

Nager à contre-courant ou comment le corpus parallèle devient une plateforme pour l'étude de la collocation

Mots-clés : Linguistique de corpus, corpus parallèle, corpus comparable, dictionnaire d'encodage, collocation, français, arabe.

La constitution des corpus ne suffit pas en soi à garantir une meilleure qualité au travail lexicographique, d'où l'intérêt à la fois d'une méthode de constitution et d'une méthode d'exploitation pour tirer profit de ces mines lexicales. Ces méthodes doivent entre autres répondre aux besoins particuliers de l'investigation lexicographique : diachronique, synchronique, etc. C'est dans le but de fabriquer un dictionnaire de collocations bilingue français – arabe, qu'une réflexion est menée sur la constitution du corpus et sur son exploitation.

Constitution de corpus :

La collocation, qui se prête tant bien que mal à la traduction et à l'appréhension d'un apprenant d'une langue étrangère, enfouit les origines du flou qui l'entoure dans sa nature profonde. Hausmann le dit clairement. D'après lui, « l'idiosyncrasie de la collocation ne se relève que dans l'optique d'une autre langue qui combine, pour exprimer le même fait, des mots différents ». C'est un phénomène contrastif qui se révèle en discours. C'est en partie la raison qui a motivé le recours à plusieurs types de corpus (corpus parallèle et corpus comparable) afin d'atteindre le but lexicographique fixé.

La langue arabe étant une langue sémitique, elle ne bénéficie pas de multiples logiciels et applications que la linguistique de corpus met à la disposition des langues indo-européennes. C'est donc une raison de plus pour envisager la conception d'une nouvelle méthode permettant l'exploitation d'un corpus qui facilite l'étude du phénomène dans la partie arabe sans le recours à des outils d'exploitation informatique.

Le choix porte donc sur l'association de plusieurs types de corpus (parallèle, comparable et monolingue) afin de tirer profit de leurs avantages respectifs et de pouvoir comparer leurs résultats dans un but lexicographique qui se révèle prometteur.

Stratégie de passerelle :

Il va de soi que le corpus ne fait pas l'approche, de cela le besoin a resurgi de développer une méthode d'exploitation qui optimise la combinaison de plusieurs corpus.

La stratégie de passerelle est une méthode d'exploitation qui repose sur l'idée que pour joindre deux collocations correspondantes potentielles dans deux langues différentes A et B, il est possible de se servir d'une passerelle. Lorsqu'on utilise cette passerelle, on arrive à l'autre bout, où il sera possible de vérifier l'existence de cette collocation correspondante potentielle.

Elle se résume concrètement en trois expressions-clés dans cette étude : corpus monolingue français comme point de départ en langue source, corpus parallèle bilingue comme passerelle entre la langue source et la langue cible et corpus comparable monolingue arabe comme corpus de contrôle.

Fondement théorique :

La stratégie de passerelle vise essentiellement à optimiser l'emploi des corpus parallèle et comparable ; les résultats des exploitations de ces deux types de corpus sont complémentaires.

Les corpus parallèle sont considérés comme point de départ des recherches contrastives, surtout en matière phraséologique. Même l'affichage de deux corpus, où chaque

énoncé est aligné en face de sa traduction, permet une consultation rapide et un repérage aisé de l'équivalence.

Les critiques à l'égard de la traduction limitent sans aucun doute la portée des corpus parallèles : ces corpus sont composés des textes et de leurs traductions. Mais les corpus parallèles ont au moins le mérite de faciliter les études contrastives.

Le corpus comparable surgit dès lors non pas comme alternative mais comme complément au corpus parallèle dans la lexicographie multilingue. Il surpasse en effet l'influence que le texte source peut exercer sur le texte cible par l'intermédiaire de la traduction. Le corpus comparable est censé refléter la langue telle quelle, y compris l'aspect extralinguistique.

Si la motivation initiale de la constitution de notre corpus et la conception de la méthode baptisée *stratégie de passerelle* sont initialement liées aux difficultés intrinsèques à l'exploitation du corpus arabe, les résultats obtenus portent fortement à croire dorénavant à une méthode adaptée à la recherche des phénomènes contrastifs. En réalité, le corpus comparable commence là où le corpus parallèle s'arrête. Mais le corpus parallèle est le prisme du lexicographe, du traducteur et du pédagogue. Certes ce postulat ne se conforme pas à la vision qui juge le corpus monolingue le plus à même pour l'étude de la collocation. D'ici à ce que notre postulat soit - nous l'espérons - confirmé, cette méthode fraiera son chemin à contre-courant.

Références bibliographiques:

- AL-QAISI, Fu'ad (2007) : *Etude des collocations entre langue et discours dans un corpus bilingue journalistique aligné*. Mémoire du master 2. Lyon : Université Lumière Lyon 2.
- AL-QAISI, Fu'ad (2015) : *Apport de la linguistique de corpus à la lexicographie bilingue (français-arabe) : Macrostructure et microstructure d'un dictionnaire de collocations*. Thèse de doctorat. Université Lumière Lyon 2.
- BEJOINT, Henri et THOIRON, Philippe (1992) : « Macrostructure et microstructure dans un dictionnaire de collocations en langue de spécialité ». In *Terminologie et Traduction*. n2/3. Luxembourg : office des publications officielles des communautés européennes. p. 513-527.
- HAUSMANN, Franz (1979) : « Un dictionnaire des collocations est-il possible? ». in *Travaux de Linguistique et de Littérature*. 17/1. p. 187-195.
- LANGLOIS, Lucie (1996) : *Bitexte, bi-concordance et collocation*. URL : <http://www.dico.uottawa.ca/theses/langlois/l1-debut.htm#resume>
- McENERY, Anthony et XIAO Zhonghua (2007) : « Parallel and Comporable Corpora : What are they up to ? URL : <http://www.research.lancs.ac.uk/portal/en/publications/parallel-and-comparable-corpora-what-are-they-up-to%28a72e1d6e-550f-49aa-b041-f6ccb2ab4fd9%29.html>
- REY-DEBOVE, Josette (2006) : « Lexicographie moderne ». In *Etudes Lexicographiques*, n5. Rabat : Association Marocaine des Etudes Lexicographiques. p. 19-32.
- SIEPMANN, Dirk (2005): « Collocation, Colligation and Encoding Dictionaries (part 1) ». In *International Journal of Lexicography*. Vol18. n4. Oxford : Oxford University Press. p. 409-443.
- SIEPMANN, Dirk (2006) « Collocation, Colligation and Encoding Dictionaries (part 2) ». In *International Journal of Lexicography*. Vol19. n1. Oxford : Oxford University Press. p. 1-38.
- SINCLAIR, John (1991) : *Corpus, Concordance, Collocation*. Oxford : Oxford University Press.
- SINCLAIR, John (2004): « Developing Linguistic Corpora: a Guide to Good Practice ». URL: <http://ahds.ac.uk/guides/linguistic-corpora/chapter1.htm>
- SVENSEN, Bo (2009) : *A Handbook of Lexicography: The Theory and Practice in Dictionary-Making*. Cambridge : Cambridge University Press p.162-163.
- TEUBERT, Wolfgang (2007) : « Corpus Linguistics and Lexicography ». In *Text Corpora and Multilingual Lexicography*. Wolfgang TEUBERT (dir.) Amsterdam : John Benjamin Publishing Company. p.109-133.
- TEUBERT, Wolfgang (1996) : « Comparable or Parallel Corpora? ». In *International Journal of Lexicography*. Vol9. n3. Oxford : Oxford University Press. p. 238- 264.
- ZANETTIN, Federicco (1998) : « Bilingual Comparable Corpora and The Training of Translators ». URL: <http://id.erudit.org/iderudit/004638ar>