lexical, or le fait d'avoir à leur disposition un panel de collocations utilisant des verbes différents en combinaison avec un même substantif leur permettrait d'une part de varier leur expression et d'éviter les répétitions, et d'autre part d'être plus précis.

Il convient à présent de s'interroger sur les possibles applications didactiques de ces résultats, dans la perspective d'élaboration d'outils d'aide à la rédaction pour des étudiants non francophones.

4. Applications didactiques

Un certain nombre d'interrogations sont soulevées à ce stade de l'observation, dans la perspective d'une exploitation didactique de ces résultats. Elles concernent le point de départ à adopter pour l'enseignement / apprentissage de ces collocations, la délimitation des unités lexicales, et la lemmatisation.

4.1. Verbes ou noms ?

Tout d'abord, et cette questions s'est posée dès le départ, avant l'étape de sélection des collocations, faut-il se concentrer sur les verbes ou sur les noms ? Il va de soi que l'accent sera mis sur la combinatoire des différentes unités, mais dans le cadre d'une didactisation des collocations, il est important de constater que les étudiants non francophones avec lesquels nous travaillons partent en général du nom, et qu'ils tentent ensuite de lui associer un verbe (avec plus ou moins de succès), verbe qui sera *être*, *avoir*, ou *faire*, qui font le plus souvent fonction de verbes support « passe partout », lorsqu'ils n'ont pas de verbe plus précis à leur disposition.

Rappelons que dans ces collocations, on considère que le nom constitue la base, et le verbe le collocatif. La démarche des locuteurs non francophones correspond donc à une certaine logique linguistique. Ainsi, « on considère que la base contrôle la collocation car, du point de vue du locuteur, c'est le collocatif qui est choisi en fonction de la base, et non l'inverse » (Polguère 2003 : 135). Ces locuteurs appliquent donc ce principe, et « piochent » dans leur vocabulaire disponible un verbe leur permettant de mettre en discours la base qu'ils souhaitent actualiser.

Du point de vue du locuteur non francophone, les verbes choisis remplissent donc la fonction de verbe-support, fonction qui est d'actualiser la base, de la faire fonctionner dans la phrase. Ces locuteurs ont tendance à utiliser des verbes sémantiquement vides (*être, avoir, faire*) qui ne rajoutent aucun sens à la base. Or, comme le montrent les collocations extraites du corpus, les verbes utilisés comme collocatifs ne peuvent généralement pas être considérés comme sémantiquement vides, ils introduisent des nuances d'expression, et l'alternance entre ces différents verbes permet d'éviter les répétitions, d'enrichir le style et surtout de se rapprocher de l'écriture d'un natif..

Or, dans la gamme d'ouvrages didactiques centrés sur le lexique ou d'aides à la rédaction pour publics non francophones existant sur le marché, la plupart ne didactisent pas ces éléments combinatoires et font abstraction de leur présence dans les discours. Les mots simples, isolés, sont souvent mis à l'honneur.

Les activités didactiques d'enseignement / apprentissage de ces collocations devront donc prendre en compte ces aspects. Ainsi, l'élaboration d'outils interactifs permet de travailler réellement sur la combinatoire (possibilité de travailler indifféremment à partir des noms ou des verbes, suivant les options choisies par les apprenants, par exemple) ; faire travailler les apprenants à partir de corpus permet également de leur faire prendre conscience de cette combinatoire. Une plate-forme a été élaborée dans le cadre du projet FULS à cet effet (plate-forme qui sera prochainement accessible en ligne).

4.2 Délimitation des unités

La question de la délimitation des unités lexicales se pose dans divers champs de la linguistique, en particulier en terminologie, traitement automatique des langues et didactique du lexique. Les collocations sur lesquelles nous avons travaillé, collectées de façon automatique, sont ainsi, selon nous, incomplètes à droite. Des collocations telles que *être un facteur, être l'objet, être un signe, présenter l'avantage* sont incomplètes pour des locuteurs non natifs si on ne précise pas qu'elle sont suivies de la préposition *de,* idem pour la préposition *à* dans des collocations comme *être un*

préalable. Mais dans ce cas, où fixer la limite à droite de ces collocations ? Doit-on également inclure les compléments privilégiés, statistiquement les plus souvent associés à ces collocations V + N ?

De même, doit-on repousser la limite à gauche de ces collocations, en collectant les sujets les plus fréquemment associés ?

Si nous considérons que l'objectif de notre étude est de recenser des collocations transdisciplinaires, caractéristiques du discours scientifique et non pas d'un discours lié à une spécialité particulière (ici la médecine), il nous semble évident que les compléments et sujets privilégiés dans les textes dépendent directement du domaine de spécialité en question. Repousser les limites à droite et à gauche risquerait de gommer le caractère transdisciplinaire de ces collocations. En revanche, intégrer dans la collocation la préposition nous semble parfaitement justifié dans le cadre d'un dispositif destiné à des apprenants non francophones. Rappelons également que nous avons choisi de travailler à partir d'une liste de collocations V + dét. + N, la simple combinatoire V + N ne nous semblant pas suffisante. Ainsi, il est primordial de considérer *être l'objet* comme une unité différente de *être un objet*, de même en ce qui concerne *présenter l'avantage* et *présenter un avantage*.

4.3.Lemmatisation

La lemmatisation pose également un certain nombre de questions, en particulier compte tenu du type de public que vise notre étude. En effet, elle gomme certaines spécificités syntaxiques. Ainsi, elle occulte le caractère très fréquent voire invariable de certains emplois : passé composé, forme passive, tournures impersonnelles, etc. Si l'on souhaite enseigner à des apprenants non francophones à manipuler, de façon active, en production, ces collocations, il convient de rétablir ces emplois dans les séquences didactiques (Tutin 2007). Toutefois, ceci repose la question de la délimitation des unités, car si l'on s'intéresse aux propriétés syntaxiques, au fonctionnement discursif, on ne travaille plus sur des unités lexicales mais sur une unité de discours beaucoup plus vaste. La lemmatisation est une étape indispensable lors de la collecte, elle permet de repérer les fréquences réelles des unités qui sinon se retrouvent « éparpillées », mais un retour au contexte est indispensable lors de l'élaboration des outils didactiques.

Une étude du contexte permet ainsi d'analyser, et donc de prévoir une exploitation didactique, des caractéristiques syntaxiques de ces collocations. Comme nous l'avons signalé ci-dessus, les ouvrages didactiques ou les dictionnaires présentent ces collocations sous une forme canonique (Tutin 2005), et ne fournissent donc pas toujours le mode d'emploi de ces collocations, pourtant indispensable si l'on se situe dans une perspective de production écrite.

Conclusions

Les collocations V+det+N extraites et analysées ici appartiennent à la phraséologie transdisciplinaire des écrits scientifiques. A ce titre, il convient de les inclure dans un outil d'aide à la

rédaction d'écrits universitaires pour étudiants non francophones, car leur maîtrise est indispensable à la production d'écrits respectant les normes du genre.

Mais rappelons que les résultats présentés et analysés ici ont été obtenus à partir d'un échantillon puisque nous avons travaillé à partir des 500 premières collocations non lemmatisées. Il importe donc de poursuivre cette étude sur la suite de la liste, tout d'abord parce que la lemmatisation fait apparaître des fréquences supérieures, et parce que, comme nous l'avons souligné à plusieurs reprises, les phénomènes les plus intéressants ne se situent pas toujours dans les hautes fréquences.

De même, il faut à présent confirmer le caractère transdisciplinaire des collocations observées dans ce corpus constitué de textes du domaine de la médecine. Pour ce faire, une comparaison avec le corpus Scientext pourrait permettre de valider leur aspect transversal.

L'interrogation principale qui demeure à ce stade de la recherche concerne la délimitation des unités : où fixer les limites à droite et à gauche des collocations, dans la perspective d'élaboration d'outils d'aide à la rédaction ? Compte tenu du fait qu'elles ne prennent tout leur sens qu'en contexte, comment présenter ces collocations dans leur mode de fonctionnement ? Comment transmettre leurs propriétés syntaxiques mais aussi sémantiques ?

Bibliographie

CARRAS Catherine, 2006, « Quelles réponses didactiques pour l'intégration des étudiants étrangers dans les cursus universitaires français? Un cas spécifique d'enseignement / apprentissage du FLE en milieu homoglotte», dans D. Abry et M. Fiévet (coord) *L'enseignement apprentissage du français langue étrangère en milieu homoglotte : spécificités et exigences*. PUG, ADCUEFE, pp33-45.

CAVALLA Cristelle. et CARRAS Catherine, 2009, « Développement d'outils pour le suivi linguistique des apprenants dans les universités grenobloises », dans Bérard E. et Borg S. *Terres de FLE*, pp 63-73.

CAVALLA Cristelle, 2008, "Les écrits universitaires : Un français spécifique pour les apprenants étrangers", dans Olivier Bertrand et Isabelle Schaffner (eds), *Apprendre une langue de spécialité : enjeux culturels et linguistiques*, Paris : Editions Ecole Polytechnique, 93-104.

CAVALLA Cristelle, 2010, Méthodologie d'apprentissage de l'écrit universitaire, dans Parpette C. et Mangiante JM, *Faire des études supérieures en français, FDLM, R&A* n°47, pp153-161.

COXHEAD Averil et NATION Paul, 2000, "The specialised vocabulary of English for academic purposes", dans *Research Perspectives on English for Academic Purposes*, Cambridge: Cambridge University Press, 252-267.

EURIN-BLAMET Simone et HENAO DE LEGGE Martine, 1992, *Pratiques du français scientifique*, Paris: Hachette.

MEL'CUK Igor, 1993, "La Phraséologie et son rôle dans l'enseignement-apprentissage d'une langue étrangère", dans *Etudes de linguistique appliquée* n°92, 82-113.

MEL'CUK Igor, CLAS André et POLGUÈRE Alain, 1995, *Introduction à la lexicologie explicative et combinatoire*, Louvain la Neuve: Duculot.

PHAL André, 1971, *Vocabulaire général d'orientation scientifique (V.G.O.S.) - Part du lexique commun dans l'expression scientifique*, Paris: Didier, Crédif.

POLGUERE Alain, 2003, *Lexicologie et sémantique lexicale*, Montréal : Presses de l'Université de Montréal.

TOLAS Jacqueline, 2004, *Le français pour les sciences*, Grenoble: Presses Universitaires de Grenoble.

TUTIN Agnès, 2005, "Collocations du lexique transdisciplinaire des écrits scientifiques : annotation et extraction des propriétés linguistiques dans la perspective d'une application didactique", dans *La phraséologie dans tous ses états – Cahiers de l'Institut de Linguistique de Louvain*, 31.2-4, Louvain La Neuve (Belgique), pp247-262.

TUTIN Agnès, 2007, « Modélisation linguistique et annotation des collocations : une application au lexique transdisciplinaire des écrits scientifiques », dans S. Koeva, D. Maurel & . Silberztein *Formaliser les langues avec l'ordinateur : de Intex à Nooj*. Besançon, Presses Universitaires de Franche Comté, pp 189-215.

---

[i] Formes et Usages des Lexiques Spécialisés (2007-2010) ; Lidilem, Grenoble, C.Cavalla : http://scientext.msh-alpes.fr/fuls/

[ii] Français sur Objectif Universitaire

[iii] ANR ; Lidilem, Grenoble, A.Tutin et F.Grossmann : http://scientext.msh-alpes.fr/scientext-site/spip.php?article1. Corpus de textes scientifiques consultable en ligne d'environ 4,8 millions de mots en français, de différentes disciplines issues des sciences humaines, expérimentales et sciences pour l'ingénieur.

[iv] Centre de Recherches en Terminologie et Traduction ; Lyon2, F.Maniez : http://recherche.univ-lyon2.fr/crtt/

[v] Logiciel permettant notamment l'élaboration de concordanciers, développé par L. Anthony www.antlab.sci.waseda.ac.jp/