Séminaire de l'Équipe Informatique Linguistique

Lundi 14 juin 2010 à 10h30

Elsa TOLONE

Institut Gaspard-Monge, Université Paris-Est

Utilisation des tables du Lexique-Grammaire dans un analyseur syntaxique

Les tables du Lexique-Grammaire (Gross 1975) constituent un lexique syntaxique très riche pour le français. Les catégories grammaticales couvertes sont les verbes, les noms prédicatifs, les expressions figées et les adverbes simples et figés. Cette base de données linguistique n'est cependant pas directement exploitable informatiquement car elle est incomplète et manque de cohérence. Chaque table regroupe un certain nombre d'entrées jugées similaires car elles acceptent des propriétés communes qui ne sont pas codées dans les tables mais uniquement décrites dans la littérature. Notre objectif est donc de définir pour chaque table ces propriétés indispensables à l'utilisation des tables dans les diverses applications de Traitement Automatique des Langues (TAL). Ensuite, nous montrons comment, grâce au travail décrit dans (Constant & Tolone 2010) et (Tolone & Sagot 2009), nous avons converti les tables du Lexique-Grammaire en un lexique au format Le*fff* (Sagot 2010), appelé lexique LGLex, permettant ainsi son intégration dans l'analyseur syntaxique FRMG (Thomasset & de La Clergerie 2005). Nous comparons les résultats de l'analyseur syntaxique FRMG sur le corpus de référence de la campagne EASy selon qu'il utilise les entrées du Le*fff* ou celles des tables du Lexique-Grammaire ainsi converties.

Références :

(Constant & Tolone 2010) Matthieu Constant et Elsa Tolone. A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables. Lingue d'Europa e del Mediterraneo, Grammatica comparata, pages 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. Actes du 7ème Language Resource and Evaluation Conference (LREC'10), 8 pages. La Valette, Malte. 2010.
(Thomasset & de La Clergerie 2005) François Thomasset et Éric de La Clergerie. Comment obtenir plus des méta-grammaires. Actes de la 12ème conférence sur le Traitement Automatique des Langues Naturelles (TALN'05), 10 pages. Dourdan, France. 2005.
(Tolone 2009) Elsa Tolone. Les tables du Lexique-Grammaire au format TAL. Actes de la conférence MajecSTIC 2009, 8 pages. Avignon, France. 2009.
(Tolone & Sagot 2009) Elsa Tolone et Benoît Sagot. Using Lexicon-Grammar tables for French verbs in a large-coverage parser. Actes du 4ème Language & Technology Conference (LTC'09), pages 200-204. Poznań, Pologne. 2009.