

Un outil pour calculer des arbres de dépendance à partir d'arbres de dérivation

Lionel Clément

TALANA, UFRL case 7003 - Université Paris 7
2, pl. Jussieu 75005 Paris FRANCE - Lionel.Clement@linguist.jussieu.fr

Résumé

Nous présentons un outil permettant de calculer un arbre de dépendance sémantique à partir d'un arbre de dérivation TAG.

Cette opération est rendue possible grâce à un algorithme de filtrage des arbres de dérivation et à une étude des propriétés sémantiques liées aux opérations de substitution et adjonction sur une grammaire lexicalisée.

Introduction

Les principes de cooccurrence prédicat-argument (PCPA) et de minimalité sémantique associés au principe de lexicalisation nous permettent d'interpréter l'arbre de dérivation en représentation sémantique de la phrase comme le font (Rambow & Joshi, 1992), (Vijay-Shanker *et al.*, 1995), (Candito & Kahane, 1998) et (Candito, 1999).

PCPA indique que tout arbre élémentaire ancré par un prédicat comporte au moins un nœud pour chacun des arguments ((Kroch & Joshi, 1985), (Abeillé, 1991)). Le principe de lexicalisation pose que tout arbre élémentaire comporte au moins un nœud feuille lexical et le principe de minimalité sémantique que tout arbre élémentaire corresponde à une unité sémantique non vide (Abeillé, 1991), (Candito, 1999).

Ainsi l'arbre de dérivation est une représentation possible de dépendances sémantiques dans la mesure où les adjonctions et substitutions constituent des opérations entre gouverneurs sémantiques et dépendants sémantiques.

Mais, même en prenant la définition de (Schabes & Shieber, 1994) des arbres de dérivation, cette représentation syntaxique ne permet pas toujours de calculer les dépendances immédiatement.

Les cas suivants semblent poser particulièrement problème :

- L'adjonction d'arbres élémentaires non modificateurs comme les auxiliaires verbaux. (fig.1)
- L'inversion de l'ordre de dépendance entre gouverneur sémantique et dépendant sémantique lors d'adjonction. C'est par exemple le cas pour une infinitive ou une complétive dominant la principale dans l'arbre de dérivation (fig.2).

Et dans le cas de l'analyse selon (Vijay-Shanker, 1987)

- Les ambiguïtés artificielles dues aux ordres multiples d'adjonctions dans la dérivation. (fig.3).

Nous avons donc développé, dans le cadre du projet FTAG (Abeillé *et al.*, 1999), un outil qui permet de calculer un arbre de dépendance sémantique de la suite analysée en tirant-parti, d'une part des informations lexicales de la grammaire lexicalisée et, d'autre part d'un ensemble de règles générales.

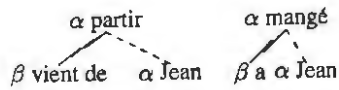


FIG. 1 – Arbres de dérivation de "Jean vient de partir", "Jean a mangé"

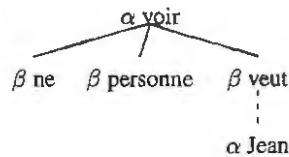


FIG. 2 – Arbres de dérivation selon (Vijay-Shanker, 1987) de "Jean ne veut voir personne"

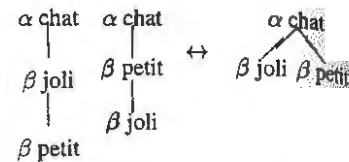


FIG. 3 – Arbres de dérivation selon (Vijay-Shanker, 1987) puis (Schabes & Shieber, 1994) de "joli petit chat"

1. Principe général

Le logiciel filtre les arbres de dérivation en fonction d'un schéma général (par exemple celui de l'adjonction d'un auxiliaire verbal ou d'une infinitive), puis applique de façon incrémentale un ensemble de règles permettant de dresser une représentation sémantique correspondante. Les fonctions effectives données par la grammaire LTAG sont directement attribuées aux arguments sémantiques.

Nous avons décrit trois règles générales qui s'appliquent pour un ensemble de familles donné :

- Règle faisant remonter la principale d'une complétive ou d'une infinitive.
- Règles aplatissant les modificateurs.
- Règle éliminant l'adjonction d'un auxiliaire de temps.

Nous expliciterons infra les deux premières.

2. Règles de calcul d'un arbre de dépendance sémantique à partir d'un arbre de dérivation TAG

2.1. Règle faisant remonter la principale d'une complétive ou d'une infinitive

A l'exception des complétives sujet, les phrases enchâssées sont décrites comme des adjonctions sur un nœud phrastique pour les complétives et infinitives et sur un nœud nominal pour les relatives.

Dans le premier cas, il en résulte naturellement que l'arbre élémentaire correspondant à la phrase matrice est dominé par l'arbre élémentaire correspondant à la phrase enchâssée dans l'arbre de dérivation comme montré fig.4

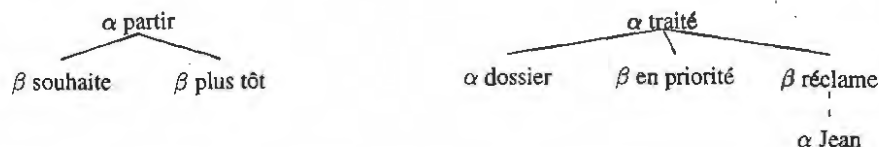


FIG. 4 – Arbres de dérivation de "Jean souhaite partir plus tôt" et "Jean réclame que son dossier soit traité en priorité"

Dans une représentation sémantique, nous voudrions voir cet ordre - s'il correspond à un ordre de dépendance sémantique - respecté pour les relatives mais pas pour les complétives et infinitives.

Nous présentons fig.5 la règle générale qui permet de traduire l'adjonction sur un nœud phrase d'un arbre élémentaire correspondant à une complétive ou infinitive. Cette règle ne s'appliquera pas aux relatives car l'arbre de dérivation ne sera pas filtré.

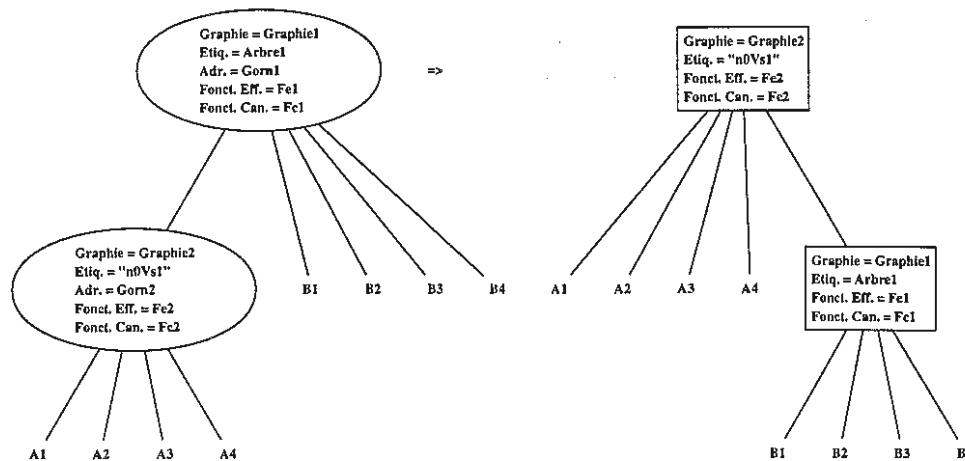


FIG. 5 - Règle faisant remonter la principale d'une complétive ou d'une infinitive. Les termes A_1, A_2, \dots, A_k et B_1, B_2, \dots, B_k correspondent à des variables libres pouvant être instanciées par un nœud.

Le résultat de l'application de la règle est simplement une réorganisation des nœuds comme le montre la fig.5.

2.2. Règles aplatissant les modifieurs

L'analyse de (Schabes & Shieber, 1994) permet de construire plus immédiatement une représentation sémantique pour les modifieurs multiples. En effet, plusieurs adjonctions peuvent avoir lieu sur le même nœud en les ordonnant, ce qui produit un arbre de dérivation "plat" comme montré fig.6.

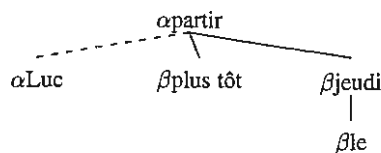


FIG. 6 - Arbre de dérivation selon (Schabes & Shieber, 1994) de "Luc part plus tôt le jeudi."

Dans le cas d'une analyse selon (Vijay-Shanker, 1987), il est possible "d'aplatir" les modifieurs de telle manière qu'il puissent être dépendants sémantiques au même niveau.

L'arbre de dérivation correspondant à une adjonction multiple contient nécessairement une adjonction sur la racine d'un arbre auxiliaire. Cette condition étant par ailleurs suffisante, elle permet de décrire la structure filtrante.

Nous présentons fig.7 la règle générale qui permet d'aplatir les modifieurs. Cette règle sera appliquée autant de fois que des modifieurs artificiellement enchâssés apparaîtront.

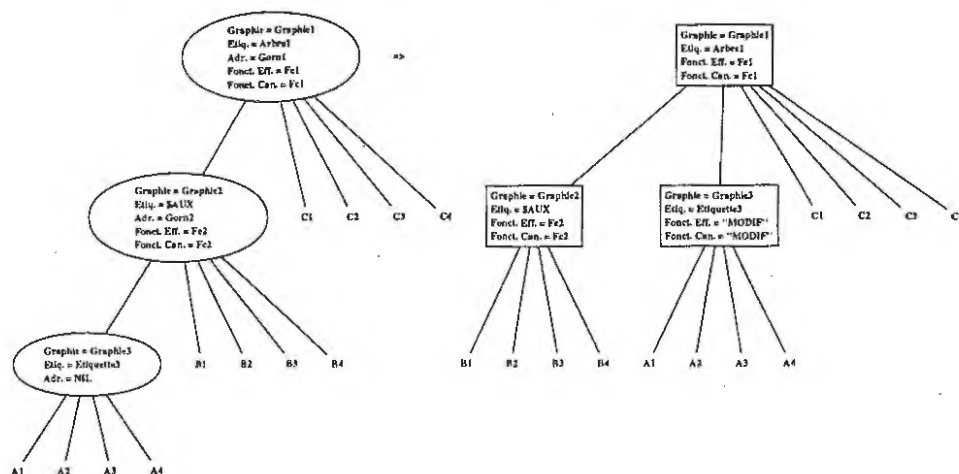


FIG. 7 – Règle aplatissant les modifieurs

Conclusion

Cet algorithme a été implémenté et fournit pour chaque analyse effectuée un arbre de dépendance sémantique. C'est donc une interface utile pour lier un niveau syntaxique à un niveau conceptuel.

Ce travail a été fait dans la perspective de l'analyse en TAG. Nous pouvons le relier aux travaux de (Danlos, 1998) où l'arbre de dérivation est calculé à partir d'un tel niveau conceptuel.

Références

- ABEILLÉ A. (1991). *Une grammaire lexicalisée d'arbres adjoints pour le français: application à l'analyse automatique*. PhD thesis, University Paris 7.
- ABEILLÉ A., CANDITO M.-H. & KINYON A. (1999). Flag: current status and parsing scheme. In *VEXTAL'99*, Venise.
- CANDITO M.-H. (1999). *Représentation hiérarchique de grammaires lexicalisées: application au français et à l'italien*. PhD thesis, University Paris 7.
- CANDITO M.-H. & KAHANE S. (1998). Can the tag derivation tree represent a semantic graph? an answer in the light of meaning-text theory. In *TAG+4*.
- DANLOS L. (1998). G-tag: un formalisme lexicalisé pour la génération de textes inspiré de tag. In *Traitement Automatique des Langues (TAL)*, volume 39.
- KROCH A. S. & JOSHI A. K. (1985). *The linguistic relevance of tree adjoining grammars*. Technical report MS-CIS-85-16, Department of Computer and Information Science, University of Pennsylvania.
- RAMBOW O. & JOSHI A. (1992). A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena. In *International Workshop on The Meaning-Text Theory*, Darmstadt. Arbeitspapiere der GMD 671. To appear in *Current Issues in Meaning-Text Theory*, Leo Wanner, editor.
- SCHABES Y. & SHIEBER S. M. (1994). An alternative conception of tree-adjoining derivation. In *Computational Linguistics*.
- VIJAY-SHANKER K. (1987). *A study of Tree Adjoining Grammars*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA.
- VIJAY-SHANKER K., WEIR D. & RAMBOW O. (1995). Parsing d-tree grammars. In *International Workshop on Parsing Technologies*.