

# Rapport d'avancement DFD - 3<sup>e</sup> année de thèse

Yannick PARMENTIER – Projet Langue Et Dialogue – Sept. 2006

## Sujet de thèse

**Titre :** « *Développement d'une architecture logicielle pour la construction sémantique à partir de Grammaires d'Arbres Adjoints de taille réelle* ».

**Problématique :** A l'heure actuelle, il n'existe pas, pour les grammaires d'arbres adjoints (*Tree Adjoining Grammars - TAG*), de système de construction sémantique supportant les grammaires de taille réelle. Un tel système permet d'associer à une phrase, non seulement sa structure syntaxique mais également sa représentation sémantique (i.e. une représentation du sens de cette phrase, généralement une formule appartenant à un langage logique).

Ce sujet de thèse vise à spécifier, développer et évaluer des outils informatiques permettant la construction sémantique basé sur les TAGs. Plus précisément, nous souhaitons mettre en place une architecture permettant notamment :

- (1) la production semi-automatique d'une grammaire TAG à portée sémantique pour le Français,
- (2) la validation de cette grammaire,
- (3) la construction automatique, par le biais de cette grammaire, d'une représentation du sens d'énoncés en langue naturelle.

## Travail Réalisé

En 1<sup>ère</sup> année de thèse, mon travail a concerné le développement d'outils pour disposer des ressources nécessaires à la construction sémantique, la ressource principale étant une grammaire intégrant une dimension sémantique. Dans cet optique j'ai développé, en collaboration avec Joseph Le Roux, doctorant du projet Calligramme, ce que l'on appelle un *compilateur de métagrammaire*<sup>1</sup>. Un tel outil permet de construire une grammaire de taille importante à partir d'une spécification factorisée (la métagrammaire). L'intérêt de notre compilateur, nommé XMG, réside dans le fait qu'il permet d'annoter la grammaire avec des informations sémantiques, et également qu'il est efficace (grammaire de taille réelle, plusieurs milliers d'arbres, compilée en 15 min sur une architecture standard).

Lors de la 2<sup>nde</sup> année de thèse, j'ai travaillé sur l'interfaçage entre XMG et un analyseur syntaxique, dans le but non seulement de valider les grammaires produites mais également de mettre en place un procédé de construction sémantique basé sur l'analyse syntaxique.

La validation de grammaires produites automatiquement, a été réalisé dans le cadre de la thèse de Benoît Crabbé au LORIA.

J'ai alors développé un module de construction sémantique utilisant un générateur d'analyseurs syntaxiques pour TAGs, générateur plus connu sous le nom de DyALog. Ce module pré-traite la grammaire produite automatiquement avec XMG pour avoir deux ensembles de règles grammaticales (les premières syntaxiques et les secondes sémantiques). Il utilise alors le système DyALog pour générer l'analyseur dont la couverture correspond à celle des règles syntaxiques et utilise le résultat de l'analyse pour resynchroniser les informations syntaxiques et sémantiques. En sortie, ce module produit une représentation sous forme prédicative du (des) sens de l'énoncé analysé.

En 3<sup>e</sup> année de thèse, nous avons travaillé au développement des ressources pour la construction sémantique. Ainsi Claire Gardent a pu étendre la métagrammaire du Français de Benoît Crabbé pour y intégrer une dimension sémantique. Cette métagrammaire a permise de générer, au moyen de XMG, la première grammaire TAG du Français de taille réelle à portée sémantique. Un lexique de taille réduite a en outre été écrit. Nous avons alors commencé l'évaluation de nos ressources et de nos outils en nous basant sur la suite de test TSNLP. Cette évaluation est toujours en cours actuellement.

Enfin, j'ai commencé la rédaction de mon mémoire de thèse (80 pages à l'heure actuelle).

---

<sup>1</sup>disponible librement à l'adresse <http://sourcesup.cru.fr/xmg>, et enregistré auprès de l'INRIA sous licence CeCILL.

## Perspectives de travail pour la 4<sup>e</sup> année de thèse

Le travail restant à faire durant cette quatrième année de thèse concerne (a) la finalisation de l'évaluation de notre procédé de construction sémantique et (b) la suite de la rédaction du mémoire de thèse.

Concernant (a), nous avons adapté nos outils pour pouvoir les utiliser en mode batch et ainsi automatiser la construction sémantique pour pouvoir traiter des corpus. Cependant, le passage à l'échelle n'est pas immédiat, puisque nous découvrons des erreurs dans les ressources qu'il nous faut corriger. Nous sommes proches d'une version stable.

Concernant (b), le plan de thèse est le suivant :

1<sup>ère</sup> partie – La construction sémantique pour Grammaires d'Arbres Adjoints : état de l'art.

1. La construction sémantique en Linguistique Informatique : de Montague aux Sémantiques Plates avec variables d'unification.
2. La construction sémantique pour Grammaires d'Arbres Adjoints : approches existantes.
  - 2.1 Les Grammaires d'Arbres Adjoints.
  - 2.2 Les approches basées sur l'arbre de dérivation.
  - 2.3 Les approches basées sur l'arbre dérivé.

2<sup>e</sup> partie – Compilation de Grammaires d'Arbres Adjoints à portée sémantique.

3. La production semi-automatique de Grammaires d'Arbres Adjoints.
4. XMG : un formalisme métagrammatical extensible.
5. XMG : un compilateur de métagrammaires multi-paradigme.
6. Cas pratique : description d'une grammaire à portée sémantique.

3<sup>e</sup> partie – La construction sémantique pour Grammaires d'Arbres Adjoints : Implantation et évaluation.

7. L'analyse syntaxique comme base de la construction sémantique.
8. Procédé d'évaluation et analyse des résultats.

## Publications

1. **Coreference handling in XMG.** Claire Gardent et Yannick Parmentier. *Actes de la Conférence jointe du Comité International sur la Linguistique Informatique et de l'Association pour la Linguistique Informatique (COLING/ACL2006), Poster*, Sydney, juillet 2006.
2. **A constraint driven metagrammar.** Joseph Le Roux, Benoît Crabbé et Yannick Parmentier. *Actes de l'atelier sur les Grammaires d'Arbres Adjoints et formalismes liés (TAG+8)*, Sydney, juillet 2006.
3. **SemTAG, the LORIA toolbox for TAG-based Parsing and Generation.** Eric Kow, Yannick Parmentier et Claire Gardent. *Actes de l'atelier sur les Grammaires d'Arbres Adjoints et formalismes liés (TAG+8), Poster*, Sydney, juillet 2006.
4. **XMG - An expressive formalism for describing tree-based grammars.** Yannick Parmentier, Joseph Le Roux et Benoît Crabbé. *Actes de la 11<sup>e</sup> Conférence de la section Européenne de l'Association pour la Linguistique Informatique (EACL 2006), Session de Démonstrations*, Trento, avril 2006.
5. **XMG : a Multi-formalism Metagrammatical Framework.** Yannick Parmentier et Joseph Le Roux. *Actes de la session étudiante de la 17<sup>e</sup> Ecole d'Ete Européenne en Logique, Linguistique et Informatique (ESSLLI 05)*, Edimbourg, août 2005.
6. **Large scale semantic construction for Tree Adjoining Grammars.** Claire Gardent et Yannick Parmentier. *Actes de la 5<sup>e</sup> conférence sur les aspects logiques de la Linguistique Informatique (LACL 05)*, Bordeaux, avril 2005.
7. **The Metagrammar Compiler : An NLP Application with a Multi-paradigm Architecture.** Denys Duchier, Joseph Le Roux et Yannick Parmentier. *Actes de la 2<sup>nd</sup> conférence internationale Oz / Mozart*, Charleroi, octobre 2004.
8. **XMG : Un Compilateur de Méta-Grammaires Extensible.** Denys Duchier, Joseph Le Roux et Yannick Parmentier. *Actes de la 12<sup>e</sup> conférence sur le Traitement Automatique des Langues Naturelles (TALN 05)*, Dourdan, juin 2005.