



Polymnie - Analyse et synthèse avec les Grammaires Catégorielles Abstraites : du lexique au discours

Sylvain Pogodalla

► To cite this version:

Sylvain Pogodalla. Polymnie - Analyse et synthèse avec les Grammaires Catégorielles Abstraites : du lexique au discours. Les rencontres du numérique de l'ANR, Nov 2016, Paris, France. hal-01417895

HAL Id: hal-01417895

<https://hal.inria.fr/hal-01417895>

Submitted on 16 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Coordinateur : INRIA Nancy Partenaires : INRIA Paris, LaBRI—CNRS, IRIT—U. Toulouse

Abstract: Our project aimed at developing grammars for text parsing and generation, with a lexical description of the syntactic, semantic, and discourse properties of words. To this end, we have used the grammatical framework of Abstract Categorical Grammars (ACG). This framework allows us to describe reversible grammars, i.e., that are suitable both for parsing and for generation.

► OBJECTIFS DU PROJET

L'informatique linguistique s'intéresse à la modélisation de la langue et à la transformation d'expressions vers des formats manipulables par les ordinateurs. L'analyse automatique consiste à passer de la forme de surface à une forme plus abstraite. La tâche réciproque est la synthèse (ou génération). Certaines applications combinent ces deux tâches (traduction, résumé...). Dans le calcul du lien entre forme et sens d'une expression interviennent : le sens des mots (sémantique lexicale), la combinaison de ces mots (analyse grammaticale), et, pour des textes plutôt que des phrases isolées, la cohérence du discours, liée à l'interprétation en contexte et à la structure rhétorique.

L'objectif de notre projet concernait la modélisation de phrases et de textes dans un paradigme compositionnel, tenant compte des aspects de dynamique et de structure, en analyse et en synthèse, et le développement de grammaires pour l'analyse et la génération de texte.

► MÉTHODOLOGIE ET RÉSULTATS

La contrainte de compositionnalité est importante pour la mise en œuvre de traitements automatiques, mais rend la modélisation de certains phénomènes difficile. Notre travail s'est réparti suivant quatre axes.

Propriétés formelles des systèmes

- Ajout d'opérateurs commutatifs pour les langues à ordre de mots relativement libres et identification de classes de langages polynômiales
- Étude de la relation entre ACG, logique linéaire du premier ordre et programmation logique (linéaire)
- Étude de systèmes de types (polymorphes, dépendants) pour la sémantique lexicale

Intégration de théories linguistiques et modélisation

- Interaction entre sémantique des événements et quantification, modélisation du pluriel
- Modularité de la modélisation sémantique (dynamisme, intensionnalité) inspirée de la théorie des langages de programmation

Harpagon	Viens confesser l'action la plus noire, l'attentat le plus horrible, qui jamais ait été commis. (...)
Valère	Monsieur, puisqu'on vous a découvert tout, je ne veux point chercher de détours, et vous nier la chose. (...)
Valère	De grâce, ne vous mettez point en colère. Quand vous m'aurez ouï, vous verrez que le mal n'est pas si grand que vous le faites.
Harpagon	Le mal n'est pas si grand que je le fais ! Quoi mon sang, mes entrailles, pendard ?
Valère	Votre sang, Monsieur, n'est pas tombé dans de mauvaises mains. (...)
Valère	(...) à persuader votre fille de me donner sa foi, et recevoir la mienne
Harpagon	(...) Que nous brouilles-tu ici de ma fille ?

- Modélisation du discours et des verbes permettant un discours rapporté, et de leur interaction avec les connecteurs de discours
- Reformulation et unification dans les ACG de théories du discours prévues pour la génération et permettant de construire des structures discursives complexes (voir l'illustration)

Expérimentation et évaluation

- Corrélations entre verbes et relations rhétoriques et développement de la base de données *Leclsie*
- Apprentissage supervisé et non supervisé pour l'identification automatique de relations discursives

Développement

- Développement et diffusion d'un logiciel de développement et d'utilisation de grammaires ACG

► CONCLUSION ET PERSPECTIVES

Polymnie a permis des avancées significatives (nouvelle interface réversible syntaxe-discours ; développement et diffusion de logiciel et de base de données ; modèles compositionnels pour la coercion, les pluriels, la quantification ; expérimentations).

Le projet a permis d'envisager de nouvelles questions scientifiques et technologiques : optimisation, algorithmique pour des fragments d'ordre supérieur, hybridation des approches symboliques (notamment pour les grammaires discursives).