

Intégration d'une dimension sémantique dans les grammaires d'arbres adjoints

Claire Gardent

CNRS/LORIA
claire.gardent@loria.fr

Résumé

Dans cet article, nous considérons un formalisme linguistique pour lequel l'intégration d'information sémantique dans une grammaire à large couverture n'a pas encore été réalisée à savoir, les grammaires d'arbres adjoints (Tree Adjoining Grammar ou TAG). Nous proposons une méthode permettant cette intégration et décrivons sa mise en oeuvre dans une grammaire noyau pour le français. Nous montrons en particulier que le formalisme de spécification utilisé, XMG, (Duchier *et al.*, 2004) permet une factorisation importante des données sémantiques facilitant ainsi le développement, la maintenance et le débogage de la grammaire.

Mots-clés : grammaire d'arbres adjoints, calcul sémantique.

Abstract

In this paper, we consider a linguistic formalism for which the integration of semantic information within a large scale grammar has not yet been realised, namely, Tree Adjoining Grammar (TAG). We propose a method for integrating this information and describe its implementation within a core TAG for French called FRAG. Furthermore, we show that the formalism used to specify the grammar, XMG, (Duchier *et al.*, 2004) allows for a strong factorisation of the semantic data, thus giving better support for the development, maintenance and debugging of the grammar.

Keywords: Tree Adjoining Grammar, Semantic Construction.

1. Introduction

L'objectif central et le plus ambitieux de la linguistique informatique a toujours été la réalisation de systèmes « intelligents » c'est-à-dire de systèmes capables de raisonner sur le sens d'un texte ou d'un énoncé. Or de tels systèmes présupposent une mise en correspondance précise entre énoncés (oraux ou écrits) et représentation sémantique. La linguistique contemporaine et en particulier, la linguistique informatique a mis en avant des formalismes grammaticaux qui permettent d'exprimer cette mise en correspondance de façon déclarative et transparente. Ceci est vrai en particulier, des grammaires syntagmatiques guidées par les têtes (Head Driven Phrase Structure Grammar ou HPSG), des grammaires catégorielles et des grammaires lexicales fonctionnelles (Lexical functional grammar ou LFG). Pour chacun de ces formalismes linguistiques, l'interface syntaxe/sémantique a été définie puis mise en oeuvre dans des grammaires à large couverture. Des algorithmes d'analyse syntaxique et de génération ont en outre été optimisés et équipés d'heuristiques (symboliques ou statistiques) qui permettent d'une part, de traiter du texte réel en temps réel et d'autre part, de déterminer parmi l'ensemble des analyses ou des réalisations possibles celle qui est la plus plausible.

Dans cet article, nous considérons un formalisme pour lequel l'intégration d'information sémantique dans une grammaire à large couverture n'a pas encore été réalisée à savoir, les grammaires d'arbres adjoints (Tree Adjoining Grammar ou TAG). Nous proposons une méthode permettant cette intégration et décrivons sa mise en oeuvre dans une grammaire d'arbres adjoints noyau du français SEMFRAG. Nous montrons en particulier que le formalisme linguistique utilisé (XMG, (Duchier *et al.*, 2004)) pour définir SEMFRAG permet une factorisation importante des données sémantiques facilitant ainsi le développement, la maintenance et le débogage de la grammaire.

2. Les grammaires d'arbres adjoints : syntaxe et sémantique

Syntaxe. Il existe plusieurs types de grammaires d'arbres adjoints. Nous considérons ici les grammaires d'arbres adjoints lexicalisées basées sur l'unification ou FLTAG (Vijay-Shanker et Joshi, 1988). Une FLTAG comprend un ensemble d'arbres élémentaires et deux opérations permettant de combiner ces arbres entre eux, l'opération de substitution et l'opération d'adjonction. Les arbres résultant d'une de ces opérations sont appelés *arbres dérivés*.

Les arbres élémentaires sont lexicalisés, c'est-à-dire qu'ils sont explicitement associés à un lemme ou une forme fléchiée. Un arbre élémentaire est soit initial, soit auxiliaire. Un *arbre initial* est un arbre dont les noeuds feuilles peuvent être des noeuds dit de substitution (marqués par \downarrow). Un *arbre auxiliaire* est un arbre dont l'un des noeuds feuilles est un noeud « pied » (marqué par \star) étiqueté par la même catégorie que le noeud racine. Les noeuds des arbres qui ne sont pas des noeuds de substitution sont étiquetés par deux structures de traits appelées TOP et BOTTOM. Les noeuds de substitution portent uniquement une structure TOP.

L'opération de substitution permet d'insérer un arbre dérivé d'un arbre initial τ_δ dans un arbre contenant un noeud de substitution τ_α : le noeud racine de τ_δ est alors identifié avec un noeud de substitution dans τ_α et les traits TOP sont unifiés ($Top_{\tau_\alpha} = Top_{\tau_\delta}$). L'opération d'adjonction permet d'insérer un arbre auxiliaire τ_β dans un arbre quelconque τ_α à un noeud n : les traits TOP_n et $BOTTOM_n$ du noeud n où se fait l'adjonction sont alors unifiés avec les traits TOP du noeud racine de l'arbre auxiliaire et les traits BOTTOM de son noeud pied respectivement ($Top_n = Top_{Root_{\tau_\beta}}$ et $Bottom_n = Bottom_{Foot_{\tau_\beta}}$).

Sémantique. Plusieurs méthodes ont été proposées pour intégrer une dimension sémantique dans les grammaires d'arbres adjoints : utilisation de TAG synchronisées (Shabes et Shieber, 1990) ; calcul sémantique à partir des arbres de dérivation (Kallmeyer, 2002) ; et calcul sémantique à partir des arbres dérivés (Frank et van Genabith, 2001 ; Gardent et Kallmeyer, 2003 ; Gardent et Parmentier, 2005). L'approche présentée ici implémente la proposition faite dans (Gardent et Kallmeyer, 2003). Dans cette approche, le lien entre syntaxe et sémantique se fait de la façon illustrée par la figure 1. Chaque arbre élémentaire est associé à une **représentation sémantique** où les arguments manquants sont des variables d'unification (les **indices sémantiques**). Ces variables apparaissent en outre sur certains noeuds de l'arbre et sont instanciées par le biais des substitutions et des adjonctions¹. Dans ce qui suit, nous désignons cette méthode de calcul sémantique en parlant de **calcul sémantique par unification** et nous appelons les indices sémantiques apparaissant dans les arbres syntaxiques **indices d'arbres**.

L'exemple donné en figure 2 illustre le traitement des principaux types de phénomènes sémantiques dont un module de calcul sémantique doit pouvoir traiter à savoir : la quantification, le contrôle, la modification et les relations prédicat/argument. (Gardent et Kallmeyer, 2003) mon-

¹ Pour plus de détails sur le calcul sémantique utilisé, cf. (Gardent et Kallmeyer, 2003).

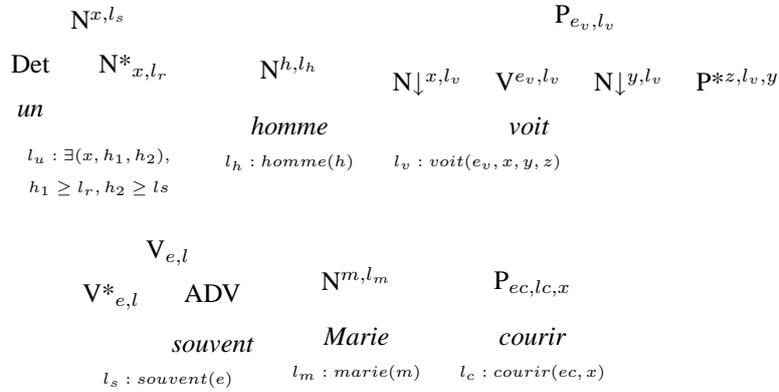


Figure 1. Une grammaire exemple

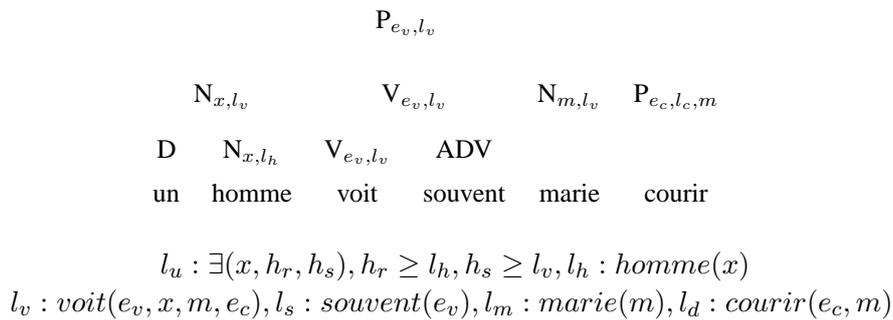


Figure 2. « Un homme voit souvent Marie courir »

tre par ailleurs, comment le calcul sémantique par unification permet de traiter de phénomènes sémantiques problématiques pour les approches TAG basées sur l'arbre de dérivation.

Les indices d'arbres sont des variables sur les individus (x, y), sur les événements (e_v, e_c) et sur les étiquettes (l_r, l_c). Les variables sur les étiquettes sont utilisées en particulier pour traiter des opérateurs à portée tels que les quantifieurs. Comme dans la grammaire HPSG (Copestake, 1999), les représentations sémantiques utilisées sont des formules de la MRS (Minimal Recursion Semantics) c'est-à-dire des formules logiques « aplaties » où les variables lambda sont remplacées par des variables d'unifications (nos « indices sémantiques »).

3. Intégration de la dimension sémantique dans une grammaire noyau du français

Afin d'intégrer une dimension sémantique du type décrit dans la section précédente dans une grammaire d'arbres adjoints, il est nécessaire :

1. d'associer chaque arbre élémentaire avec une représentation sémantique. Par exemple, dans la grammaire donnée en figure 1, l'arbre pour *Marie* est associé à la représentation sémantique $l_m : marie(m)$;
2. de décorer les arbres avec des indices sémantiques. Par exemple, le noeud racine de l'arbre pour *Marie* est associé aux indices sémantiques m et L ;
3. d'identifier les indices sémantiques décorant les noeuds d'un arbre (les *indices d'arbres*) avec les indices correspondants dans la représentation sémantique associée à cet arbre (les *indices sémantiques*). Dans la figure l'identification est indiquée par l'utilisation d'une variable identique. Dans une grammaire comme SEMFRAG où la dimension syntaxique (l'arbre) est décrite indépendamment de la dimension sémantique (la formule MRS), cette identification doit être explicitée.

L'intégration de cette information pour l'ensemble des arbres (plusieurs milliers) constituant une grammaire d'arbres adjoints est une tâche complexe. Dans ce qui suit, nous montrons comment une factorisation de l'information la facilite. Nous commençons par décrire la grammaire syntaxique utilisée (FRAG). Nous présentons ensuite le formalisme mis en oeuvre pour la spécifier à savoir XMG. Nous montrons enfin comment XMG peut être utilisé pour représenter et surtout pour factoriser la dimension sémantique et l'interface syntaxe/sémantique d'une grammaire TAG.

3.1. FRAG – une grammaire d'arbres adjoints pour le français

FRAG (Crabbé et Duchier, 2004) est une grammaire d'arbres adjoints développée pour un fragment noyau du français dans le cadre de XMG. Générée automatiquement par le compilateur XMG à partir d'une méta-grammaire, FRAG couvre les constructions verbales, nominales et adjectivales de base et comporte environ 4 200 arbres. Évaluée sur la suite de tests TSNLP (Test Suite for Natural Language Processing), elle génère 76% des items grammaticaux, rejette 83% des items agrammaticaux et exhibe un taux d'ambiguïté moyen de 1,64 analyses par phrases.

3.2. XMG – un système de description et de compilation pour les formalismes linguistiques

Pour factoriser l'information syntaxique et/ou sémantique contenue dans SEMFRAG, nous utilisons le système XMG. Ce système comprend trois grandes composantes : (i) des langages

de description linguistique (pour la syntaxe et pour la sémantique) permettant la spécification d'une méta-grammaire, (ii) un langage de contrôle permettant de combiner les descriptions linguistiques créées et (iii) des modules de compilation permettant de produire une grammaire à partir d'une méta-grammaire (*i.e.*, un ensemble de descriptions linguistiques).

Langages de description. Le langage de description syntaxique est un langage de descriptions d'arbres dont les noeuds peuvent être décorés par des structures de traits (cf. (Crabbé et Duchier, 2004) pour une description plus détaillée de ce langage). Une originalité importante de ce langage de description est l'utilisation des *couleurs* pour contrôler la combinatoire des descriptions d'arbres. Dans une description d'arbre, chaque variable de noeud peut être soit rouge, soit blanche, soit noire. Lors de la génération des arbres satisfaisant une description, la coloration des variables de noeuds restreint l'ensemble des solutions possibles selon le schéma donné ci-dessus. En pratique, l'utilisation des couleurs permet d'éviter des identifications explicites entre noeuds qui, dans une méta-grammaire de taille importante deviendraient vite fastidieuses. Nous verrons en particulier (section 3.3) que les couleurs permettent un encodage fortement factorisé de la projection des indices dans les arbres élémentaires TAG.

	● _B	● _R	○ _W
● _B	⊥	⊥	● _B
● _R	⊥	⊥	⊥
○ _W	● _B	⊥	○ _W

XMG permet également de décrire la dimension sémantique d'une grammaire et plus particulièrement, les formules de la MRS. Le langage de représentation sémantique est le suivant (avec l, p, e_i représentant des variables d'étiquettes, de prédicats et d'objets (individus ou évènements) respectivement ; et \geq représentant la relation de portée entre étiquettes) :

$$Description ::= l : p(e_1, \dots, e_n) \mid \neg l : p(e_1, \dots, e_n) \mid e_i \geq e_j \quad (1)$$

Langage de contrôle. Le langage de contrôle XMG permettant de décrire des *relations entre fragments* (syntaxiques et/ou sémantiques). Ce langage permet en particulier le nommage (2), l'héritage, la conjonction et la disjonction (3) de fragments. Il permet également un traitement sophistiqué des interactions entre héritage et nommage. En particulier, les variables **exportées** par un fragment sont des variables globales à tous les enfants de ce fragment².

$$Clause ::= Name \rightarrow Goal \quad (2)$$

$$Goal ::= Description \mid Goal \wedge Goal \mid Goal \vee Goal \mid Goal \wedge Goal \quad (3)$$

Compilation. La phase de compilation permet de produire une grammaire TAG à partir d'une méta-grammaire. Elle se fait en plusieurs étapes. Un premier module produit pour chaque fragment défini dans la méta-grammaire sa version extensionnelle c'est-à-dire son contenu après traitement des relations entre fragments (héritage, conjonction, disjonction, etc.). Un second module permet de produire les arbres satisfaisant une description d'arbre (il est utilisé pour produire les arbres TAG à partir des descriptions d'arbres définies par la méta-grammaire). Enfin, un troisième module filtre des structures linguistiquement incorrectes à partir d'un certain nombre de critères linguistiques (*e.g.*, un arbre élémentaire ne peut pas contenir plus d'un site d'extraction).

3.3. Intégration factorisée d'une dimension sémantique dans une méta-grammaire TAG

Pour construire SEMFRAG nous intégrons une dimension sémantique dans la méta-grammaire TAG développée au moyen d'XMG pour générer FRAG (cf. section 3.1).

² Le langage XMG permet également de bloquer l'exportation d'une variable spécifique si besoin est.

3.3.1. Association arbre élémentaire/représentation sémantique

Chaque arbre élémentaire décrit par la méta-grammaire est associé à une **représentation sémantique**. En outre, les **indices sémantiques** c'est-à-dire les variables représentant les arguments sémantiques d'un prédicat, sont chacun associés à un nom à portée globale qui sera nécessaire pour pouvoir correctement identifier indices d'arbres et indices sémantiques (cf. section 3.3.3). Par exemple, les arbres élémentaires associés à un verbe intransitif seront associés à la représentation sémantique typique d'une relation unaire impliquant un évènement à savoir : $l : p(e)$; $l : theta_1(e, x)$ avec p et $theta_1$ des variables dont la valeur sera donnée par le lexique (par exemple, $p = \text{dormir}$ et $theta_1 = \text{agent}$ dans le cas du verbe intransitif *dormir*). Le premier argument sémantique du prédicat (e.g., x dans $l : p(e)$; $l : theta_1(e, x)$) est de plus nommé *arg1*.

Pour l'ensemble de la grammaire, 8 représentations sémantiques suffisent :³

- | | |
|---|--|
| (1) a. $l : p(e)$; $l : theta_1(e, x)$ | (Verbe intransitif) |
| b. $l : p(e)$; $l : theta_1(e, x)$; $l : theta_2(e, y)$ | (Verbe transitif) |
| c. $l : p(e)$; $l : theta_1(e, x)$; $l : theta_2(e, y)$; $l : theta_3(e, z)$ | (Verbe ditransitif) |
| d. $l : p(v)$ | (Adjectif unaire, nom commun, auxiliaire, adverbe) |
| e. $l : p(x, y)$ | (adjectif binaire, préposition unaire, nom relationnel non déverbal) |
| f. $l : p(x, y, z)$ | (adjectif ternaire) |
| g. $l : p(x, h_2, h_3)$; $h_2 \geq l_n$; $h_3 \geq l_v$ | (déterminant quantifiant) |
| h. $l : p_1(x, l_2, h_3)$; $l_2 : p_2(x)$; $h_3 \geq l_v$ | (SN quantifiant) |

L'association entre description syntaxique et description sémantique se fait de façon factorisée soit au niveau des familles d'arbres TAG (pour les verbes), soit dans les couches supérieures de la hiérarchie d'héritage (pour les autres catégories du discours).

Factorisation par le biais des familles TAG. La description $n0Vn1$ décrit la famille TAG du même nom c'est-à-dire l'ensemble des arbres représentant les configurations syntaxiques possibles pour un verbe transitif. Si ces configurations sont multiples (environ 150)⁴, la sémantique d'un verbe transitif est toujours la même : un verbe transitif dénote une relation entre un évènement et deux participants⁵. Cette constante sémantique est capturée dans SEMFRAG par l'association de la représentation sémantique des verbes binaires (1b) nommée ici *binaryRel* avec la description $n0Vn1$:

$$\begin{aligned}
 n0Vn1 &\rightarrow binaryRel \wedge \\
 &(\quad dian0Vn1Active \vee dian0Vn1Passive \vee dian0Vn1dePassive \vee dian0Vn1ShortPassive \\
 &\quad \vee dian0Vn1ImpersonalPassive \vee dian0Vn1middle \vee dian0Vn1Reflexive \quad)
 \end{aligned}
 \tag{4}$$

Factorisation par héritage. Lorsque plusieurs descriptions syntaxiques héritent d'une description commune D et ont la même sémantique S , S est associée à la description commune la plus haut placée dans la hiérarchie. Par exemple dans SEMFRAG, les prépositions à un argument sont décrites par une dizaine de descriptions syntaxiques distinctes (selon la nature et la position des arguments syntaxiques possibles) mais toutes ces descriptions ont en commun une description sous spécifiée reflétant l'existence pour chacune des prépositions possibles d'un noeud de substitution (correspondant à l'argument attendu) et d'un noeud pied (correspondant à

³ Les conventions suivantes sont utilisées : v est une variable quelconque, x, y, z sont des variables portant sur des individus, e_i des variables événementielles, h_i des « poignées » c-à-d, des variables permettant de sous-spécifier la portée des quantificateurs et l_i des variables portant sur les étiquettes étiquettant les formules des sémantiques plates.

⁴ Les 150 possibilités correspondent à la combinaison des formes verbales (actif, passif, impersonnel, réflexif, moyen) et des réalisations possibles pour les arguments (canonique, questionné, pronom relatif, clitique, etc.).

⁵ Comme nous le verrons par la suite, dans certains cas comme le passif sans agent, l'un des arguments est existentiellement quantifié.

l'item modifié par le syntagme prépositionnel introduit par la préposition). Comme cette classe est héritée par l'ensemble des descriptions terminales pour les prépositions unaires, nous associons la représentation sémantique des prépositions avec celle-ci plutôt qu'avec les descriptions terminales. Ainsi pour ces prépositions, l'association sémantique/syntaxe est faite par une seule déclaration.

La figure (3) donne une idée de la factorisation permise par la méta-grammaire : la première colonne (Sémantique) réfère aux représentations sémantiques données en (1), la deuxième au nombre d'appels de ces représentations dans la méta-grammaire, la troisième donne une liste (non exhaustive) des constructions syntaxiques associées à la représentation sémantique pointée par la première colonne et la dernière colonne donne un exemple illustratif. Cette figure montre une forte factorisation de l'association entre arbres élémentaires et représentations sémantiques puisque 75 appels suffisent à associer une représentation sémantique à chacun des 4 200 arbres présents dans la grammaire.

Sémantique	Nombre d'appels	Constructions	Exemples
1a	6	Verbes intransitifs	<i>Jean dort</i>
1b	22	Verbes transitifs	<i>Jean mange une pomme</i>
1c	21	Verbes ditransitifs	<i>Jean donne un livre à Marie</i>
1d	10	Adjectifs unaires Dét. Non Quantifiants Copule Modaux Verbes à explétifs	<i>Jean est joyeux ; Un joyeux départ</i> <i>Ce chat</i> <i>Jean est directeur/heureux/à Paris</i> <i>Il faut un ingénieur/partir</i> <i>Il neige</i>
1e	5	Noms relationnels non déverbaux Adjectifs binaires	<i>Le chef du département</i> <i>Un enfant fort en maths</i>
1f	3	Adjectifs ternaires	<i>Un enfant supérieur à Luc en maths</i>
1g	7	Dét. Quantifiants	<i>Un, aucun, chaque</i>
1h	1	SN quantifiants	<i>Personne, aucuns</i>

Figure 3. Factorisation des représentations sémantiques

3.3.2. Déclaration des indices d'arbres

Dans le calcul sémantique par unification proposé pour TAG par (Gardent et Kallmeyer, 2003 ; Gardent et Parmentier, 2005), les arbres syntaxiques sont décorés d'indices sémantiques coréférencés avec les indices apparaissant dans les représentations sémantiques associées à ces arbres. Ces indices spécifient la relation entre constituants syntaxiques d'une part, et foncteurs et arguments sémantiques d'autre part. La déclaration des indices d'arbres comprend d'une part, la déclaration des **noeuds ressources** (ceux qui dominent un constituant pouvant fournir un argument sémantique à un foncteur) et d'autre part, la déclaration des **noeuds besoin** (ceux qui indiquent une exigence pour un argument sémantique).

Noeuds besoin. Intuitivement, les noeuds besoin reflètent un besoin pour un argument sémantique. Par exemple, dans l'arbre pour *voir* de la figure 1, le noeud sujet est un noeud besoin car il est associé au premier argument sémantique (x) de la relation binaire ($voir(e,x,y)$) associée à cet arbre⁶ Dans les grammaires d'arbres adjoints, l'identification des noeuds besoin est simple :

⁶ Comme le montre (Copestake *et al.*, 1999), dans le calcul sémantique basé sur l'unification, les variables d'unification correspondent aux variables lambda. Les variables événementielles sont ici des constantes skolémisées si bien que la relation binaire $voir(e,x,y)$ correspond au lambda terme $\lambda x \lambda y. \exists e. voir(e, x, y)$.

ce sont les noeuds de substitution (qui indiquent une relation prédicat/argument) et les noeuds pieds (qui indiquent une relation modifieur/modifié). Dans la méta-grammaire, chacun de ces noeuds est associé à un indice⁷ et une étiquette sémantique. En outre, indice et étiquette sont chacun associés à un nom à portée globale qui pourra être utilisé pour identifier indices sémantiques et indices d'arbres (cf. section 3.3.3). Autant que possible, la spécification des noeuds besoin se fait dans les niveaux supérieurs de la hiérarchie si bien que 59 spécifications suffisent à la déclaration des noeuds besoin pour l'ensemble de la grammaire.

Noeuds ressources. Les noeuds ressources sont des noeuds syntaxiques qui fournissent l'indice sémantique d'un foncteur sémantique. Ainsi les noeuds racines de la plupart des arbres élémentaires sont des noeuds ressources. Mais du fait de l'adjonction, certains noeuds non terminaux seront eux aussi des noeuds ressources. Comme le montre (Frank et van Genabith, 2001), ces noeuds sont les noeuds apparaissant sur la projection du noeud ancre. Par exemple dans la figure 1, l'arbre pour *voir* a pour noeuds ressources le noeud racine (étiqueté par les indices e_v, l_v) et le noeud V (puisque'il apparaît sur la projection du noeud ancre *voir*).

L'information sémantique étiquettant les noeuds ressources est factorisée dans la description syntaxique *ProjectionSem* qui décrit un arbre local où le noeud $n1$ est le père de $n2$ et où $n1$ et $n2$ sont associés au même indice et à la même étiquette (cf. figure 4). Cette description est spécialisée au niveau des catégories syntaxiques par héritage – ainsi par exemple la description *ProjectionSemV_P* hérite de *ProjectionSem* et décrit un arbre local où le noeud $n1$ est décoré avec la catégorie P et $n2$ avec la catégorie V . Les noms de noeuds $n1$ et $n2$ étant exportés par la description *ProjectionSem*, les noeuds $n1$ et $n2$ apparaissant dans la définition de *ProjectionSemV_P* sont interprétés comme identiques aux noeuds $n1$ et $n2$ introduits par *ProjectionSem* (cf. section 3.2). Le fragment décrit par *ProjectionSemV_P* est ainsi celui donné dans la partie droite de la figure. Les arbres locaux représentant la projection d'indices sé-

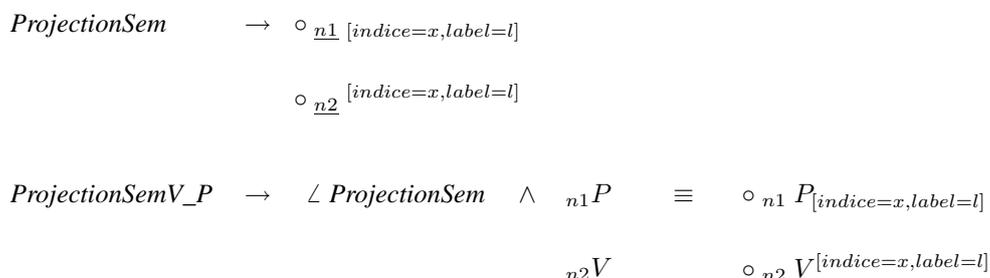


Figure 4. Projection des indices sémantiques

mantiques sont ensuite combinés avec les fragments d'arbres représentant l'épine dorsale des foncteurs syntaxiques. Figure 5 donne l'exemple des constructions supports à tête nominale. Dans ce cas, ce sont les couleurs (cf. section 3.2) qui contrôlent la combinatoire : parce que les noeuds P et V de la description *FormePredicative* sont de couleur noire alors que les noeuds P et V de la description *ProjectionSemV_P* sont de couleur blanche, lors de génération des arbres satisfaisant la description *FormePredicative*, noeuds blancs et noeuds noirs seront identifiés (un arbre contenant des noeuds blancs est mal formé) garantissant ainsi une solution unique, celle donnée dans la partie droite de la figure.

⁷ Les arguments phrastiques infinitivaux sont associés aux deux indices : l'indice introduit par l'infinitive et celui du contrôleur (cf. figure 1).

$$\begin{array}{l}
 \text{FormePredicative} \rightarrow \text{ProjectionSemV_P} \wedge \bullet P \quad \equiv \quad \bullet P_{[indice=x,label=l]} \\
 \quad \quad \quad \quad \quad \quad \quad \quad \bullet V \quad \quad \quad \quad \quad \quad \bullet V_{[indice=x,label=l]} \\
 \quad \quad \quad \quad \quad \quad \quad \quad V \quad \bullet_{xPred} \quad \quad \quad V \quad \bullet_{xPred} \\
 \text{FormePredicativeNominale} \rightarrow \text{ProjectionSemN_V} \wedge \text{FormePredicative} \wedge xPred.cat = N \\
 \quad \quad \quad \equiv \bullet P_{[indice=x,label=l]} \\
 \quad \quad \quad \quad \quad \quad \bullet V_{\substack{[indice=x,label=l] \\ [indice=x2,label=l2]}} \\
 \quad \quad \quad \quad \quad \quad V \quad \quad \quad \bullet N_{xPred}^{[indice=x2,label=l2]}
 \end{array}$$

Figure 5. L'épine dorsale des constructions supports à tête nominale

3.3.3. Identification des indices d'arbres et des indices sémantiques

L'identification des indices d'arbres et des indices sémantiques se fait par le biais de contraintes d'interface qui permettent d'identifier des valeurs appartenant à des dimensions linguistiques différentes (i.e., l'arbre syntaxique et la représentation sémantique). Par exemple, pour la classe *nOVnI* mentionnée plus haut, l'identification des indices d'arbres et des indices sémantiques se fera comme suit :

$$\begin{array}{l}
 nOVnI \rightarrow \text{binaryRel}[arg0=E,arg1=X,arg2=Y] \wedge \\
 \quad (\text{dianOVnIActive}[vbI=E,subjectI=X,objectI=Y] \vee \text{dianOVnIPassive}[vbI=E,subjectI=Y,cagentI=X] \\
 \quad \vee \text{dianOVnIdePassive}[vbI=E,subjectI=Y,deobjectI=X] \vee \text{dianOVnIShortPassive}[vbI=E,subjectI=Y] \\
 \quad \vee \text{dianOVnIImpersonalPassive}[vbI=E,objectI=Y] \vee \text{dianOVnImiddle}[vbI=E,subjectI=Y] \\
 \quad \vee \text{dianOVnIReflexive}[vbI=E,subjectI=X,reflI=X])
 \end{array} \quad (5)$$

En d'autres termes, l'identification des indices d'arbres et des indices sémantiques repose sur l'attribution de noms globaux aux indices sémantiques ainsi qu'aux indices d'arbres couplée avec une identification appropriée des variables identifiées par ces noms. Par exemple, la variable nommée *arg1* dans la définition de la représentation sémantique *binaryRel* (i.e., x dans est $l : p(e); l : \theta_1(e, x); l : \theta_2(e, y)$) est identifié par unification avec l'indice d'arbre correspondant à la variable nommée *subjectI* (i.e., x dans l'arbre pour *voir* de la figure 1) dans l'arbre dénoté par la description *dianOVnIActive*.

4. Conclusion

L'élaboration de systèmes TAL intelligents présuppose une mise en correspondance précise entre énoncés et représentations sémantiques. Nous avons montré comment étendre une grammaire TAG avec une dimension sémantique permettant cette mise en correspondance. Nous avons également montré que cette extension de la grammaire peut se faire avec une factorisation forte facilitant ainsi le développement et la maintenance de la grammaire. Le travail en cours porte sur une extension du lexique et de la grammaire visant le traitement (analyse et génération) de paraphrases ainsi que sur l'évaluation du système existant par le biais de l'analyse et de la génération. La grammaire SEMFRAG a ainsi été couplée avec deux analyseurs (l'analyseur Dyalog de Eric de la Clergerie et l'analyseur LLP2 de Azim Roussanaly) et un réalisateur de surface, le réalisateur GenI (Gardent et Kow, 2005). En analyse, nous comparons l'efficacité des deux méthodes de calcul sémantique décrites par (Gardent et Parmentier, 2005). En génération, nous visons à mesurer et contrôler la sur-génération.

Remerciements

Je remercie Yannick Parmentier pour son aide précieuse concernant l'utilisation d'XMG lors du développement de la grammaire TAG à dimension sémantique SEMFRAG.

Références

- COPESTAKE A. (1999). « The (new) LKB system ». available from <http://www-linguistics.stanford.edu/~aac/doc5-2.pdf>.
- COPESTAKE A., FLICKINGER D., SAG I. et POLLARD C. (1999). « Minimal Recursion Semantics. An Introduction ». Manuscript, Stanford University.
- CRABBÉ B. et DUCHIER D. (2004). « Metagrammar Redux ». In *International Workshop on Constraint Solving and Language Processing - CSLP 2004, Copenhagen*.
- DUCHIER D., LE ROUX J. et PARMENTIER Y. (2004). « The MetaGrammar Compiler : An NLP Application with a Multi-paradigm Architecture ». In *Second International Mozart/Oz Conference - MOZ 2004, Charleroi, Belgique*.
- FRANK A. et VAN GENABITH J. (2001). « GlueTag. Linear Logic based Semantics for LTAG ». In M. Butt et T. H. King (éds.), *Proceedings of the LFG01 Conference*. Hong Kong.
- GARDENT C. et KALLMEYER L. (2003). « Semantic construction in FTAG ». In *Proceedings of the 10th meeting of the European Chapter of the Association for Computational Linguistics*. Budapest, Hungary.
- GARDENT C. et KOW E. (2005). « Generating and selecting grammatical paraphrases ». In *Proceedings of the 10th ENLG*. Aberdeen, Scotland.
- GARDENT C. et PARMENTIER Y. (2005). « Large scale Semantic construction for Tree Adjoining Grammar ». In *Proceedings of Logical Aspects in Computational Linguistics*. Bordeaux, France.
- KALLMEYER L. (2002). « Using an Enriched TAG Derivation Structure as Basis for Semantics ». In *Proceedings of TAG+6 Workshop*. Venice, p. 127 – 136.
- SHABES Y. et SHIEBER S. (1990). « Synchronous Tree Adjoining Grammar ». In *Proceedings of COLING '90*.
- VIJAY-SHANKER K. et JOSHI A. (1988). « Feature structures based tree adjoining grammars ». In *Proceedings of COLING*. Budapest, Hungary, p. 714-719.