



Linx

Revue des linguistes de l'université Paris X Nanterre

68-69 | 2013

Corpus et apprentissage du français

L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE

Sonia Di Vito



Édition électronique

URL : <http://linx.revues.org/1519>

DOI : 10.4000/linx.1519

ISSN : 2118-9692

Éditeur

Université Paris Ouest – département
Sciences du langage

Édition imprimée

Date de publication : 19 novembre 2013

Pagination : 159-176

ISSN : 0246-8743

Référence électronique

Sonia Di Vito, « L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE », *Linx* [En ligne], 68-69 | 2013, mis en ligne le 19 novembre 2015, consulté le 01 octobre 2016. URL : <http://linx.revues.org/1519> ; DOI : 10.4000/linx.1519

Ce document est un fac-similé de l'édition imprimée.

© Tous droits réservés

L'utilisation des corpus dans l'analyse linguistique et dans l'apprentissage du FLE

Sonia Di Vito, Università della Tuscia

1. Introduction

Dans le domaine de l'apprentissage des langues, on assiste depuis quelques décennies à l'essor des nouvelles technologies qui remplacent de plus en plus les supports traditionnels pour deux raisons principales : d'une part, l'emploi d'outils multimédia permet des possibilités infinies d'intégration des différents supports (textes, images, photos, audio, vidéo ou 3D, etc.) ; d'autre part, l'interactivité développe la communication homme-machine, donnant ainsi à l'apprenant une certaine autonomie dans le choix de son parcours d'apprentissage (Kraif, 2003).

Plus récemment, l'apport des nouvelles technologies a permis à la linguistique de corpus de se développer, tout d'abord dans le monde anglo-saxon, puis en France. Dès ses débuts, la linguistique de corpus a été considérée non seulement comme un excellent instrument pour mener des observations sur les phénomènes linguistiques¹, mais aussi comme une méthodologie pour favoriser une approche opérationnelle dans le domaine de l'enseignement-apprentissage. Dans la littérature on a appelé ceci l'approche des « trois I » (illustration – interaction – induction) (McEnery & Xiao, 2010). Par

¹ Une importance capitale a été accordée aux variables dont dépend le contexte d'énonciation, à savoir ce que Firth (1957 : 28) appelle le « *context of situation* », notamment la relation entre les occurrences linguistiques et le monde extérieur (Robins, 1971 : 33-46). En outre, les réflexions dans ce domaine ont amené d'une part à affiner les analyses linguistiques (des logiciels de concordances facilitent considérablement le travail du linguiste-chercheur), d'autre part à rendre plus rapide la collecte des données linguistiques permettant ainsi de créer des ressources très vastes (certains corpus atteignent les centaines de millions de mots).

illustration on entend le fait de présenter à la classe d'étudiants en langues étrangères, des données réelles ; *l'interaction* concerne la deuxième phase du travail des étudiants qui échangent leurs opinions et leurs observations ; enfin *l'induction* permet à l'étudiant de se créer sa propre « règle » (qui sera affinée au fur et à mesure que le corpus de données s'élargit) sur le phénomène étudié.

Dans notre étude nous nous interrogerons tout d'abord sur les possibilités et les limites de la méthodologie de l'apprentissage sur corpus (désormais ASC – cf. Boulton & Tyne, 2014), en mettant en évidence les points de repère théoriques que nous fournit cette nouvelle approche ; puis nous porterons notre attention sur des questions d'ordre méthodologique qui pourront aider le didacticien à utiliser les outils mis à disposition par la linguistique de corpus pour favoriser l'apprentissage des langues étrangères (et en particulier du FLE). Dans une troisième partie nous présenterons des exemples concrets d'application de l'ASC à la didactique des langues, notamment pour ce qui concerne l'étude de patrons syntaxiques, du lexique de spécialité et des relations sémantiques (la synonymie).

2. Points de repère théoriques en linguistique de corpus

Johns (1991) proposait d'utiliser la linguistique de corpus, en particulier l'étude des concordances, dans une salle de classe, comme laboratoire pratique pour l'apprentissage des langues. Il parlait alors d'un processus qui, pour reprendre ses mots, « *needs to be driven by access to linguistic data*² » et créait l'expression « *data-driven learning* » (DDL) pour décrire cette approche.

Bernardini (2004 : 17) a souligné comment « *lo studio della lingua può essere visto come un processo induttivo nel quale il significato e la forma devono essere associati*³ », et cela selon des schémas que l'étudiant se construit lui-même en associant ses connaissances du monde à la langue/culture qu'il est en train d'étudier. A ces schémas pragmatiques correspondent des formes d'expressions établies, auxquelles l'étudiant est exposé grâce aux corpus et à l'analyse des concordances qui deviennent ainsi des instruments idéaux pour observer et apprendre les associations entre formes et contenus qui sont adoptées par une communauté linguistique donnée.

Avec cette approche l'étudiant interroge ses propres intuitions, pose ses propres hypothèses et les vérifie par l'analyse de données réelles. Les avantages d'une telle approche sont nombreux sur le plan de l'intérêt suscité chez l'étudiant qui est amené à faire des observations à partir de ses intuitions ; il devient alors acteur dans la construction de son propre savoir. Certes, le travail de l'étudiant ne pourra se faire qu'en semi-autonomie : il sera guidé par son professeur dans la découverte de la langue étrangère et il « apprendra à apprendre » (Bernardini, 2004 : 16) grâce à des exercices

² « [le processus d'apprentissage] a besoin d'être appuyé par l'accès à des données linguistiques » (notre traduction).

³ « L'étude de la langue peut être considérée comme un procédé inductif grâce auquel l'étudiant apprendra à établir une corrélation entre le signifié et le signifiant » (notre traduction).

qui développent ses capacités d'observation et d'interprétation des *patterns* ou patrons d'emploi des mots.

Mais quels sont les principes épistémologiques sur lesquels se basent cette discipline et cette approche didactique qui requièrent, en quelque sorte, une attitude de « chercheur » de la part des étudiants ?

2.1. La linguistique de corpus et l'école contextualiste britannique

Le courant britannique de la linguistique de corpus trouve ses fondements théoriques dans l'école contextualiste dont le précurseur est Firth. Son point de départ est le fait qu'aucun énoncé ne peut être compris en dehors du contexte culturel et situationnel dans lequel il est produit, sans tenir compte des informations d'ordre sociologique (âge, sexe, milieu) ainsi que des informations qui concernent l'acte énonciatif (genre de discours, endroit où il a été produit, buts de l'énonciation, etc.). D'où l'exigence de fonder ses analyses linguistiques sur les observations de « ce qui est visible » (Williams, 2003 : 35).

Les disciples de Firth ont contribué à diffuser ses idées innovantes, en particulier Sinclair qui a été le premier véritable maître à penser de la linguistique de corpus et qui a permis que cette discipline prenne réellement son essor. Son approche de la langue consiste à l'observer, à faire des hypothèses sur son fonctionnement et à prêter attention à « *what is going on*⁴ » (Hanks, 2008 : 220). Une autre proposition innovante de Sinclair consiste à considérer que les analyses linguistiques ne doivent pas se baser sur des textes particuliers, mais sur de grands ensembles de données dans la mesure où c'est seulement dans de très grands corpus que le chercheur peut trouver des régularités qui permettent de décrire le fonctionnement de la langue.

Certains spécialistes du domaine considèrent la linguistique de corpus comme une méthodologie applicable à différentes disciplines (McEnery & Wilson 2001) ; d'autres, au contraire, pensent que la linguistique de corpus est une discipline à part entière dans la mesure où elle a des présupposés épistémologiques évidents (Sinclair, 1991 ; Leech, 1992 ; Tognini-Bonelli, 2001). Cette dernière perspective d'analyse consiste, en effet, en une étude attentive et objective se basant sur des données linguistiques auxquelles on accède via un concordancier ; c'est donc une étude qui permet de généraliser les résultats obtenus à la totalité de la langue, à condition, bien sûr, que le type de corpus⁵ analysé le permette. Ces deux attitudes face à la linguistique de corpus ne sont pas totalement contradictoires et peuvent être tout à fait complémentaires (Williams, 2005 : 13).

Au-delà des réflexions théoriques sur la manière d'observer la langue, il était nécessaire de mettre au point les critères à adopter pour analyser les données

⁴ « ce qui se passe » (notre traduction).

⁵ Dans Habert et al. (1997) nous retrouvons une classification et une définition des typologies des corpus ainsi que les opérations à effectuer dans leur constitution et leur étiquetage.

linguistiques. Dans les paragraphes suivants nous analyserons de plus près les concepts proposés par Sinclair pour aider le chercheur et l'apprenant-chercheur à mettre en évidence les faits de langue saillants étudiés à partir des corpus.

2.2. Les critères d'analyse

L'analyse des *KWIC* (*Key Word In Context* – mot-clé en contexte) montre clairement qu'il existe des régularités observables à partir de quatre points de vue différents : un mot-clé peut être décrit selon ses collocations, ses colligations, ses préférences sémantiques et ses prosodies sémantiques.

Le concept de collocation est issu de la théorie du contextualisme firthien, selon laquelle la signification et l'emploi d'un mot peuvent être étudiés à partir de l'observation de son contexte linguistique, le co-texte, qui met en évidence les liens syntaxiques et sémantiques caractérisant le mot-clé et ses co-occurents. Les collocations sont analysées sur l'axe paradigmatique (l'axe vertical) qui met en évidence les patrons associés le plus fréquemment au mot-clé⁶.

Associé au concept de collocation, nous trouvons celui de colligation qui concerne les relations grammaticales et syntaxiques qui s'instaurent entre le mot-clé et son co-texte. Le chercheur peut retrouver ce genre d'information sur l'axe syntagmatique, sur chaque ligne de concordance.

Quant à la notion de préférence sémantique, elle relève des « sélections » que le mot-clé opère d'un point de vue sémantique. En effet, de nombreuses recherches ont mis en évidence que le mot-clé peut avoir comme co-occurents des mots qui partagent certaines caractéristiques sémantiques (Stubbs, 2002 : 449), des co-occurents qui ont donc des propriétés sémantiques communes.

Enfin, dans une étude sur *happen* en anglais, Sinclair⁷ avait remarqué que ce verbe avait dans la plupart des cas des collocatifs à valeur négative. Reprenant cette constatation, Louw (1993) a avancé l'hypothèse suivante : tout comme la présence de certains phonèmes dans un mot peut exercer une influence sur les phonèmes qui l'entourent, la connotation particulière d'un mot exerce une influence sur le sémantisme des mots qui constituent son cotexte, cette connotation positive ou négative étant la prosodie sémantique. Louw définit la prosodie sémantique comme une « forme de signification » (Louw, 2000), classifiable comme étant positive ou négative, qui révèle le point de vue du locuteur dans des situations pragmatiques

⁶ Dans le domaine de la linguistique francophone, quand on parle de « collocations » on entend, dans la plupart des cas, les caractéristiques qui ont été mises en évidence au niveau du figement ou du semi-figement des expressions de la langue et qui sont définies par Mel'čuk et ses associés de l'Université de Montréal comme des « unités préfabriquées à mi-chemin entre locutions et combinaisons libres » (Tutin, 2005). Il s'agit d'associations lexicales préconstruites composées d'une base qui conserve son sens premier et d'un collocatif qui a le rôle de donner une signification différente à l'expression. Les associations lexicales telles que *café noir*, *peur bleue*, *avoir une portée*, *satisfaire un besoin* sont considérées comme des collocations.

⁷ L'analyse est disponible sur le site <http://www.twc.it/happen.html>.

spécifiques. Il a de même démontré que les prosodies sémantiques sont liées aussi à des phénomènes d'ordre syntaxique (la transitivité, la réflexivité) et pragmatique (l'expression de l'ironie et du manque de sincérité).

Ces critères, ou étapes, grâce auxquels on arrive à identifier des unités de signification plus larges (Sinclair, 1991 : 110), peuvent être utilisés dans un parcours d'apprentissage des langues étrangères, dans la mesure où ils mettent en évidence des phénomènes relevant du système linguistique tels que des patrons syntaxiques, des co-occurents lexicaux ou des préférences sémantiques qui aident le locuteur à désambiguïser les synonymes.

3. Questions d'ordre méthodologique

Il est important d'aborder maintenant quelques questions d'ordre méthodologique concernant l'ASC appliquée au FLE. Il existe des concepts avec lesquels les acteurs de l'enseignement/apprentissage, enseignants et étudiants, doivent se familiariser, notamment les notions-clés de l'analyse sur corpus et le traitement des données brutes issues d'un concordancier.

3.1. Les notions-clés de l'analyse sur corpus

Pour que les étudiants puissent tirer profit d'un enseignement basé sur corpus, il semble nécessaire de les familiariser non seulement avec les termes qui indiquent les critères d'analyse, mais aussi avec les notions qui leur permettront d'interpréter les concordances (cf. Kübler, 2014).

Dans un premier temps, il convient d'aborder le concept de concordance. Le concordancier nous permet d'effectuer une recherche par mots-clés qu'il présente sur des lignes que l'on appelle une concordance. Une telle présentation des données nous offre la possibilité d'observer le corpus avec deux perspectives : l'une verticale correspondant à l'axe paradigmatique, l'autre horizontale correspondant à l'axe syntagmatique. L'axe vertical (Tognini-Bonelli, 2001 : 3) représente la totalité des occurrences du mot-clé dans le corpus, à savoir l'ensemble des emplois possibles de ce mot dans la langue, sa dimension paradigmatique. Par conséquent le corpus, s'il est considéré sur l'axe vertical, devient l'expression de la *langue*. L'axe horizontal, en revanche, représente l'emploi du mot-clé dans un contexte linguistique précis et particulier, nous sommes donc au niveau syntagmatique : une ligne de concordance correspond en effet à l'instance de *parole*⁸.

Sur le plan de la didactique du FLE, il est important que les apprenants (qu'ils soient débutants ou avancés) apprennent à observer les deux axes dans la mesure où

⁸ Ces deux types de lecture des concordances favorisent un changement de perspective dans la relation entre *langue* et *parole*, considérées autrefois comme deux entités dans une relation dichotomique : elles sont désormais dans une relation dialogique, car le corpus présente les deux entités simultanément.

c'est à partir de l'étude de ces deux dimensions qu'ils pourront découvrir les régularités qui leur permettront de systématiser et d'intérioriser les règles d'usage de la langue. Mais il est bien évident que le type et le niveau de difficulté des activités varieront en fonction de la compétence des apprenants.

Une autre notion à analyser et à interpréter est la fréquence des mots ou expressions recherchés. En effet c'est la fréquence, à savoir le nombre d'occurrences d'un mot-clé, qui nous indique si tel ou tel mot est peu ou beaucoup utilisé.

3.2. L'extraction des concordances et le travail d'épuration

Est-il nécessaire de filtrer le corpus en vue de l'enseignement d'une langue étrangère ? Dans les phases initiales de la « recherche » de l'apprenant en FLE, une enquête menée sur des données importantes, non filtrées, pose problème. Un nombre trop élevé d'occurrences risque, entre autres, de décourager l'apprenant qui pourrait être amené à analyser des centaines de contextes ; c'est plutôt à un étudiant en linguistique que l'on pourrait demander de faire ce type d'exercice.

Il convient donc de limiter voire de filtrer (ou tout simplement de présélectionner ou didactiser) les données auxquelles ont accès les apprenants. Il s'agit souvent de fournir à l'étudiant en FLE des lignes de concordances dans lesquelles il peut reconnaître les phénomènes d'ordre syntaxique et/ou lexical qui font l'objet de son apprentissage (cf. Kerr, 2013). À travers ses observations, conduites sur la base de questions *ad hoc* préparées par l'enseignant, il arrivera à expliciter la règle d'emploi de ces mots (pour les exemples voir ci-dessous). Le traitement que l'enseignant doit opérer sur les lignes de concordances avant de les utiliser avec ses étudiants consiste donc à éviter des occurrences qui pourraient poser problème.

À titre d'exemple, nous citons un travail de préparation dans lequel nous avons éliminé toutes les occurrences qui risquaient de compliquer l'interprétation des concordances, voire d'induire en erreur nos étudiants de FLE⁹. Cette opération de « nettoyage » a été faite sur la liste des occurrences obtenues pour le mot-clé *beaucoup (de)*¹⁰, avec le concordancier en ligne du site de Tom Cobb, Lextutor (<http://www.lexutor.ca/conc/fr>). En observant cette liste (Figure 1) nous remarquons que :

- dans les constructions des lignes de 2 à 8, *beaucoup* apparaît toujours suivi de la préposition *à* ; l'apprenant pourrait être amené à penser que cette colligation est propre au mot *beaucoup*, ne comprenant pas que la préposition fait partie de la locution verbale (« *laisser à désirer* » – « *devoir à* + quelqu'un ou quelque chose » – « *rester à faire* » – « *ressembler à* ») et que *beaucoup* se rapporte au verbe ;

⁹ Il s'agit d'une activité préparée pour des étudiants italophones en première année universitaire dans un module de *Lingua e traduzione francese* (Langue et traduction française).

¹⁰ L'objectif de l'activité, qui est présentée dans la section 4.1.1., était de retrouver les règles d'emploi de *très* et *beaucoup (de)* qui correspondent en italien au seul mot *molto*.

- dans la ligne 17 (*beaucoup confiance*), le mot *confiance* fait partie de la locution « *faire confiance* ». Nous avons supprimé cette occurrence du fait de l'utilisation particulière de l'intensifieur dans l'expression « faire confiance ».

1. ... aventure originale les tente	beaucoup: 200 chercheurs nous ont déjà ...
2. ... cessus décisionnel qui laisse	beaucoup à désirer. Ces manquements, f ...
3. ... cérolés". Nous devons surtout	beaucoup à l'entreprise. "Performant" ...
4. ... ssait que de 3% par an - doit	beaucoup à l'arrivée du cédérom il y a ...
5. ... amélioration mais il en reste	beaucoup à faire. Une analyse sérieuse ...
6. ... in survenu le 29 janvier doit	beaucoup à Berlin et Téhéran, influent ...
7. ... ssait que de 3% par an - doit	beaucoup à l'arrivée du cédérom il y a ...
8. ... té et de l'autre, ressemblent	beaucoup à un mouvement de Tai Chi. Ce ...
9. ... L'occasion. Huguette dépense	beaucoup, aime manger des plats import ...
10. ... pliquait : « Moi, je n'ai pas	beaucoup aimé car j'ai un faible pour ...
11. ... 'est passé? De mon côté, j'ai	beaucoup appris. Je fus témoin d'une j ...
12. ... et les chiites libanais doit	beaucoup au rôle joué par l'Allemagne ...
13. ... C'est quelque chose qui plaît	beaucoup aux politiques aussi, mais ce ...
14. ... conjoncturel de 1997 n'a pas	beaucoup changé la donne: le salaire m ...
15. ... monde : ceux qui en attendent	beaucoup comme ceux qui le redoutent l ...
16. ... rd-américaine, a probablement	beaucoup compté. - Après New York et ...
17. ... vous constaterez qu'ils font	beaucoup confiance au nom du gouvernem ...
18. ... malencontreusement, polluent	beaucoup. Contrairement aux algues, ...
19. ... Saint-Laurent, n'ont pas reçu	beaucoup d'attention et les problèmes ...
20. ... l est très important de boire	beaucoup d'eau durant la cure. CONT ...
21. ... es. Dans les récentes années,	beaucoup d'entre nous avons perdu notr ...
22. ... malheureusement, c'est ce que	beaucoup d'entre eux ne comprennent pa ...
23. ... e fumons pas, si nous faisons	beaucoup d'exercice et si nous diminu ...
24. ... de 100 espèces quoique, pour	beaucoup d'observateurs, c'est un chif ...
25. ... de l'été et durant l'automne,	beaucoup d'oiseaux de rivage migrent p ...
26. ... sation. Ce processus consomme	beaucoup d'oxygène et, dans ces condit ...
27. ... le congrès de Dijon. Il a dit	beaucoup d'amabilités à mon propos - l ...
28. ... j'aime beaucoup Paris, j'y ai	beaucoup d'amis et je m'y sens bien. U ...
29. ... e façon brutale. Par exemple,	beaucoup d'animaux sont en voie d'exti ...

Figure 1 : Liste des concordances brutes de beaucoup

Nous avons donc éliminé certaines occurrences dans la mesure où les apprenants de niveau débutant ne sont en général pas capables de faire une telle réflexion métalinguistique qui suppose une connaissance plutôt avancée ou en tout cas approfondie de la langue¹¹.

Le tri des lignes de concordance se fait donc en fonction de l'objectif d'apprentissage et du niveau des étudiants.

4. Applications

Après avoir abordé les questions de méthode, nous donnerons une série d'exemples d'activités basées sur l'ASC qui ont été proposées à un public italien. Il faut préciser tout d'abord que la réflexion sur les collocations, très utile pour un apprenant en langues étrangères pour comprendre combien les emplois sont liés par des contraintes, ne passe pas obligatoirement par l'utilisation des corpus et concordances.

C'est le cas, par exemple, de la « *Lexical Approach* » introduite par Michael Lewis (1993, 1997), qui propose une approche de l'enseignement du lexique dans laquelle les collocations jouent un rôle central. L'enseignement des collocations consiste non

¹¹ En fait, un travail de réflexion sur l'ensemble complet des occurrences aurait pu être fait avec des apprenants de niveau intermédiaire, afin de développer leur capacité de réfléchir aux structures que la langue nous propose.

seulement à fournir aux étudiants un échantillon de possibilités lexicales pour accroître leur vocabulaire, mais surtout à leur faire découvrir les associations qui existent entre les mots. Morgan Lewis (2000 : 20) propose un exemple d'activité à développer en classe, non pas pour corriger mais pour collecter (« *don't correct – collect* »). C'est le cas de la présentation des collocations du mot *exam*, que le professeur donne ou fait découvrir à l'apprenant sans recourir au support d'un corpus. A la fin de cette activité, l'apprenant pourra utiliser le concept *exam* car il aura à sa disposition l'ensemble des mots et des associations nécessaires pour pouvoir en parler (Morgan Lewis, 2000 : 21).

Voyons, pour appliquer cet exemple au français, les collocations du substantif *examen* qui posent problème pour un apprenant italien. L'enseignant peut écrire au tableau tous les verbes que les étudiants associent au mot-clé et contrôler que le sémantisme des différentes collocations soit assimilé. Nous aurons donc une liste comme celle-ci :

Avoir (à passer)	}	un examen
Avoir (son) = réussir (fam.)		
Passer		
Etre reçu à		
Rater (son)		
Echouer à		

À travers cette liste de collocations, écrite simplement au tableau, l'étudiant sera amené non seulement à connaître le sémantisme des diverses expressions, mais aussi les colligations des formes verbales associées au mot *examen* (le verbe approprié, la présence ou non du déterminant associé à un verbe donné, etc.).

Comme conclusion de l'activité, l'enseignant peut proposer un exercice à trous pour vérifier que les étudiants aient bien compris les différences sémantiques de ces expressions.

Tout en reconnaissant l'intérêt de l'approche lexicale de Michael Lewis pour l'enseignement systématique du lexique, nous nous concentrerons ici sur l'utilisation pédagogique des collocations à partir des corpus et des concordances, ce qui permet, entre autres, de fournir aux étudiants un corpus avec un minimum de contexte.

Nous avons pu constater que l'introduction du corpus en classe aide à développer le plaisir de la recherche et de la découverte en autonomie. C'est pourquoi nous préférons demander aux étudiants d'interroger les concordances à partir d'un corpus donné pour découvrir une règle de grammaire ou l'emploi d'un mot et de ses synonymes dans des contextes variés et cela en fonction de leur niveau et des objectifs de formation.

4.1. Apprendre les patrons syntaxiques

Nous trouvons très efficace de proposer aux apprenants de niveau A1-A2 des analyses sur corpus, pour leur faire apprendre des colligations (ou structures

syntaxiques) qui ne coïncident pas dans les deux langues. C'est le cas, par exemple, de *très* et *beaucoup (de)* et de *entre* et *parmi*.

4.1.1. Le cas de *très* et *beaucoup (de)*

Très et *beaucoup (de)* correspondent en italien à *molto*, mot variable avec une fonction adverbiale ou adjectivale. Si le passage du français à l'italien ne pose pas problème quant à l'emploi de ces formes, l'inverse s'avère plus problématique, comme en attestent les nombreuses erreurs que font encore les étudiants en troisième année de Licence.

Pour aider les étudiants à systématiser l'emploi de *beaucoup*, *beaucoup de* ou *très* dès le début de leur parcours de formation, nous leur avons demandé d'observer les concordances des mots *beaucoup* et *très* à l'aide d'une série de questions :

1. Quels sont les mots qui constituent le contexte gauche et droit de *très* et *beaucoup* ?
2. Quelle est la catégorie syntaxique de ces mots ?
3. Y a-t-il des constructions récurrentes ?
4. À quelle catégorie grammaticale correspond le co-texte à gauche de *beaucoup* ?
5. Les mots qui suivent la construction *beaucoup de* sont-ils au pluriel, au singulier... ?

Ensuite nous avons demandé aux étudiants de rassembler les collocatifs qui ont des éléments en commun, comme les adverbes en *-ment*, les formes verbales simples et conjuguées, etc. Une fois fini le repérage des collocatifs, nous leur avons demandé de classer tous les co-occurents dans deux tableaux, l'un pour *très*, l'autre pour *beaucoup (de)* (Figure 2) afin d'identifier les colligations et de découvrir les règles d'emploi.

Collocations	Colligations	Traduction en italien
beaucoup de	beaucoup de + substantif au singulier et au pluriel	molto molta molti molte
beaucoup (de personnes) travaillent beaucoup voient beaucoup racontent beaucoup ne parviennent pas	Beaucoup (de personnes) + verbe conjugué	molti lavorano molti vedono
tente beaucoup dépense beaucoup polluent beaucoup il en fait beaucoup les organisme impliqués peuvent beaucoup	Verbe (forme simple)+ beaucoup	spende molto inquinano molto
j'ai beaucoup appris je n'ai pas beaucoup aimé a probablement beaucoup compté ont beaucoup diminué	Auxiliaire + beaucoup + participe passé	ho imparato molto non mi è piaciuto molto sono diminuiti molto
semblent beaucoup nuire pourraient beaucoup profiter	verbe conjugué + beaucoup + verbe (infinitif)	sembrano nuocere molto

Figure 2 : Tableau des colligations et des collocations de *beaucoup (de)*

En se basant sur l'observation de la récurrence et de la position des mots les étudiants vont systématiser les occurrences et parvenir ainsi du point de vue cognitif à une généralisation et à une mise en évidence des exceptions à la règle générale.

4.1.2. Le cas de *entre* et *parmi*

Pour le couple *entre* et *parmi*, nous avons remarqué le même problème qu'avec *très* et *beaucoup (de)*, dans la mesure où ces deux termes ont un seul équivalent en italien, *fra* (ou encore *tra*). En effectuant la recherche dans Lextutor, nos étudiants de niveau B1-B2 ont analysé les collocatifs de ces deux mots dans des concordances (voir Figure 3 et Figure 4). En nous aidant de questions qui concernent les collocatifs de ces deux mots (« Quels sont les mots qui apparaissent à droite de *entre* ? et de *parmi* ? »), nous avons ainsi amené les étudiants à réfléchir sur les mots qui viennent à droite du mot-clé. Dans le cas de *entre* pour les étudiants habitués aux explications données par la plupart des grammaires il est évident que cette préposition entraîne la présence de deux éléments qui se complètent (par ex. *entre 10 et 50 kilomètres, entre 40 et 60 personnes, entre 5 % et 10 %*, etc.).

```
arison) Click keyword for more context
prévu l'émission commençant ENTRE 0 heure et 0 h 40 suivant les
l dans la stratosphère soit ENTRE 10 et 50 kilomètres d'altitude
s de six relevés différents ENTRE 15 heures et 17 heures, dans t
rouge aux côtés des Afghans ENTRE 1979 et 1989, ben Laden est un
Dans la minorité municipale ENTRE 1989 et 1995, il avait combatt
, vend 3 milliards d'actifs ENTRE 1993 et 1994 et montre des qua
cot met en avant le montant ENTRE 200 millions et 400 millions d
a lourde indemnité, évaluée ENTRE 200 millions et 400 millions d
ir entre 30 et 50 à Lyon et ENTRE 260 et 350 à Paris (Le Monde d
lés à la pollution de l'air ENTRE 30 et 50 à Lyon et entre 260 e
ter main forte aux secours. ENTRE 40 et 60 personnes auraient ét
che, ceux qui ont recueilli ENTRE 5 % et 10 % des suffrages ont
ats ceux qui ont recueilli ENTRE 5 % et 10 % des voix ont vu la
```

Figure 3 : Liste brute des occurrences du mot *entre*

Ensuite nous avons donné aux étudiants une autre liste de concordances avec des éléments pluriels à droite du mot *entre* (*une histoire d'amour entre deux enfants, des signes de désaccord entre généraux refont surface, des consultations intensives ont lieu entre les alliés, la solidarité entre les pays, la loi aboutit à une plus grande justice entre les candidats, un pacte d'actionnaire a été signé entre les associés*), leur permettant d'observer d'autres emplois. Pour chercher à classifier ces emplois différents, nous avons posé d'autres questions (« Comment définiriez-vous la signification de ces termes ? Quel concept expriment-ils ? ») concernant le sémantisme des termes qui se trouvent à gauche de *entre*, à savoir sur *histoire d'amour, désaccord, consultations, solidarité, justice, pacte*. Ce qui ressort des réponses des étudiants c'est qu'il s'agit dans tous les cas d'une « relation entre éléments ». Il semble donc que la présence de ce sème relie les

deux types de collocatifs de *entre* (coordination entre deux éléments et relation entre éléments)¹².

En revanche, associés à *parmi* (Figure 4), nous trouvons des éléments qui représentent un ensemble de personnes, de choses (*parmi les plus anciens de la Chine, parmi les plus riches, parmi les médicaments, parmi les meilleurs spécialistes, etc.*), souvent associés à des superlatifs. Ce terme est « toujours suivi de l'article défini ou adjectif démonstratif qui actualisent l'idée d'ensemble défini dans lequel on choisit, on isole, on extrait un élément » (Hédiard, 2003 : 116-117).

is et menacés de l'auteur, il reste **PARMI** les livres d'Elsa Morante quelques trésors d
ction "Bouquins" de Robert Laffont. **PARMI** les livres sur elle et son entourage, signal
"rumeurs de corruption", y compris **PARMI** les magistrats (Le Monde du 30 juillet). L'e
ohibé n'a finalement été découvert **PARMI** les médicaments saisis dans un camion de l'é
r-né de la famille des deux-roues, **PARMI** les meilleurs spécialistes de la planète. Mê
e au titre des personnes désignées **PARMI** les membres du Conseil d'Etat, de la Cour de
ressortissants américains figurent **PARMI** les morts et six autres sont portés disparus
isième défi concerne l'inquiétude, **PARMI** les nations les plus riches, que la concurre
t leur indique les aires de fêtes. **PARMI** les nouveautés, la possibilité de s'initier
: visiter d'autres régions du pays. **PARMI** les numéros qui ont marqué l'histoire de l'é
non seulement dans la foule, mais **PARMI** les paras. Ceux-ci auraient ainsi pu riposte
s d'un village au coeur du Quercy. **PARMI** les plaisirs annoncés : la Fanfare d'Assier,
rque à bouche sheng, un instrument **PARMI** les plus anciens de la Chine (ses premières

Figure 4 : Liste brute des occurrences du mot *parmi*

Enfin, pour évaluer le degré d'acquisition de la règle que les étudiants viennent de découvrir, nous leur avons soumis un exercice à trous réalisé par le même logiciel, présentant des phrases complètes extraites du même corpus.

4.2. Apprendre les patrons lexicaux

Notre expérience nous montre que l'étude des concordances peut aider les étudiants à améliorer leur compétence en expression écrite : ils apprennent non seulement à enrichir leur lexique mais surtout à tenir compte des associations, des contraintes d'emploi et du critère de fréquence entre des mots ou expressions presque équivalents en français, et à éviter des calques à partir de l'italien, écueil fréquent vu la proximité entre les deux langues. À partir de la correction des devoirs écrits que nos étudiants de niveau B1-B2 doivent réaliser sur des sujets d'actualité, tels que l'écologie, nous avons constaté des erreurs dans l'emploi des collocatifs associés à certains termes tels que l'adjectif d'intensité employé avec le mot *actualité*, ou encore une tendance à employer systématiquement des expressions proches de l'italien (*en ce qui concerne* vs *pour ce qui concerne*). L'enseignant doit aider l'étudiant à observer des collocations appropriées (en considérant aussi la notion de fréquence) et à utiliser, de nouveau, par

¹² Voir aussi Hédiard (2003).

exemple, le concordancier en ligne dans Lextutor. Le but de cette approche est de faire en sorte que l'étudiant se corrige lui-même à partir des concordances ou de modèles d'écrits produits par des scripteurs natifs.

4.2.1. Traduction de *estrema attualità*

L'analyse des concordances permet de sensibiliser les apprenants au problème des associations entre mots, en l'occurrence des combinaisons entre substantif et adjectif.

Il s'agit d'un exercice proposé à un étudiant de troisième année de Licence (niveau B2 du CECRL) comme recherche d'une alternative à l'expression qu'il a employée. L'étudiant s'est laissé influencer par l'italien et a traduit littéralement *estrema attualità* par *extrême actualité*. Or, nous lui avons proposé une mini-recherche sur le concordancier en ligne, grâce à laquelle il a constaté que l'expression qu'il a choisie est peu présente dans un corpus d'un million de mots et que le seul adjectif qui est en cooccurrence avec le substantif *actualité* est *brûlante* en position postposée¹³. Les résultats de sa recherche sur corpus ont été confirmés avec une requête sur un moteur de recherche qui lui a fourni un nombre élevé d'occurrences (1 840 000 occurrences) de la collocation avec l'adjectif postposé à côté d'un emploi plus restreint (78 800 occurrences) de « brûlante actualité ».

4.2.2. Pour ce qui concerne vs en ce qui concerne

Une autre activité de « recherche » a été proposée à des étudiants de niveau B2 liée à l'étude de la fréquence des locutions *pour ce qui concerne* vs *en ce qui concerne*, en raison de l'influence que l'expression italienne *per quanto riguarda* exerce sur le choix de l'une des deux expressions françaises.

La première tournure n'est pas fautive mais on peut constater à la lumière des recherches effectuées sur le concordancier en ligne et sur les moteurs de recherche qu'elle est beaucoup moins fréquente que la deuxième. En effet, on trouve 25 occurrences de l'expression *en ce qui concerne*, dans le corpus de Lextutor, par rapport aux 4 occurrences de l'expression *pour ce qui concerne*, plus proche de l'italien *per quanto riguarda*.

À travers l'étude des fréquences l'étudiant peut donc se rendre compte de l'utilisation effective des deux locutions et choisir celle qui est la plus employée dans la langue cible. Même si la seule fréquence paraît problématique (par exemple, pour l'étude de certains « faux » synonymes où les aspects qualitatifs sont importants), elle permet néanmoins à l'apprenant de choisir entre deux voire plusieurs formes potentiellement en concurrence.

¹³ Nous avons repéré aussi deux occurrences avec l'adjectif antéposé. Dans ce cas le figement de cette expression est plus marqué, même si l'expression est plus rare.

4.2.3. Apprendre le lexique de spécialité.

L'apprentissage du lexique de certains domaines disciplinaires est une phase incontournable dans la formation en langue de spécialité. L'analyse des lignes de concordances de mots-clés relevant d'un domaine spécifique facilite l'apprentissage du lexique et en particulier l'acquisition des collocations.

Citons, par exemple, le cas des mots *développement durable*, *pollution*, *dégâts* qui sont très liés au domaine de l'environnement, mots que nous avons traités pendant nos cours de français dans le cadre de la *laurea specialistica* (Master 1) en Sciences de la communication publique et d'entreprise¹⁴. Le cours était centré sur l'apprentissage du français de spécialité, c'est pour cela que nous avons abordé les sujets sur la base des matières fondamentales de la filière en nous appuyant sur des textes accessibles en ligne. Parmi les différents sujets abordés, il y en avait un qui concernait les problèmes de l'environnement et des énergies renouvelables.

La leçon comportait trois parties : tout d'abord, à partir du texte numérisé nous avons effectué l'extraction des mots les plus fréquents qui ont permis aux étudiants d'avancer des hypothèses sur le thème abordé dans le document. Ensuite nous avons étudié le texte pour parvenir à une compréhension globale, puis détaillée à l'aide de questions ouvertes ou à choix multiple. Nous sommes passés enfin à l'étude des concordances pour faire émerger, par induction, les collocations des mots les plus représentatifs du domaine et pour amener les étudiants à les réemployer dans des tâches de traduction ou de production écrite (essai argumentatif sur l'environnement, par exemple).

Dans cette dernière phase, nous nous sommes concentrés, par exemple, sur les collocations du mot *développement*. Dans le domaine de l'environnement, il existe deux adjectifs qui peuvent être associés à ce terme, *durable* et *soutenable*. Cette dernière variante est un calque de l'anglais *sustainable*, que nous retrouvons aussi en italien (*sostenibile*). Or, si au niveau de la compréhension, l'emploi de *soutenable* ne pose pas problème, au niveau de la fréquence d'emploi, il est beaucoup moins fréquent que *durable* (509 000 résultats pour *soutenable* et 44 900 000 pour *durable* dans Google). À travers les données sur la fréquence et les listes des concordances, les étudiants ont ainsi pu apprendre le collocatif le plus utilisé du mot *développement*¹⁵.

4.3. Apprendre à désambigüiser les synonymes

Dans notre cours de sémantique lexicale pour des étudiants en Licence de langues étrangères (niveau B2), nous avons abordé l'analyse de la relation de synonymie à travers l'étude des concordances en utilisant non seulement les concepts

¹⁴ L'examen consistait en une traduction de l'italien en français.

¹⁵ Cet exemple nous confirme l'importance de l'étude des fréquences : elles montrent que, même si une expression dans la langue de l'apprenant a son équivalent dans la langue étudiée, il ne va pas de soi que cette dernière soit la plus utilisée. Il est donc intéressant de comparer les fréquences des mots d'un point de vue inter-linguistique (Hédiard, 2007 : 286-287).

de collocation et de colligation, mais aussi la notion de préférence sémantique. En l'occurrence, nous avons proposé une réflexion sur certains verbes qui expriment la cause et nous avons montré comment ces concepts fournissent des critères efficaces pour désambiguïser certains mots qui sont traités, dans des dictionnaires de langue et de synonymes, comme des synonymes absolus mais qui en réalité ne sont que des parasyonymes.

Pour cette étude, nous avons utilisé un corpus d'environ 1 million de mots que nous avons constitué dans le cadre de notre recherche doctorale à partir d'articles tirés de la presse en ligne¹⁶. Les verbes que nous avons proposés aux étudiants sont *causer*, *engendrer*, *entraîner*, *générer*, *susciter* et *provoquer*.

À travers l'analyse des concordances, les étudiants, à l'aide d'un questionnaire, ont mis en évidence les différentes collocations associées à chacun de ces verbes et ont rempli une grille (Tableau 5) en fonction de la fréquence, des collocations et des préférences sémantiques, critères qui se sont avérés les plus pertinents pour désambiguïser ces synonymes.

Verbe	fréquences	collocations ¹⁷	préférences sémantiques
<i>causer</i>	68 occurrences	<i>gaz, problèmes, pollution, malformations, dommages, dégâts</i>	domaine de l'environnement
		<i>cancer, problèmes, lésions</i>	domaine de la médecine
		<i>émotions, préjugés, irritation, tollé, malaise</i>	domaine des sentiments
<i>engendrer</i>	40 occurrences	<i>inquiétudes, inégalité, refus, stress, troubles, souffrances, désespoir, malaise, frustrations</i>	domaine des sentiments
		<i>mouvements de population, problèmes, exclusion sociale, courants impétueux, initiatives et actions, résistance</i>	domaine social
		<i>coûts, livraisons, euros, chiffre d'affaire</i>	domaine économique
		<i>pollution, dégâts</i>	domaine de l'environnement
<i>entraîner</i>	86 occurrences	<i>dévalorisation, baisse, surconsommation, frais, restrictions budgétaires, chute des profits</i>	domaine économique
		<i>suppression, retrait, chute du mur, évacuation, ajustement structurel, démission, changement radical, dessaisissement, mort</i>	domaine social
<i>générer</i>	36 occurrences	<i>résidus, pollution, déchets, effet de serre, pollution, environnement favorable</i>	Domaine de l'environnement
		<i>profits, revenus, emplois, ventes, gains, marché, commerce, dollars</i>	domaine économique

¹⁶ *Le Monde*, *Le Monde Diplomatique*, *Le Figaro*, *L'Express*, *Le Parisien*, *Les Échos*, *Cybersciences*, *Éléments*, *Le Quotidien du Médecin*. Tous les articles utilisés ont été publiés entre 2004 et 2006.

¹⁷ Certains mots à signification générale (*gaz, problèmes, préjugés*) ont été insérés dans un domaine sémantique particulier du fait du contexte sémantique dans lequel ils figuraient.

<i>provoquer</i>	170 occurrences	<i>tensions, panique, inquiétudes, agacement, tollé, sécurité, préoccupations, rejet, désempolement, malaise, inégalités, amertume, confiance, mécontentement</i>	Domaine des sentiments
		<i>aggravation, problèmes de santé, pandémie, typhoïde, choléra, troubles rénaux, lésions, allergies, cancer, maladie chronique</i>	domaine de la médecine
		<i>crise, émeutes, changement de régime, catastrophe sociale, dévastations, changement de comportement, fins de l'empire, départ des familles, déplacement, désertions, révolte, fuite</i>	domaine social
		<i>incendie, perturbations, inondations, variations de température, émission de méthane, dommages, changement climatiques, marée noire, noyade</i>	domaine de l'environnement
<i>susciter</i>	104 occurrences	<i>craintes, stupéfaction, aigreur, enthousiasme, déception, incertitudes, panique, émotion, indignation, espoir, sympathie</i>	domaine des sentiments
		<i>réactions, critiques, interrogations, exposition, opposition, mobilisation, intervention</i>	actions de désapprobation, discussion sur un sujet

Tableau 5 : Désambiguïsation des synonymes

Cette activité nous a permis de montrer aux étudiants que chacun des verbes sélectionne des termes appartenant à des domaines spécifiques et que l'analyse sur corpus peut nous aider, en présence de plusieurs synonymes, à choisir le mot le plus approprié¹⁸.

5. Conclusion

Dans cet article nous avons exposé les raisons pour lesquelles nous avons adopté l'ASC dans nos cours de FLE et nous avons montré comment nous avons adapté la démarche à des publics et à des contextes différents. Au cours des différentes expériences, nous avons pu constater que l'étudiant est capable de devenir « chercheur et artisan de son apprentissage » avec cette approche et qu'il est particulièrement motivé. Il développe une autonomie qu'il transfère aussi à d'autres secteurs de l'apprentissage. Il faut ajouter que l'apprenant tire profit de cette façon d'apprendre parce qu'il conceptualise et systématise, à travers des activités de réemploi, les règles de fonctionnement qu'il a découvertes et les mémorise plus aisément. En dernier lieu, l'habitude d'observer les faits de langue l'aide à développer une meilleure capacité d'induction ainsi qu'une conscience linguistique et métalinguistique plus affûtée.

¹⁸ Les activités de recherche proposées aux apprenants ont été conçues à partir des erreurs relevées dans leurs devoirs (surtout des essais et des traductions). Elles se sont basées surtout sur les analyses des collocations des synonymes (par exemple *atteindre* et *rejoindre*, en italien *raggiungere*, ou bien *an* et *année*, en italien *anno*).

Et comme c'est généralement le cas pour toute nouvelle approche, il existe néanmoins des problèmes, des limites. En effet, si les activités ne sont pas suffisamment guidées, elles risquent de susciter un sentiment de désorientation chez l'apprenant, surtout s'il n'est pas très avancé dans son apprentissage (Gavioli, 2001 : 108) et si le mot ou phénomène étudié présente un fonctionnement complexe et difficile à interpréter. Il faut ajouter, en outre, que l'enseignant doit investir beaucoup de temps dans la préparation des cours basés sur l'ASC (du moins dans un premier temps), étant donné qu'il doit « préparer » non seulement les données que les étudiants analyseront (corpus, extraction des concordances, listes des fréquences, etc.), mais aussi le matériel de support pour ces analyses (tableaux, questions pour guider la recherche, etc.).

Il est évident que l'ASC ne constitue pas une nouvelle panacée (cf. Boulton & Tyne, 2014) : elle ne peut pas couvrir tous les aspects de l'apprentissage, ni remplacer d'autres pratiques d'enseignement (liées, par exemple, au développement de l'expression). Il faut éviter que les moments de réflexion sur la langue occupent trop de place dans la classe au détriment de toutes les autres activités visant à développer différentes aptitudes langagières. Il revient à l'enseignant d'évaluer à quel moment et avec quelle méthodologie il peut adopter cette démarche, stimulante et riche.

Bibliographie

- BERNARDINI, S. (2004), « Corpora in the classroom », dans J. Sinclair (éd.), *How to use Corpora in Language Teaching*, Amsterdam, John Benjamins, p. 15-36.
- BOULTON, A. (2007), « Esprit de corpus : promouvoir l'exploitation de corpus en apprentissage des langues », *Texte et corpus*, 3, p. 37-46, http://web.univ-ubs.fr/corpus/jlc5/ACTES/ACTES_JLC07_boulton.pdf.
- BOULTON, A. (2011), « Data-driven learning: the perpetual enigma », dans S. Roszkowski & B. Lewandowska-Tomaszczyk (éds.), *Explorations across Languages and Corpora*, Francfort, Peter Lang, p. 563-580.
- BOULTON, A. & TYNE, H. (2014), *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*, Paris, Didier.
- FIRTH, J. R. (1957), *Papers in Linguistics 1934-51*, Oxford, Oxford University Press.
- GAVIOLI, L. (2001), « The learner as a researcher: Introducing corpus concordancing in the classroom », dans G. Aston (éd.), *Learning with Corpora*, Houston, Athelstan, p. 108-137,
- GROSS, G. (1996), *Les expressions figées en français*, Paris, Ophrys.
- HABERT, B., NAZARENKO, A. & SALEM, A. (éds.) (1997), *Les linguistiques de corpus*, Paris, Armand Colin.
- HANKS, P. (2008), « The lexicographical legacy of John Sinclair », *International Journal of Lexicography*, 21/3, p. 219-229.

- HÉDIARD, M. (2007), « La notion de fréquence du point de vue de la didactique d'une langue voisine », dans M. Hédiard (éd.), *Linguistica dei corpora. Strumenti e applicazioni*, Cassino, Edizioni dell'Università degli studi di Cassino, p. 285-305.
- HÉDIARD, M. (2003), « Pour une grammaire 'motivée' à travers l'analyse des collocations », *Rassegna Italiana di Linguistica Applicata*, 35/1-2, p. 113-125.
- JOHNS, T. (1991), « Should you be persuaded: Two examples of data-driven learning », dans T. Johns & P. King (éds.), *Classroom Concordancing, English Language Research Journal*, 4, p. 1-16.
- KERR, B. (2013), « Grammatical description and classroom application: Theory and practice in data-driven learning », *Bulletin suisse de linguistique appliquée*, 97, p. 17-39.
- KRAIF, O. (2003), « Propositions pour l'intégration d'outils TAL aux dispositifs informatisés d'apprentissage des langues », *Lidil*, 28, p. 153-165.
- KÜBLER, N. (2014), « Mettre en œuvre la linguistique de corpus à l'université : vers une compétence utile pour l'enseignement/apprentissage des langues ? » *Recherches en didactique des langues et des cultures : Les Cahiers de l'Acedle*, 11/ 1, p. 37-77.
- LEECH, G. (1992), « Corpora and theories of linguistic performance », dans J. Svartvik (éd.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium*, Berlin/New York, Mouton de Gruyter, p. 105-122.
- LEWIS, Michael (1993), *The Lexical Approach*, Hove, Language Teaching Publications.
- LEWIS, Michael (1997), *Implementing the Lexical Approach: Putting Theory into Practice*, Hove, Language Teaching Publications.
- LEWIS, Michael, (éd.) (2000), *Teaching Collocation*, Hove, Language Teaching Publications.
- LEWIS, Morgan (2000), « There is nothing as practical as a good theory », dans M. Lewis (éd.), *Teaching Collocation*, Hove, Language Teaching Publications, p. 10-27.
- LOUW, B. (1993), « Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies », dans M. Baker, G. Francis & E. Tognini-Bonelli (éds.), *Text and Technology: In Honour of John Sinclair*, Amsterdam, John Benjamins, p. 157-76.
- LOUW, B. (2000), « Contextual prosodic theory: Bringing semantic prosodies to life », dans C. Heffer, H. Sauntson & G. Fox (éds.), *Words in Context: A Tribute to John Sinclair on his Retirement*, Birmingham, University of Birmingham, p. 48-94.
- MCENERY, T. & WILSON, A. (2001), *Corpus Linguistics*, Edimbourg, Edinburgh University Press.
- MCENERY, T. & XIAO, R. (2010), « What corpora can offer in language teaching and learning », dans E. Hinkel (éd.), *Handbook of Research in Second Language Teaching and Learning*, Londres, Routledge, p. 364-380.
- ROBINS, R. H. (1971), « Malinowski, Firth and the 'Context of Situation' », dans E. Ardener (éd.), *Social Anthropology and Language*, Londres, Tavistock, p. 33-46.
- SINCLAIR, J. (1991), *Corpus, Concordance and Collocation*, Oxford, Oxford University Press.
- STUBBS, M. (2002), « Two quantitative methods of studying phraseology in English », *International Journal of Corpus Linguistics*, 7/2, p. 215-244.
- TOGNINI-BONELLI, E. (2001), *Corpus Linguistics at Work*, Amsterdam, John Benjamins.

Sonia Di Vito

TUTIN, A. (2005), « Le dictionnaire des collocations est-il indispensable ? » *Revue Française de Linguistique Appliquée*, 10/2, p. 31-48.

WILLIAMS, G. (2003), « Les collocations et l'école contextualiste britannique », dans F. Grossman & A. Tutin (éds.) *Les collocations. Analyse et traitement*, Amsterdam, De Werelt, p. 33-44.

WILLIAMS, G. (2005), « Introduction », dans G. Williams (éd.), *La linguistique de corpus*, Rennes, Presses Universitaires de Rennes, p. 13-18.