Seminar of Computional Linguistics Group

Monday 14th June 2010 at 10h30

Elsa TOLONE

Gaspard-Monge Institute, University of Paris-Est

Using Lexicon-Grammar tables for French in a parser

Lexicon-Grammar tables (Gross 1975) are a very rich syntactic lexicon for the French language. The covered parts of speech are verbs, predicatives nouns, frozen expressions and simples and frozen adverbs. This linguistic database is nevertheless not directly suitable for use by computer programs, as it is incomplete and lacks consistency. Tables are defined on the basis of features which are not explicitly recorded in the lexicon. These features are only described in literature. Our aim is to define for each tables this essential properties to make them usable in various Natural Language Processing (NLP) applications. Then, we describe how, thanks to the work described in (Constant & Tolone 2010) and (Tolone & Sagot 2009), we converted the Lexicon-Grammar tables of French verbs into the Le*fff* format (Sagot 2010), named LGLex lexicon, which allowed us to integrate it into the FRMG parser (Thomasset & de La Clergerie 2005). We compare the results of the FRMG parser on the EASy reference corpus depending on whether it relies on the entries of the Le*fff* or those of the converted Lexicon-Grammar tables.

References :

(Constant & Tolone 2010) Matthieu Constant and Elsa Tolone. A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables. Lingue d'Europa e del Mediterraneo, Grammatica comparata, pp. 79-93. Aracne. 2010.

(Gross 1975) Maurice Gross. Méthodes en syntaxe : Régime des constructions complétives. Hermann. Paris, France. 1975.

(Sagot 2010) Benoît Sagot. The Le*fff*, a freely available and large-coverage morphological and syntactic lexicon for French. Proceedings of the 7th Language Resource and Evaluation Conference (LREC'10), 8 pp. La Valette, Malta. 2010.

(Thomasset & de La Clergerie 2005) François Thomasset and Éric de La Clergerie. Comment obtenir plus des méta-grammaires. Proceedings of the 12nd Traitement Automatique des Langues Naturelles conference (TALN'05), 10 pp. Dourdan, France. 2005.

(Tolone 2009) Elsa Tolone. Les tables du Lexique-Grammaire au format TAL. Proceedings of MajecSTIC 2009 conference, 8 pp. Avignon, France. 2009.

(Tolone & Sagot 2009) Elsa Tolone and Benoît Sagot. Using Lexicon-Grammar tables for French verbs in a large-coverage parser. Proceedings of the 4th Language & Technology Conference (LTC'09), pp. 200-204. Poznań, Poland. 2009.