

Proposition CILPR 2013
Section 16 – Projets en cours, ressources et outils nouveaux

Statistiques et motifs séquentiels
Le cas des suites *Nom Préposition Nom, mot à mot, pétale par pétale*

Véronique Magri-Mourgues
Université de Nice, UMR 7320, Bases, Corpus et Langage

Enjeu de la communication

La communication se propose d'évaluer l'outil statistique pour identifier et analyser les suites *Nom préposition Nom*, avec pour critère restrictif que les deux substantifs sont identiques. L'enjeu essentiel de cette recherche est de montrer que ce type de structures peut être utilisé à des fins de classifications.

Corpus d'étude

La base qui sert de corpus exploratoire à cette recherche est constituée par un extrait de *Frantext* catégorisé : il comprend quatre groupes génériques répertoriés dans *Frantext* : d'une part deux ensembles narratifs, le roman et les mémoires, d'autre part la poésie et le théâtre afin que les genres principaux soient tous représentés.

Méthodologie

Le premier travail consiste à repérer les suites recherchées. Il s'avère que toutes les prépositions ne sont pas aptes à entrer dans ce type de construction. Seules seront retenues *à, après, contre, par, pour, sur* ainsi que les structures complexes *de...à* et *de...en*. Une réflexion pourra s'engager sur le statut des prépositions et sur leur portée sémantique. Le rendement des structures est évalué en observant les variations de substantifs de part et d'autre des prépositions : la préposition peut varier tandis que les noms restent identiques et l'expression conserve globalement le même sens (*mot à mot, mot pour mot ; bouche à bouche, bouche contre bouche* ; la préposition peut être la même tandis que les substantifs peuvent varier (*nuit après nuit, soir après soir* : le sens exprime ici l'idée de succession ; *rendre haine contre haine, corps à corps, nez contre nez* : le sens se module en l'idée d'échange, d'opposition, de contact).

L'analyse a aussi à composer avec les structures lexicalisées ; on s'intéressera ainsi aux critères et aux fonctions statistiques qui autorisent une interprétation sur le degré de figement des structures. Entre phraséologie et style d'auteur, conçu comme appropriation des unités linguistiques, entre langue et discours, ces unités seront évaluées en fonction de leur productivité dans le corpus d'étude. L'observation des hapax – ou unités de fréquence 1 – sera engagée comme possible critère distinctif de l'innovation lexicale. L'étude amorcera par ce biais une délimitation et une possible définition des unités lexicalisées.

On envisagera par ailleurs l'environnement cotextuel des unités considérées afin d'observer la fonction syntaxique assurée dans la phrase ainsi que la place occupée dans la séquence textuelle. On verra si une fonction structurante peut leur être attribuée.

Comme le corpus est constitué de quatre ensembles génériques, il sera loisible de voir enfin comment ces unités phrastiques se distribuent et si une répartition générique peut être proposée. On observera par exemple que l'ensemble *Poésie* est le plus propice à l'innovation lexicale et établit des liens privilégiés avec le sémantisme de la succession par exemple.

Cette étude sera menée en utilisant les représentations de l'analyse factorielle, de l'analyse arborée. Un logiciel sera mis à contribution, *Hyperbase*, logiciel d'analyse hypertextuelle mis au point par Étienne Brunet, qui a été professeur à l'Université de Nice et membre de l'UMR 7320, *Bases, Corpus et Langage*.

Bibliographie

BIBER D., CONRAD S., CORTES V. (2004). "If you look at... : Lexical bundles in University teaching and textbooks", *Applied Linguistics* 25/3, Oxford University Press, 371-405.

BRUNET É. (2011). *Hyperbase*. Logiciel hypertexte pour le traitement documentaire et statistique des corpus textuels. Manuel de référence. Disponible: <ftp://ancilla.unice.fr/manuel.pdf>

BRUNET É. (2007). *Fréquences et séquences. Mise en œuvre dans Hyperbase*. In: *Lexicometrica: Topographie et topologie textuelles*. <http://lexicometrica.univ-paris3.fr/numspeciaux/special9/brunet.pdf>.

BRUNET É (1995) - « *Cardiogrammes* », in François RASTIER (dir.), *L'analyse thématique des données textuelles*, Paris : Didier, p. 107-145.
<http://www.revue-texto.net/Parutions/Analyse-thematique/Brunet.pdf>

LONGRÉE D., LUONG X., MELLET S. (2008). « Les motifs : un outil pour la caractérisation topologique des textes », *JADT*, 733-744.

GROSS G. (1996). *Les expressions figées en français-Noms composés et autres locutions*, Paris : Ophrys.

Langages (2009), n° 173. « Approches récentes de la préposition ».

LE PESANT D. (2006). « Esquisse d'une classification syntaxique des prépositions simples du français », in D. Leeman & C. Vaguer (dir.). *Modèles Linguistiques XXVII-1*, vol. 53. Toulon : Université du Sud et Editions des Dauphins.

MEJRI S. (1997). *Le figement lexical. Descriptions linguistiques et structuration sémantique*, Tunis : Publications de la Faculté des lettres Manouba 1997, 632 p.

MELIS L. (2003). *La Préposition en français*, Paris : Ophrys.

RENOUF, A., & SINCLAIR J.-M. (1991). "Collocational frameworks in English. In K. Aijmer, & B. Altenberg", *English corpus linguistics*, New York : Longman, p. 128-143.