

NOM Prénom : Gasiglia Nathalie

Nature de la mission : participation à un colloque, les *Journées Internationales de Linguistique de Corpus*

Lieu et date : Université de Bretagne Sud (Lorient), 4-6 septembre 2013 (présence effective le 6)

Frais de mission attribués par le laboratoire : 371,70 euros

Description de la mission (par ex. résumé de l'intervention proposée/activités de recherche réalisées au cours de la mission...) :

Les *Journées Internationales de Linguistique de Corpus* réunissent régulièrement, depuis 2001, des chercheurs qui abordent de manières variées des analyses linguistiques, psycholinguistiques, sociolinguistiques, etc. ou qui documentent en corpus des descriptions lexicographiques, ou encore (plus occasionnellement) qui explorent des corpus dictionnaires pour des études métalexographiques.

Relevant de ce dernier type, le travail de Mahfoud Mahmoud (Université de Rouen) [*Sens et interprétations du mot justice dans un corpus de dictionnaires bilingues de la période coloniale en Algérie*], exploite le texte de 13 dictionnaires français-arabe de l'époque coloniale afin d'analyser les emplois du nom *justice* dans les articles consacrés à ce mot, et dans l'ensemble des définitions ou l'ensemble des textes. Il fait ainsi ressortir la prépondérance des emplois relatifs à la justice française sur ceux relatifs aux justices indigènes.

Articulant une approche lexicographique ancrée sur les CPA de Patrick Hanks et taliste basée sur l'exploitation des patrons syntaxiques fournis par le dictionnaire *Les verbes français* (Jean Dubois & Françoise Dubois-Charlier), l'analyse d'un corpus de textes scientifiques français proposée par Rui Yan et Agnès Tutin (Université Stendhal Grenoble III) [*Un modèle lexicographique basé sur le corpus Scientext comparé avec d'autres ressources lexicographiques : l'exemple des constructions verbales*] vise le repérage des constructions régulièrement attestées pour une sélection de verbes afin d'élaborer des descriptions lexicographiques susceptibles de soutenir les efforts consentis par de jeunes chercheurs sinophobes qui rédigent leurs premières productions scientifiques en français.

D'autres approches outillées ont des visées moins directement applicatives et peuvent soutenir des analyses littéraires ou linguistiques.

Marie-Hélène Lay (Université de Poitiers) présentait un outil d'exploration de corpus qu'elle a développé pour et en collaboration avec des collègues littéraires [*AnaLog, outil de*

représentation des textes au service des humanités numériques], en attachant une attention toute particulière à leurs besoins documentaires et à leurs souhaits en matière d'exploitation des informations fournies par cet outil.

D'autres exposés se concentraient sur des données extraites sur des bases statistiques ou avec des méthodologies d'analyse quantitatives : le travail que nous avons présenté Ann Bertels (KU Leuven, ILT & QLVL) et moi [*Conjuguer des approches quantitatives et linguistiques pour l'analyse d'un corpus footballistique*], celui de Itsuko Fujimura (Université de Nagoya) et Nobushige Aoki (Université de Gunma) [*Typologie des collocations françaises et anglaises basée sur Log-r*] ou encore celui de Hai Hieu Vu et Jeanne Villaneau (IRISA, UBS), et de Farida Saïd (LMBA, UBS) [*Utilisation des liens wikipedia pour la détection automatique des concepts d'un domaine*].

La contribution présentée par Ann Bertels et moi relatait une recherche effectuée durant le stage d'Ann Bertels à l'UMR STL :

Nous avons montré comment nous avons conjugué des approches quantitatives et linguistiques afin d'améliorer la pertinence des résultats d'une étude des valeurs d'unités lexicales ou supralexicales d'un corpus de commentaires de matchs de football. Dans un premier temps, nous avons procédé à une double analyse quantitative qui nous a permis d'évaluer, sur des bases cooccurentielles, la valeur sémantique des mots-clés du corpus footballistique. Ensuite, ces observations quantitatives ont été complétées par une analyse qualitative des mots-clés les plus spécifiques. Dans le but de mieux comprendre le comportement combinatoire des unités simples, nous avons procédé finalement à une exploration des unités polylexicales pour une sélection de mots-clés utilisés comme mots-pôles. Cette exploration fait intervenir les cooccurents les plus pertinents d'un point de vue statistique, ainsi que les constituants syntaxiques (chunks) repérés automatiquement. La comparaison des résultats permet de mettre en évidence la complémentarité des approches quantitatives et linguistiques.

Bénéfice de la mission (pour le chercheur/l'enseignant-chercheur, pour le laboratoire) :

Après le stage qu'Ann Bertels a effectué au début de l'année 2013 à l'UMR STL et que nous avons consacré à l'extraction d'information dans un corpus de commentaires footballistiques que j'avais constitué avec Hans Paulussen (ITEC, KULAC) et à leur manipulation en exploitant un outillage statistique, nous étions désireuses de présenter nos travaux et de voir comment ils peuvent être appréciés. Les *Journées Internationales de Linguistique de Corpus* constituaient un lieu stimulant pour ce type d'exposé.