

**Proposition de communication dans le cadre du colloque international :  
« Lexique(s) et genre(s) textuel(s) : approches sur corpus »**

Université Grenoble Alpes, 25-28 septembre 2018

**Abidi Rim**

[rimaabidi300@gmail.com](mailto:rimaabidi300@gmail.com)

*Institut Supérieur des Langues de Tunis*

**Vers une modélisation contrastive des verbes de mouvement et de déplacement à l'aide  
des corpus textuels**

À l'instar de la sociolinguistique, la psycholinguistique, le traitement automatique du langage TAL, la linguistique de corpus, née dans la deuxième moitié du XX<sup>e</sup> siècle, fut d'emblée promue discipline autonome, reconnue et en pleine expansion dans la sphère des domaines des recherches scientifiques grâce à son exploitation de l'outil informatique et des méthodologies innovantes. À ce propos, certains tentent de définir la linguistique de corpus en spécifiant qu'elle « est conçue et développée comme méthodologie visant à faire des découvertes sur une langue en général, pour soutenir la linguistique appliquée dans sa tâche. Les outils statistiques mis en place sont destinés à mettre à jour ce qu'il y a de commun entre toutes les occurrences d'un item lexical »<sup>1</sup>.

La réflexion développée dans la présente communication s'inscrit dans le cadre d'une étude générale proposée par le colloque, portant sur *la modélisation des unités lexicales à l'aide des corpus textuels*. En effet, il est question d'une modélisation linguistique prenant compte au mieux des spécificités syntaxiques et lexico-sémantiques relatives à l'emploi des verbes de mouvement et de déplacement dans une perspective contrastive, et ce, en exploitant différentes bases de données lexicales et textuelles. D'ailleurs, parmi les problématiques inhérentes à cette nouvelle branche, fondée sur une systématique scientifique computationnelle et baptisée la *linguistique de corpus*, nous soulevons celles de la représentativité et de l'exhaustivité de ces corpus constitués à partir de plusieurs ressources lexicologiques.

Ces multiples ressources exigent ipso facto un traitement automatique qui s'opère à travers des logiciels concordanciers tels que Antconc, WordSmith, Sketch Engine, ainsi que des logiciels aux fonctionnalités plus développées et plus profitables pour les investigations linguistiques comme *TermoStat* ou d'*INTEX*, appelé également *NooJ*. En procédant à une analyse des données longitudinales, le recensement de notre exemplier, qui se veut à la fois quantitatif et qualitatif, a porté sur des bases de données textuelles telle que *Frantext*<sup>2</sup>, ou les ressources lexicales spécifiques comme *NOVLEX* afin d'élucider et de modéliser le fonctionnement linguistique et cognitif propre aux unités étudiées.

---

<sup>1</sup> W. Teubert, *La linguistique de corpus : une alternative*. *Semen. Revue de sémio-linguistique des textes et discours* 27, 2009. < <https://semen.revues.org/8914> > [consulté le 10/03/2018].

<sup>2</sup> Nous avons fait référence à un exemplier assez large (environ 16 000 exemples, toutes valeurs confondues, dont 4 300 cas diversifiés étudiés plus en détails dans l'*annexe 1*) d'occurrences recueillies principalement dans *Frantext*, et un certain nombre pris dans *Google Books*, nous avons opté pour une version numérique en fichiers PDF, gravés sur un CD-ROM.

Dans cette perspective, notre travail explicatif des différentes typologies verbales : verbe d'état vs verbe d'action propose de désambigüiser ce découpage aberrant fondé sur l'emploi partiel d'un verbe d'action, et spécifie qu'il s'agit de deux types de procès : mouvement vs déplacement. À ce propos, c'est la définition d'Andrée Borillo qui précise que « le mouvement d'un objet peut être vu comme un changement de posture ou de position, mais qui ne va pas jusqu'à entraîner un véritable déplacement de l'endroit – de la "place" – où il se trouve » par opposition au déplacement qui « entraîne une modification des relations spatiales d'un objet avec son support des instants temporels successifs »<sup>3</sup>.

Il en ressort une réanalyse dudit découpage traditionnel pour le suppléer par une tripartition verbale cognitive inspirée des développements de Jean-Pierre Desclès, qui propose dans *langages applicatifs, langues naturelles et cognition* :

- « Les *situations statiques* de localisation (spatiale et temporelle) [...] expriment la position d'un objet par rapport à un lieu ».
- « Les *situations cinématiques* décrivent des mouvements dans un référentiel spatio-temporel (*bouger, se déplacer, rouler*) ou des changements d'états attribués à un objet (*grandir*) ».
- « Les *situations dynamiques* expriment non seulement des mouvements ou des changements d'état mais elles supposent une force externe, qui rend les modifications possibles »<sup>4</sup>.

Pour ce faire, nous avons procédé à un échantillonnage fructueux de ces verbes selon des annexes annotées qui se sont constituées à partir de ressources variées. Cette recherche laborieuse aboutissant à un corpus sélectif et réflexif a pu confirmer certaines hypothèses émises au départ afférentes à leurs régularités logico-formelles, à l'échelonnement de leurs valeurs sémantiques à partir des données lexicologiques et a permis, en dépit de la méthodologie contrastive de l'étude, de déceler une affinité opératoire entre l'approche linguistique classique et la sémantique cognitive.

De fait, il est impensable de procéder à une analyse scientifique du langage sans prendre en considération cet incontournable outil heuristique, considéré comme « la seule approche qui peut prétendre à être une *linguistique de la parole* »<sup>5</sup>, pour attester de la validité des processus examinés. Les problématiques théoriques, qui ont longtemps préoccupé plusieurs linguistes, se sont fondées avant tout sur un examen pratique de corpus pour décider de la recevabilité ou de l'agrammaticalité des énoncés en vue d'une récapitulation des normes canoniques de l'usage de la langue étudiée. Cela s'est opéré, en général, selon une démarche déductive qui s'est établie à partir d'un système d'inférences inspiré de la tradition aristotélicienne. Or, depuis la fin de la seconde guerre mondiale, les technologies numériques ont connu un essor prodigieux ayant accéléré l'exploitation des corpus. Ces derniers cessent d'être un simple moyen de validation ou adjuvant en matière pour devenir désormais l'objet central de la recherche linguistique appliquée misant davantage sur le raisonnement inductif dans l'acquisition de l'information sollicitée. Il est en d'autres termes, l'outil de mesure du bon usage, la preuve irréfutable qui atteste des multiples possibilités langagières. Il constitue la référence i.e. l'équivalent numérique des compétences développées du locuteur natif, le seul garant du bon usage. En cela,

---

<sup>3</sup> Borillo (A.), *L'espace et son expression en français*, Éditions Ophrys, Gap/Paris, 1998, p. 38.

<sup>4</sup> Desclès (J.-P.), *Langages applicatifs, langues naturelles et cognition*, Hermès, Paris, pp. 281- 292, 1990.

<sup>5</sup> W. Teubert, *op.cit.* < <https://semen.revues.org/8914> > [consulté le 10/03/2018].

dans nos exercices de recensement des distributions recevables, proprement dites grammaticales ou agrammaticales, cet outil de travail se conçoit d'ores et déjà comme l'indice de scientificité de l'analyse.

Cette investigation autour des approches sur corpus n'est aucunement mise, mais nous invite à repenser le devenir des traitements analytiques qui s'y réfèrent dans l'ère numérique. À cette fin, nous estimons que son bon fonctionnement est tributaire d'une maîtrise des objectifs de recherche qui tire avantage des nouvelles technologies numériques TAL, toujours renouvelées et actualisées (grâce à de nombreux logiciels disponibles sur la Toile). Le traitement automatique du langage devrait pouvoir s'appliquer à toutes les productions (textuelles et orales) répertoriées dans différents supports afin que ces investigations LTT (lexicologie, terminologie, traduction) appliquées sur corpus puissent accéder à une fiabilité optimisée dans leur élaboration.

**Mots-clés :** Modélisation – corpus textuels – lexicographie – verbe – mouvement – déplacement.

### **Bibliographie :**

Descès Jean-Pierre, *Langages applicatifs, langues naturelles et cognition*, Hermès, Paris, 1990.

DROUIN Patrick, LANGLAIS Philippe, *Évaluation du potentiel terminologique de candidats termes*, <http://lexicometrica.univ-paris3.fr/jadt/jadt2006/PDF/034.pdf>

KOEVA Svetla, MESFAR Slim, SILBERZTEIN Max, *Formalising Natural Languages with NooJ 2013: Selected Papers from the NooJ 2013 International Conference*.

L'Homme Marie-Claude, « Ressources lexicales, terminologiques et ontologiques : une analyse comparative dans le domaine de l'informatique », *Revue française de linguistique appliquée* 2008/1 (Vol. XIII), p. 97-118.

Maisonnasse Loïc *et al.*, « Approche statistique versus approche linguistique pour l'indexation sémantique des documents multilingues », *Document numérique* 2011/2 (Vol. 14), p. 193-214.

RASTIER François, « Enjeux épistémologiques de la linguistique de corpus », Lorient, septembre 2002, Actes à paraître sous la direction de Geoffrey Williams aux *Presses Universitaires de Rennes*, p. 2. < [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html) >

SINCLAIR John McHardy, *Corpus, Concordance, Collocation*, Oxford, Oxford University Press, 1991

SINCLAIR John McHardy, JONES Susan, DALEY Robert, *English Collocation Studies: The OSTI Report*, Londres - New York, Continuum, 2004.

Max Silberztein, « La formalisation du dictionnaire LVF avec NooJ et ses applications pour l'analyse automatique de corpus », *Langages* 2010/3 (n° 179-180), p. 221-241.

TEUBERT Wolfgang, *La linguistique de corpus : une alternative. Semen. Revue de sémio-linguistique des textes et discours* 27, 2009. < <https://semen.revues.org/8914> >

TUTIN Agnès, « Les collocations lexicales : une relation essentiellement binaire définie par la relation prédicat-argument », *Langages* 2013/1 (n° 189), p. 47-63.