

## **Elaboration d'un dictionnaire électronique des phrasèmes nominaux de la langue générale**

**Inès SFAR  
Université de Sousse-TIL  
Université Paris 13-LDI**

Un certain consensus semble exister sur le fait que tout système de traitement automatique du langage naturel se doit de disposer de dictionnaires électroniques aussi complets et explicites que possible. De même, il est pleinement admis à l'heure actuelle, que ce soit dans l'élaboration de produits industriels que dans la majorité des systèmes universitaires, qu'il est nécessaire de décrire des suites contenant des caractères séparateurs (dites phrasèmes, composés, unités phraséologiques...), ce qui a conduit à consacrer une grande énergie à la construction de bases de données lexicales recensant en extension ces unités et permettant, le cas échéant, de les aligner à des variantes ou à des équivalents de traduction. Dans bon nombre de cas, l'élaboration de ces outils se fait de façon quasi-automatique, soit en ayant recours à des procédés de base statistique, soit en procédant à la reconnaissance automatique des termes par des systèmes de repérage axés sur le contexte. Pour notre part, tout en reconnaissant les immenses avantages de ces procédures pour ce qui est de l'extraction de candidats-composés, il nous semble que l'appellation "dictionnaire" devrait être strictement réservée à des bases lexicales élaborées par des linguistes-lexicographes.

Plusieurs centres de recherche ont ressenti le besoin de disposer de répertoires exhaustifs de noms composés de la langue générale, insuffisamment décrits dans les différentes traditions lexicographiques qui ont toujours eu tendance à sous-estimer et leur nombre et leur caractère non compositionnel. Si la question semble bien assise d'un point de vue théorique, la pratique en lexicographie électronique ne semble pas en avoir tiré toutes les conséquences.

Nous comptons présenter les résultats d'un grand projet de coopération scientifique (PPICS) réalisé entre deux équipes française et espagnole: le laboratoire LDI (CNRS-UMR 7187) de l'Université Paris 13 et le laboratoire FlexSem de l'Université Autonome de Barcelone. Le travail conjoint de ces deux équipes qui ont consacré des efforts soutenus et coordonnés à l'étude et recensement des suites phraséologiques nominales en langue générale pour quatre langues de travail, a donné lieu à des dictionnaires monolingues coordonnés dépassant les 500 000 entrées (français-espagnol-catalan), dont nous nous contentons de présenter la partie française. Et cela parce que, d'une part, ces suites constituent un écueil en traitement automatique du fait de l'imprévisibilité de leurs composantes, parce qu'elles posent des difficultés particulières en traduction-adaptation (où elles peuvent, de surcroît, induire à de graves erreurs) et parce qu'il s'agit de suites qui peuvent présenter une remarquable opacité sémantique qui exige de disposer, non seulement d'équivalents de traduction, mais aussi de paraphrases intralinguistiques permettant de normaliser des textes venant de différents horizons d'une même communauté linguistique.

Notre méthodologie bénéficie de moyens automatisés d'exploitation de corpora textuels, aussi bien statistiques (en collaboration avec des spécialistes en *data mining*) qu'algebriques,

mais il s'agit d'une démarche essentiellement lexicographique dont l'essentiel du travail concerne la description linguistique. Les principaux repères théoriques qui la sous-tendent sont, d'une part, le modèle des classes d'objets développé au propre LDI de Paris 13 et, d'autre part, la pratique lexicographique la plus récente, introduisant une approche linguistique qui tient compte du phrasème dans ses contextes. Le dictionnaire électronique que nous exposons présente une macrostructure et une microstructure standard par rapport à celles qui ont été employées dans d'autres projets de grande envergure et permettent ainsi un échange immédiat de données entre groupes.

### **Bibliographie**

- Blanco, X., Mejri, S. (2006) : *Les noms de professions. Approches Linguistiques, Contrastives et Appliquées*, Servei de Publicacions de la Universitat Autònoma de Barcelona, ISBN : 84-490-2476-5, 226 pages.
- Blanco, X., Buvet, P.-A., Gavriilidou, Z. (2001) : Détermination et Formalisation, John Benjamins Publishing Company: Amsterdam/Philadelphia ISBN: 90-272-3133-8 (Europa) / 1 58811 096 6 (USA), pp. 1-345.
- Blanco, X., Buvet, P.-A. (2004) : Verbes supports et significations grammaticales. Implications pour la traduction espagnol-français. *Lingvisticæ Investigationes. Revue internationale de linguistique française et générale*. Tome XXVII : II. John Benjamins Publishing Company: Amsterdam/Philadelphia.
- Blanco, X., Buvet, P.-A. (2000) : De l'analyse syntactico-sémantique du lexique à la traduction automatique, *Bulag 25*. Centre Lucien Tesnière: Université de Franche-Comté, pp. 69-88.
- Blanco, X., Buvet, P.-A. (1999) : À propos de la traduction automatique des déterminants de l'espagnol et du français. *Meta. Journal des Traducteurs* 44:4. Les Presses de l'Université de Montréal, pp. 525-545.
- Blanco, X., Lajmi, D. (2004) : Dictionnaire électronique français-espagnol-catalan-arabe des noms des professions et des métiers (en colaboración con Douha Lajmi), *Premières Journées Scientifiques des Réseaux de Chercheurs de l'Agence Universitaire de la Francophonie*, Ouagadougou (Burkina Faso), 1-5 juin 2004. Agence Universitaire de la Francophonie et Université d'Ouagadougou.
- Blanco, X., Buvet, P.-A., Gavriilidou, Z., Greenfield, P. (2000) : Comment traduire les « bruits » ? Mejri, S.; Clas, A.; Baccouche, T.; Gross, G. (eds): *La traduction : théories et pratiques*. Publications de l'École Normale Supérieure: Tunis, pp. 21-42.
- Blanco, X., Buvet, P.-A., Gavriilidou, Z. (1999) : Analyse comparée des modifieurs figés en grec moderne, français et espagnol. *Studies in Greek Linguistics. Proceedings of the 20<sup>th</sup> Annual Meeting of the Department of Linguistics*. Aristotle University of Thessaloniki, p. 87-98
- Blanco, X., Buvet, P.-A. (1998) : Perspectives pour la traduction automatique des déterminants *Actes de la Cinquième Conférence Annuelle sur le Traitement Automatique des Langues Naturelles*. Département d'Intelligence Artificielle et Médecine, Pitié-Salpêtrière: Paris, pp. 239-242