

Sujet de stage de M2R :
Le repérage automatique des gloses

Encadrement : Augusta Mela (Univ. de Montpellier 3) et Mathieu Roche (Univ. de Montpellier 2)

Équipe TAL, LIRMM, UMR 5506, Université Montpellier 2
Augusta.Mela@univ-montp3.fr et mathieu.roche@lirmm.fr

1. Travail demandé

À partir de descriptions réalisées par les linguistes sur les formes et les sens du phénomène linguistique de *la glose*, le stagiaire vérifiera sur différents types de corpus de textes numérisés, dans quelle mesure les descriptions de forme proposées permettent effectivement de discriminer les gloses et il dressera un bilan, par formes et par types de textes, de ces mesures. L'étude effectuée au cours de ce stage sera principalement centrée sur des textes en français. Cependant, des comparaisons avec l'anglais pourraient également être menées.

Le stagiaire utilisera un outil d'annotation morpho-syntaxique de corpus (comme l'étiqueteur d'E. Brill) et mettra au point des filtres ou patrons (patterns) des séquences qui peuvent constituer des gloses.

2. Définition de la glose

Qu'il s'agisse d'évoquer une langue étrangère ou spécialisée (1), de procéder à une explication didactique (2), ou de s'assurer que l'interlocuteur attribue la signification adéquate au mot employé, le langage courant fournit de nombreux exemples de commentaires en situation parenthétique qui traduisent, expliquent le sens des mots en discours :

(1) Si l'on admet que l'état d'un électron n'est pas entièrement décrit par sa position et sa vitesse de translation dans l'espace, mais qu'il est animé en outre d'un pivotement sur lui-même, *ou « spin »*, mouvement essentiellement quantifié : son moment cinétique est d'un demi quantum et crée un moment magnétique égal à un magnéton de Bohr. (Hist. gen. sciences, 1964, t.3, vol.2)

(2) Ce sont la dépigmentation, *c'est-à-dire l'absence quasi totale des éléments colorés dermiques qui s'opposent normalement à l'action nocive des rayons ultra-violetts d'origine solaire* et l'anophtalmie, *ou réduction de l'appareil oculaire allant le plus souvent jusqu'à sa disparition complète*. (Geze, La spéléologie scientifique, 1965)

On appelle gloses ces commentaires sur le mot. Les gloses relèvent donc d'une activité métalinguistique. Cette activité métalinguistique se manifeste dans le discours écrit par des marques typographiques comme les guillemets et par des éléments lexicaux comme *appelé, c'est-à-dire, signifiant, ou*. Dans ce cas, c'est non seulement la fonction métalinguistique qui est marquée mais aussi la nature même de la relation de sémantique lexicale mise en jeu : équivalence avec *c'est-à-dire, ou* ; spécification du sens avec *au sens* ; nomination avec *dit, baptisé*, hyponymie avec *en particulier, comme*, hyperonymie avec *et/ou autre(s)*, etc.

Ces éléments lexicaux peuvent se combiner à d'autres indices métalinguistiques : d'autres éléments lexicaux comme *terme, mot*, des marques typographiques comme la virgule, la parenthèse, le tiret, les guillemets, à l'écrit, une intonation spécifique ou un geste mimant les guillemets, à l'oral.

Grâce à l'existence de ces marques et aux propriétés syntaxico-sémantiques des constructions auxquelles elles sont régulièrement associées, un repérage automatique des gloses est réalisable (Mela,2004 ; Mela,2005).

3. Perspectives d'application

Les perspectives d'application du repérage des gloses sont nombreuses :

- Pour l'acquisition de termes, de néologismes et leurs définitions :

Les gloses peuvent servir à la fois à pointer les termes spécialisés, les mots nouveaux, les emprunts non "codifiés", les sigles et acronymes, et leurs définitions ;

- Pour la classification de textes :

par superposition des listes des termes pointés par la glose et de celles des terminologies existantes ;

- Pour la typologie des textes : quelles formes de glose pour quel type de média ? quel type de situation de communication (vulgarisation, pédagogie) ? et comment les gloses se manifestent suivant le média ?

Références

MELA A. (2004). Linguistes et 'talistes' peuvent coopérer : repérage et analyse des gloses. *Revue Française de Linguistique Appliquée*, IX(1), « Linguistique et informatique : nouveaux défis », B. Habert (resp.), <<http://www.univ-montp3.fr/~amela/Publications.html>>.

MELA A. (2005). Le repérage automatique des gloses de nomination seconde. *Langues et langage*, 12, «Les marqueurs de la glose». <<http://www.univ-montp3.fr/~amela/Publications.html>>.

STEUCKARDT, A. et NIKLAS-SALMINEN, A. (éd.) (2004) : Le mot et sa glose. *Langues et langage*, 9, Aix-en-Provence, Publications de l'Université de Provence.

STEUCKARDT, A. et NIKLAS-SALMINEN, A. (éd.) (2005) : Les marqueurs de la glose. *Langues et langage*, 12, Aix-en-Provence, Publications de l'Université de Provence.