



Modélisation de la coordination dans les grammaires d'interaction

Joseph Le Roux, Guy Perrier

► **To cite this version:**

Joseph Le Roux, Guy Perrier. Modélisation de la coordination dans les grammaires d'interaction. Traitement Automatique des Langues, Lavoisier (Hermes Science Publications), 2007. <inria-00185237>

HAL Id: inria-00185237

<https://hal.inria.fr/inria-00185237>

Submitted on 5 Nov 2007

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.

Modélisation de la coordination dans les grammaires d'interaction

Joseph Le Roux* — Guy Perrier**

* *INPL-LORIA*
Campus scientifique – BP 239
54506 Vandoeuvre-lès-Nancy Cedex
joseph.leroux@loria.fr

** *Université Nancy2 - LORIA*
Campus scientifique – BP 239
54506 Vandoeuvre-lès-Nancy Cedex
perrier@loria.fr

RÉSUMÉ. Nous présentons une modélisation de la coordination à l'aide des grammaires d'interaction. La notion de polarité permet de définir l'interface d'une structure syntaxique comme sa capacité à se combiner à d'autres structures syntaxiques en termes de besoins et de ressources. La coordination de deux conjoints est alors vue comme la superposition des interfaces de leurs structures syntaxiques. Ces structures sont représentées à l'aide de la notion de description d'arbres et peuvent ainsi intégrer la sous-spécification pour se combiner de façon particulièrement souple.

ABSTRACT. We present a model of coordination within Interaction Grammars. The notion of polarity lets us define the interface of a syntactic structure as its capacity to combine with other syntactic structures in terms of requirements and resources. The coordination of two conjuncts is then viewed as the superposition of their syntactic structure interfaces. Those structures are represented by means of tree descriptions to integrate underspecification so they combine in a very flexible way.

MOTS-CLÉS: coordination, grammaires catégorielles, grammaires d'interaction, grammaires formelles.

KEYWORDS: coordination, Categorical Grammars, Interaction Grammars, formal grammars.

1. Introduction

Nous proposons de modéliser la coordination dans le formalisme des grammaires d'interaction (GI)(Perrier, 2004). Même si les GI couvrent à la fois la sémantique et la syntaxe des langues, nous nous contenterons ici d'aborder les aspects syntaxiques de la coordination tout en sachant qu'il n'est pas possible d'étudier la syntaxe sans avoir en ligne de mire la sémantique.

La difficulté du problème réside tout d'abord dans le fait que l'on est loin du consensus dans la communauté sur l'analyse du phénomène (Carston et Blake-more, 2003; Godard, 2005). Ensuite, même en choisissant un point de vue linguistique particulier, nous sommes loin d'avoir fait l'essentiel du chemin. Le caractère particulièrement subtil de la coordination donne bien du fil à retordre à la plupart des formalismes, qui sont en général obligés d'introduire des mécanismes plus ou moins *ad hoc* pour la modéliser. En effet, si on choisit une approche syntagmatique, on se heurte au problème de la coordination de non-constituants quand on cherche à représenter l'arbre syntaxique d'une phrase contenant une expression coordonnée. L'approche en termes de dépendances ne permet pas d'exprimer simplement les dépendances entre les mots d'une phrase avec coordination sous forme d'un arbre ; en effet, la coordination vise à superposer les dépendances non résolues des deux conjoints, ce qui sort du cadre de représentation de la syntaxe au moyen d'un arbre de dépendance.

Les formalismes s'adaptent avec plus ou moins de bonheur à ce phénomène complexe, les plus expressifs et les plus souples comme HPSG par exemple (Abeillé, 2003) y parvenant mieux que d'autres plus rigides et plus limités dans leur pouvoir d'expression comme TAG (Sarkar et Joshi, 1996; Seddah et Sagot, 2006).

Les grammaires catégorielles (GC), dans leurs différentes variantes (Morrill, 1994; Moortgat, 1996; Steedman, 2000), réussissent à modéliser de façon élégante des aspects de la coordination qui posent problème à la plupart des autres formalismes ; c'est le cas notamment de la coordination de non-constituants, de séquences de compléments (Dowty, 1988) ou la coordination disparate (Bayer et Johnson, 1995). Leur atout est que le type d'une expression, qui peut être un syntagme complet ou déficient, représente sa capacité à se combiner avec d'autres expressions, son interface syntaxique en quelque sorte. La coordination de deux expressions exige qu'elles aient la même interface, c'est-à-dire le même type, et effectue la superposition de ces interfaces, ce qui fait que l'expression coordonnée résultante a aussi ce type commun. L'inconvénient des GC, dans leur noyau que sont les grammaires de Lambek, est leur trop grande rigidité. Ainsi, ordre et valence des mots ne sont pas dissociés, ce qui fait qu'on ne peut ajouter ou supprimer des éléments qu'en périphérie d'une expression et pas en plein milieu. Cela pose des problèmes, par exemple pour le traitement des modificateurs de phrase à place relativement libre ou pour l'extraction médiane.

Divers enrichissements du formalisme de base ont été proposés pour résoudre ces problèmes (Morrill, 1994; Moortgat, 1996; Steedman, 2000; Lecomte et Retoré, 2001). Certains ont donné lieu à expérimentation sur de gros corpus de référence arborés par apprentissage automatique de grammaires (Hockenmaier, 2006; Moot, 2007).

Notre démarche, si elle accorde aussi une place essentielle à l'expérimentation, diffère en ce sens que nous cherchons à formaliser de la façon la plus complète et la plus fine une grammaire de façon manuelle. Dans cette optique, des questions comme la capacité d'un formalisme à faciliter l'écriture de grammaires à large couverture ou la limitation de la surgénération sont importantes. C'est aussi pourquoi nous avons choisi dans un premier temps d'évaluer la pertinence de la grammaire construite sur une suite de phrases tests positives et négatives plutôt que sur un gros corpus réel. Notre proposition de modélisation de la coordination s'inscrit dans cette démarche.

Cette proposition se fonde sur la notion de polarité, qui est au cœur des GI et qui touche à la sensibilité aux ressources de la langue naturelle. Dans une structure syntaxique, les polarités permettent d'exprimer son interface comme son aptitude à se combiner avec d'autres structures en termes de besoins (polarités négatives) et ressources (polarités positives). Cette notion de polarité est déjà au cœur des GC même si elle n'y apparaît pas de façon explicite. Les GI introduisent d'une part un raffinement de la notion de polarité en la descendant du niveau des syntagmes au niveau des traits : une même expression peut alors être porteuse à la fois de polarités positives et négatives. Par ailleurs, elles introduisent de la souplesse dans la représentation des structures syntaxiques et dans la façon de les composer à l'aide du concept de *description d'arbres* (Rogers et Vijay-Shanker, 1992).

Les structures syntaxiques de base du formalisme sont les descriptions d'arbres syntaxiques (DAS). Une DAS est une façon compacte de représenter une famille d'arbres par leur spécification commune ou, autrement dit, de représenter un arbre sous-spécifié. Une DAS est étiquetée de traits polarisés et son interface est formée de sa partie active, celle qui est porteuse de polarités positives ou négatives qui vont servir à opérer la combinaison avec d'autres DAS.

Le principe qui constitue la ligne directrice de l'article est que deux expressions sont coordonnables si leurs structures syntaxiques, représentées sous forme de DAS, possèdent la même interface et leur coordination consiste à superposer ces interfaces.

Dans la section 2, nous circonscrivons les phénomènes syntaxiques relatifs à la coordination que nous prendrons en compte. Dans la section 3, nous donnerons une brève présentation informelle des GI. Dans la section 4, nous aborderons le cœur de l'article avec nos propositions de modélisation de la coordination. Dans la section 5, nous comparons nos propositions avec celles des GC et de HPSG. Enfin, dans la section 6, nous décrivons l'implantation de la grammaire à l'aide de l'outil XMG.

2. Phénomènes étudiés

Dans cette section, nous distinguons les différents sous-phénomènes auxquels nous nous intéresserons dans la suite de l'article. Le premier – et le plus consensuel – est la coordination de constituants, présentée par les phrases de l'exemple (1) dans lesquelles deux propositions indépendantes, deux syntagmes nominaux et deux syntagmes adjec-

tivaux sont coordonnés. Des constituants similaires doivent toujours pouvoir se coordonner. La similitude s'entend comme la compatibilité de nature et de fonction¹.

- (1) a) Il pleut et il neige.
 b) Jean et son frère discutent.
 c) une chaussure rouge et vert

L'exemple (1c) pourrait également s'interpréter comme une coordination lexicale où les deux adjectifs coordonnés se comportent comme un mot unique (Abeillé, 2006). Nous ne considérons pas ce type de construction dans cet article.

Ensuite viennent les coordinations de non-constituants, qui posent problème à la majorité des formalismes puisqu'il s'agit de donner un statut *officiel* à des fragments de constituants. Selon les approches linguistiques, ces coordinations peuvent être considérées comme des constructions elliptiques où une partie d'un des conjoints a été effacée et qu'il faut retrouver ou, au contraire, comme la coordination de deux segments qui partagent une partie de la phrase, qui est donc factorisée, et que la conjonction distribue. Par exemple, dans la phrase *Jean boit du vin et mange* (2a), soit on considère que l'on coordonne deux propositions mais que le sujet de la seconde a été effacé ou alors que l'on coordonne deux groupes verbaux² ayant le même sujet, ici *Jean*.

- (2) a) [Jean boit du vin et ~~Jean~~ mange] vs. Jean [boit du vin et mange].
 b) [Jean aime ~~la compétition~~ mais Marie déteste la compétition] vs. [Jean aime mais Marie déteste] la compétition.

De plus, on va distinguer les coordinations de non-constituants, avec *montée de næud* où la partie factorisée est en périphérie de syntagme (comme dans les exemples précédents), de celles où elle se situe à l'intérieur (phénomène dit de trou verbal ou *gapping*), comme dans (3) où le verbe *être* manque à la seconde proposition. Ce type de coordination est réservé aux propositions.

- (3) Jean est informaticien et Pierre mécanicien.

Puis, on remarque que les séquences de compléments dépendant d'une même tête (*argument clusters*) peuvent se coordonner, comme par exemple en (4a). Ici, les conjoints sont des groupes formés d'un objet direct et d'un complément d'attribution. Ces séquences ne forment pas de constituants. L'exemple (4b) montre que ces constructions ne sont pas réservées aux arguments verbaux.

- (4) a) Jean donne des fleurs à Marie et des bonbons à Pierre.
 b) La destruction de la gare routière par les bombes et de la gare ferroviaire par les tanks rend l'accès à la ville difficile.

1. La nature ne suffit pas : **Jean a été introduit par Pierre et par la grande porte*.
 2. Nous faisons l'hypothèse que le groupe verbal n'est pas un constituant.

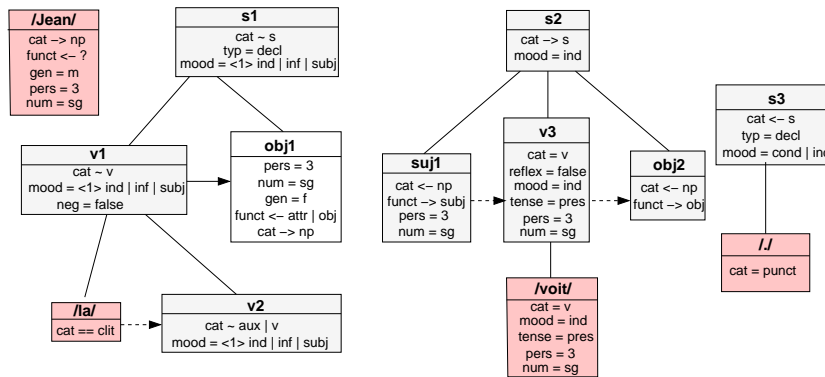


Figure 1. DAS D_1 associée à la phrase Jean la voit. par une grammaire d'interaction particulière

Les séquences coordonnées ne sont pas toujours constituées exclusivement d'arguments comme le montre l'exemple suivant, où l'on coordonne deux séquences composées d'un syntagme nominal objet direct et d'un marqueur temporel adjoint.

(5) Jean voit sa soeur lundi et son frère mardi.

Jusqu'à présent les conjoints étaient de même nature. Cependant, on peut dans certains cas coordonner des objets de natures différentes s'ils peuvent remplir la même fonction comme dans (6) où l'on coordonne un nom commun et un adjectif attribut.

(6) Pierre est bûcheron et très costaud.

On qualifie de disparate ce type de coordination. Il existe deux manières de le traiter : soit on ajoute des règles *ad hoc* et ce type de coordination constitue l'exception de la règle générale qui veut que les conjoints soient de même nature, soit l'on autorise ces coordinations comme naturelles et il faut donc un système d'unification plus fin pour donner une catégorie à ce type d'expressions et à celles avec lesquelles elles se combinent. Mais cette extension du système peut remettre en cause certains fondements d'un formalisme, comme le fait (Sag, 2002) pour HPSG. Cet aspect mérite d'être développé en détail, c'est pourquoi dans la suite de l'article nous n'étudierons que les cas où les conjoints sont de même nature.

Nous ne parlerons pas non plus, dans cet article, de la modélisation des coordinations à redoublement, des coordinations incidentes (Abeillé, 2003), des énumérations, des coordinations comportant plus de deux conjoints, ni des spécificités de certaines conjonctions de coordination comme *ni*, *car*, *mais*, par exemple.

3. Les grammaires d'interaction

Nous présenterons les GI de façon relativement informelle, le but étant que le lecteur en ait une compréhension suffisante pour appréhender la suite de l'article. Pour une présentation plus complète, il pourra se reporter à (Perrier, 2004). Néanmoins, nous avons apporté certaines modifications au formalisme, la principale étant un raffinement du système des polarités par introduction de la polarité *virtuelle*, notée \sim .

3.1. Les descriptions d'arbres syntaxiques

Les descriptions d'arbres syntaxiques (DAS) sont les objets de base du formalisme. Une DAS est un ensemble de nœuds et de relations de domination et de pré-séance entre ces nœuds. Les nœuds représentent des syntagmes et les relations des contraintes relatives à la position de ces syntagmes les uns par rapport aux autres. Les propriétés morphosyntaxiques de ces syntagmes sont exprimées par des structures de traits attachées aux nœuds. La figure 1 montre D_1 , un exemple simplifié de DAS qui est associée à la phrase *Jean la voit*. D_1 est divisée en quatre composantes connexes, chacune étant associée à un mot de la phrase. Les signes de ponctuation étant considérés aussi comme des mots, la dernière composante de la DAS est associée au point.

Les nœuds peuvent être *pleins* ou *vides*, selon qu'ils correspondent à une forme phonologique pleine ou pas. Cette propriété peut aussi ne pas être spécifiée³. Les nœuds vides représentent, dans une construction spéciale, la trace qu'occuperaient des syntagmes dans une construction canonique⁴. Ainsi dans l'exemple D_1 , la partie de la DAS relative au pronom personnel clitique *la* comporte un nœud vide obj_1 représentant la position canonique de l'objet direct juste après le noyau verbal v_1 , et cette position est vide car l'objet est ici fourni par le pronom clitique *la*.

Certains nœuds pleins ne représentent pas des syntagmes composés mais des mots qui sont les ancres des descriptions correspondantes. Ceux-ci sont représentés par des rectangles foncés et les mots associés figurent en en-tête des rectangles.

Les relations entre les nœuds d'une description peuvent avoir les types suivants :

relations de domination immédiate : $N > M$ signifie que le syntagme M est un constituant immédiat de N . Ainsi, le syntagme s_2 est composé de trois constituants immédiats : $subj_1$, v_3 et obj_2 .

relations de domination sous-spécifiée : $N \overset{\sim}{>} M$ signifie que le syntagme M est inclus dans N à une profondeur indéterminée. À la limite, M peut s'identifier

3. Sur les graphiques présents dans cet article, nous ne faisons pas de différence entre les nœuds pleins et ceux pour lesquels cette propriété n'est pas spécifiée. Ils sont tous représentés par des rectangles colorés alors que les nœuds vides sont représentés par des rectangles blancs.

4. En utilisant le mot «trace», nous ne faisons pas référence à une théorie linguistique particulière, celle de Chomsky notamment, mais nous voulons simplement représenter certaines positions argumentales canoniques qui ne sont pas occupées dans certaines constructions particulières. Cela permet de se ramener à un cadre uniforme et de simplifier les grammaires.

à N . Cette relation permet d'exprimer une dépendance syntaxique non bornée. On peut contraindre la relation qui devient $N \overset{*}{>} [f_1 = v_1, \dots, f_n = v_n]M$ avec la signification suivante : tout syntagme qui est inclus dans N et qui contient M doit avoir sa structure de traits qui s'unifie avec $[f_1 = v_1, \dots, f_n = v_n]$. Cela permet notamment de modéliser les barrières à l'extraction.

relations de préséance immédiate : $N \prec M$ signifie que le syntagme N précède immédiatement le syntagme M dans l'ordre linéaire des mots. Ainsi le noyau verbal v_1 précède immédiatement la trace vide obj_1 de l'objet du verbe.

relations de préséance sous-spécifiée : $N \overset{*}{\prec} M$ signifie que le syntagme N précède le syntagme M au sens strict. Par exemple, le clitique la précède le verbe v_2 auquel il s'adjoit mais entre les deux, il peut s'intercaler un autre clitique.

relations fils gauche et fils droit : $N > [M \dots]$ signifie que le syntagme M est le constituant immédiat de N le plus à gauche. De la même façon, $N > [\dots M]$ signifie que le syntagme M est le constituant immédiat de N le plus à droite. On peut ainsi exprimer le fait que le point final se trouve complètement à droite de la phrase qu'il ponctue par la relation $s_3 > [\dots / \cdot /]$.

Ce système de relations n'est pas figé et rien n'interdit de l'enrichir.

3.2. Le système des traits polarisés

L'originalité des GI est que les traits morphosyntaxiques associés aux syntagmes sont polarisés. Classiquement dans les grammaires d'unification, un trait est associé à une valeur dans un couple (*trait, valeur*). Dans les GI, un trait est associé à une polarité et à une valeur dans un triplet (*trait, polarité, valeur*). Une polarité peut être \leftarrow , \rightarrow , $=$ et \sim pour dire qu'un trait est *négatif*, *positif*, *neutre* ou *virtuel*. Un trait négatif représente une ressource attendue, un trait positif une ressource disponible et un trait neutre une propriété qui ne se comporte pas comme une ressource consommable. Dans la DAS D_1 , le nœud $subj_1$ est porteur de trois types de polarités : les traits $cat \leftarrow np$ et $funct \rightarrow subj$ signifient que le nœud attend un groupe nominal pour lui fournir la fonction *sujet*.

La polarité *virtuelle* $f \sim v$ n'existait pas dans la version précédente des GI (Perrier, 2004). Il a semblé utile de l'introduire pour une meilleure modélisation de l'adjonction de modificateurs à un syntagme. Nous reprenons en cela l'idée de (Nasr, 1995) avec son système de polarités noires et blanches. Auparavant, l'adjonction de modificateurs était effectuée comme dans beaucoup de formalismes (grammaires d'arbres adjoints, GC . . .) par ajout d'un niveau supplémentaire dans l'arbre syntaxique où était le syntagme modifié : à la place d'un nœud de catégorie X était inséré un arbre formé d'une racine de catégorie X et de ses fils : le modificateur et le syntagme initial de catégorie X objet de la modification. Si cette introduction d'un niveau supplémentaire est parfois justifiée, le plus souvent elle vient introduire une complexité artificielle qui pose problème pour la modélisation. Une autre façon de faire, qui est utilisée dans certains formalismes (grammaires de dépendance, grammaires de substitution de descriptions (Rambow *et al.*, 2001)), consiste à ajouter le

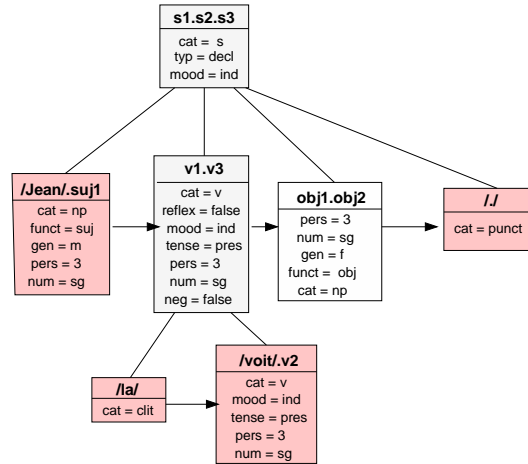


Figure 2. Modèle M_1 de la DAS D_1 correspondant à la phrase Jean la voit.

modificateur comme fils supplémentaire du nœud qu’il modifie sans rien changer au reste de l’arbre syntaxique dans lequel se situe le nœud modifié. En anglais, on parle alors de *sister adjunction* (Rambow *et al.*, 2001). Dans la DAS D_1 , on a appliqué cette nouvelle vision des modificateurs au clitique *la* qui peut être considéré comme un modificateur du noyau verbal auquel il s’adjoint. C’est pourquoi les traits *cat* des nœuds s_1 , v_1 et v_2 sont tous virtuels, ce qui signifie qu’ils doivent nécessairement se combiner avec des traits qui ne sont pas virtuels.

Les valeurs d’un trait t sont des ensembles finis d’atomes $\{v_1, \dots, v_n\}$ ⁵ pris dans un domaine fini D_t associé à t . Intuitivement, cela signifie que t peut prendre indifféremment une des valeurs atomiques v_1, \dots, v_n . La valeur d’un trait peut être munie d’un indice utilisé pour exprimer le partage de cette valeur par d’autres traits. C’est le cas dans la figure 1 pour le trait *mood* qui est partagé entre les nœuds s_1 , v_1 et v_2 .

3.3. Modèles saturés et minimaux d’une DAS

Une DAS peut être considérée comme un arbre syntaxique sous-spécifié ou comme un ensemble de contraintes définissant une famille d’arbres syntaxiques et chacun de ces arbres peut être vu comme un modèle de la description correspondante.

5. Lorsque l’ensemble se réduit à un singleton $\{v_1\}$, on identifie le singleton avec son unique élément v_1 ; sinon, on l’écrit sous forme d’une disjonction $v_1 | \dots | v_n$.

Pour préciser le rapport entre une description et ses modèles, il est nécessaire de préciser la forme de ces modèles, c'est-à-dire des *arbres syntaxiques*. Un arbre syntaxique est un arbre fini ordonné⁶ dont chaque nœud est muni d'une structure de traits de la forme $[t_1 = v_1, \dots, t_n = v_n]$ avec $v_i \in D_{t_i}$.

La définition d'un modèle d'une description s'inscrit dans la théorie des modèles pour la logique des prédicats, l'univers d'interprétation d'une description étant restreint aux arbres syntaxiques tels qu'ils viennent d'être définis. Un modèle d'une description d'arbre polarisé D est un couple (T, I) dans lequel T est un arbre syntaxique et I une fonction d'interprétation qui associe chaque nœud de D à un nœud de T préservant à la fois les contraintes de domination, de préséance et d'étiquetage (voir (Perrier, 2004) pour une définition formelle).

Ainsi, la figure 2 présente un modèle M_1 de la DAS D_1 . La fonction d'interprétation I est représentée implicitement : l'en-tête de chaque nœud est une fusion des noms de ses antécédents dans D_1 .

Néanmoins, la notion générale de modèle n'est pas suffisante pour deux raisons :

- toute description qui a au moins un modèle en a une infinité, mais tous ne sont pas pertinents à représenter la syntaxe d'une phrase ; pour cela, ils doivent avoir une propriété de *minimalité*, en ce sens qu'ils doivent ajouter un minimum d'information par rapport à la description qu'ils représentent ;

- les polarités n'interviennent pas du tout dans la notion générale de modèle ; or, elles ont été introduites pour exprimer la différence entre structures syntaxiques saturées et structures syntaxiques non saturées ; les structures saturées représentent la syntaxe de phrases grammaticales, ce qui est exprimé par une propriété particulière des modèles que nous désignons par le terme de *saturation*.

Définissons plus précisément ces propriétés de minimalité et de saturation.

- Un modèle (T, I) d'une DAS D est *minimal* si tout couple de nœuds père-fils dans T interprète un couple de nœuds de D liés par une relation de domination immédiate et si tout nœud porteur d'un trait $t = v$ dans T interprète un nœud de D porteur d'un trait t .

- Un modèle (T, I) d'une DAS D est *saturé* si pour tout nom de trait t porté par un nœud N de T , l'ensemble des traits de nom t de D portés par un nœud interprété par N vérifie l'une ou l'autre des propriétés :

- il contient exactement un trait positif et un trait négatif ;
- il ne contient ni trait positif ni trait négatif, et au moins un trait neutre.

Cette propriété de saturation donne une sémantique aux polarités : alors que les polarités neutres ne jouent aucun rôle, toute polarité positive doit nécessairement fusionner avec une polarité négative dans un modèle saturé et réciproquement, et une polarité

6. Dans un arbre ordonné, chaque nœud a l'ensemble de ses fils totalement ordonné et cet ordre est ensuite propagé à leurs descendants.

virtuelle doit nécessairement fusionner avec une qui ne l'est pas. Ce système de polarités est très voisin de celui que S. Kahane propose pour polariser les formalismes linguistiques les plus classiques (Kahane, 2006).

Il est facile de vérifier que l'arbre syntaxique M_1 de la figure 2 est un modèle minimal et saturé de la description D_1 de la figure 1.

3.4. Construction pas à pas de modèles saturés et minimaux d'une DAS

La définition qui vient d'être présentée nous permet de vérifier si un arbre syntaxique donné est un modèle saturé et minimal d'une DAS donnée mais elle ne nous fournit pas le moyen de construire un tel modèle. Or, pour réaliser les contraintes de polarités, il est nécessaire que certains nœuds de la description initiale soient fusionnés dans le modèle résultant. Au lieu de réaliser cette fusion d'un seul coup, nous allons le faire pas à pas en considérant les nœuds deux par deux. C'est cette opération de *fusion de deux nœuds* qui va constituer la base du calcul qui va permettre de construire des modèles saturés et minimaux de descriptions.

La fusion de deux nœuds entraîne l'unification des structures de traits associées. Cette notion est classique pour les grammaires d'unification où les structures de traits sont des ensembles de couples (*trait, valeur*), mais pour les grammaires d'interaction, ces structures sont des ensembles de triplets (*trait, polarité, valeur*). Unifier deux traits polarisés (t, p_1, v_1) et (t, p_2, v_2) nous donne un nouveau trait polarisé $(t, p_1 + p_2, v_1 \cap v_2)$ mais il est nécessaire de définir ce que nous entendons par la somme $p_1 + p_2$ des deux polarités p_1 et p_2 . Cette définition doit être conforme à la notion de modèle saturé telle qu'elle a été décrite précédemment. Pour cela, nous avons besoin d'introduire une polarité supplémentaire, la polarité *saturée*, notée \leftrightarrow . Cette polarité, qui n'ajoute rien au formalisme, est seulement là pour contrôler qu'un trait du modèle interprète au plus un trait positif et un trait négatif. Alors, la somme $p_1 + p_2$ de deux polarités p_1 et p_2 est donnée par la table suivante où une case vide signifie que l'addition n'est pas définie pour les valeurs de p_1 et p_2 correspondantes.

	←	→	~	=	↔
←		↔	←	←	
→	↔		→	→	
~	←	→	~	=	↔
=	←	→	=	=	↔
↔			↔	↔	

La table étant symétrique, l'addition des polarités est commutative. On vérifie aussi qu'elle est associative. Ces deux propriétés sont essentielles lorsque l'on conçoit la construction de modèles saturés et minimaux de DAS comme une suite de fusions de nœuds. Cela veut dire que l'ordre des fusions n'a pas d'incidence sur le résultat, ni l'ordre dans lequel on considère deux nœuds à fusionner entre eux.

Compte tenu de notre définition d'un modèle saturé d'une DAS, une suite de fusions de nœuds menant à un modèle saturé doit viser à éliminer tous les traits positifs,

négatifs et virtuels d'une DAS pour ne conserver que les traits neutres ou saturés.

– Les traits positifs et négatifs vont être éliminés par des fusions particulières que nous appellerons des *neutralisations*. Une fusion de deux nœuds N_1 et N_2 est une *neutralisation* si ceux-ci sont respectivement porteurs d'un trait $(t \rightarrow v_1)$ et d'un trait $(t \leftarrow v_2)$. Dans le nœud fusionné, ceux-ci sont remplacés par le trait $(t \leftrightarrow v_1 \cap v_2)$.

– Les traits virtuels vont être éliminés par des fusions que nous appellerons des *réalisations*. Une fusion de deux nœuds N_1 et N_2 est une *réalisation* si ceux-ci sont respectivement porteurs d'un trait $(t \sim v_1)$ et d'un trait (t, p_2, v_2) où p_2 est différent de \sim . Dans le nœud fusionné, ceux-ci sont remplacés par le trait $(t, p_2, v_1 \cap v_2)$.

Nous avons maintenant une méthode pour construire les modèles saturés et minimaux d'une description donnée : elle consiste à itérer les neutralisations et les réalisations jusqu'à obtention d'une DAS ne contenant que des traits neutres et saturés.

L'intérêt de la méthode est qu'elle conduit à spécifier de plus en plus la DAS de départ de telle façon que la détermination du modèle cherché devient de plus en plus simple. Si la grammaire est bien conçue, dans la plupart des cas, la DAS finale coïncide avec son modèle. Par exemple, c'est ainsi que l'on peut passer de la DAS D_1 de la figure 1 jusqu'à son modèle M_1 donné par la figure 2 .

4. Modélisation de la coordination

Dans cette section, nous présentons le principe général de notre modélisation que nous déclinerons selon les sous-phénomènes présentés en section 2.

4.1. Le principe de superposition des interfaces

Il existe plusieurs propositions de modélisation de la coordination (Abeillé, 2003) que l'on peut diviser en deux groupes : symétriques et asymétriques.

Dans l'approche asymétrique, la conjonction et l'un des conjoints forment un syntagme qui vient modifier l'autre conjoint, en d'autres termes l'un des conjoints est plus proche de la conjonction. Souvent l'un des conjoints est le spécificateur et l'autre le complément. Cependant, cette modélisation implique que de nombreux traits sont partagés entre le spécificateur, le complément et le reste du syntagme, ce qui est tout à fait inhabituel et contraire aux hypothèses communément admises (Borsley, 2005). Avec cette approche, la coordination de non-constituants doit s'interpréter comme la coordination d'un constituant complet et d'un constituant auquel il manque quelque chose, comme le présentent (Beavers et Sag, 2004).

Dans l'approche symétrique, les deux conjoints sont à égale distance de la conjonction et sont habituellement de nature et de fonction équivalentes. Ils forment avec la conjonction un segment de même type. Cette approche laisse plus de liberté quant à la modélisation de la coordination de non-constituants. On peut comme précédemment

y voir une coordination de constituants avec effacement de matériel ou alors admettre que l'on peut coordonner des segments quelconques du moment qu'ils ont le même *comportement syntaxique*.

L'approche symétrique est celle qui est choisie pour modéliser la coordination dans les GC (Steedman, 1985; Steedman, 1990; Morrill, 1994). C'est aussi celle que nous proposons de prendre comme point de départ pour les GI. Dans les GC, les conjonctions ont le type polymorphe $\forall X, (X \setminus X) / X^7$. Ce type indique qu'une conjonction cherche à sa droite et à sa gauche des segments de même type X et attribue à leur réunion le type X . Malheureusement, ce type n'est pas suffisamment contraint et il peut être nécessaire de trouver un fragment du calcul du second ordre qui convienne mieux, tant du point de vue calculatoire que linguistique (Emms, 1993).

Cette modélisation s'étend naturellement aux coordinations disparates si l'on ajoute deux constructeurs de type additifs pour le syncrétisme et l'ambiguïté (Bayer et Johnson, 1995).

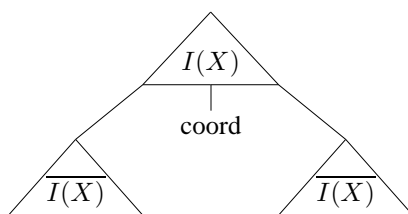


Figure 3. La forme générale des DAS associées aux conjonctions de coordination

Nous voulons donc coordonner des conjoints qui ont le même comportement syntaxique, ce qui revient à avoir le même type dans les GC et à avoir la même interface dans les GI. Nous avons également besoin de la notion de description duale.

Définition 1 Nous appelons interface d'une DAS ce qui subsiste de cette DAS si l'on ne considère que la partie qui doit interagir avec d'autres DAS, c'est-à-dire les traits polarisés et les traits virtuels ainsi que les nœuds qui les contiennent – les nœuds utiles – et les relations entre ces nœuds.

La duale d'une DAS est une DAS où les polarités sont opposées (\rightarrow devient \leftarrow et réciproquement; \sim devient $=$). Les traits neutres restent neutres mais, comme nous appliquons cette notion aux interfaces, ils ne seront pas considérés.

Nous supposons que pour être coordonnés, les conjoints doivent avoir des DAS de même interface. La coordination doit neutraliser leurs interfaces et également présenter cette même interface au reste de la phrase. À partir de ces hypothèses, nous

7. Par associativité, ce type est équivalent à $\forall X, X \setminus (X / X)$ et il n'y a donc pas de préférence pour un conjoint.

pouvons établir un patron général pour une DAS associée aux conjonctions de coordination, présenté sur la figure 3. Cette DAS est constituée de trois parties :

- deux parties dites basses et destinées à neutraliser les conjoints. Elles sont construites à partir de $\overline{I(X)}$, la duale de l’interface commune des conjoints.
- une partie dite haute destinée à interagir avec le reste de la phrase. Elle est constituée de l’interface commune des deux conjoints, notée $I(X)$.

Il faut nuancer ce schéma. Les trois parties contiennent souvent des nœuds qui ne figurent pas dans l’interface ni sa duale mais qui servent à *placer* correctement les nœuds de l’interface les uns par rapport aux autres. Également, des traits neutres, c’est à dire absents de l’interface ou sa duale, sont ajoutés pour permettre de modéliser les règles d’accords entre les conjoints et le groupe coordonné qui sont propres à chaque langue et même à chaque type de syntagme.

Dans le cas de la coordination elliptique, il n’y a pas symétrie de la forme de surface et il est nécessaire de faire apparaître la duplication de l’élément présent dans le premier conjoint et élidé du second dans la DAS associée à la conjonction de coordination, dans le but de retrouver une certaine symétrie.

4.2. Les coordinations de constituants simples

Dans la situation la plus facile à comprendre, l’interface de chaque conjoint se résume à un seul nœud. C’est le cas des coordinations de constituants simples. La DAS associée contient donc trois nœuds polarisés, comme dans la figure 4 qui représente un emploi de la conjonction pour une coordination de syntagmes nominaux.

Cette DAS contient également des traits neutres qui modélisent les règles d’accords. Ces traits ne dépendent pas de notre modélisation générale de la coordination. Ici nous avons affaire à la coordination de deux syntagmes nominaux à la troisième personne où le premier est masculin. L’expression coordonnée est donc également au masculin et à la troisième personne, du pluriel quel que soit le nombre de chaque conjoint.

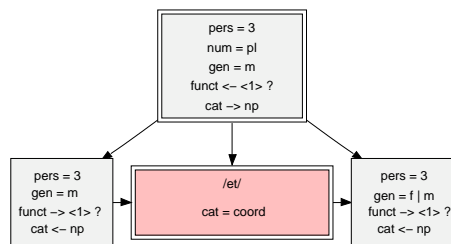


Figure 4. Une DAS pour et coordonnateur de syntagmes nominaux

4.3. La coordination de modificateurs

Dans cette partie, nous nous intéressons à la coordination de modificateurs en reprenant l'idée d'une coordination en trois parties, basée sur la notion d'interface. Nous avons choisi une modélisation classique et non l'utilisation de la *sister adjunction* avec les traits virtuels. Nous allons prendre comme exemples les adjectifs épithètes. La principale différence avec la coordination de constituants simples est que, dans le cas précédent seul le nœud racine était polarisé et donc l'interface se réduisait à un nœud. Dans le cas des modificateurs, le nœud racine et un de ses fils (le plus à droite ou le plus à gauche selon le modificateur) ont pour le trait *cat* la même valeur, mais des polarités opposées, de manière à consommer le mot modifié et à renvoyer la même structure de traits. Leur interface est donc constituée de ces deux nœuds.

Pour la coordination de modificateurs, la partie haute $I(X)$ de la DAS n'est donc plus atomique mais est un arbre unaire où nœud père et nœud fils ont des polarités opposées. Une telle description est présentée sur la figure 5. On peut observer que les parties basses contiennent un nœud ne faisant pas partie de l'interface. Ce nœud sert à placer correctement les adjectifs et à ne sélectionner que des épithètes droites. Un exemple de coordination d'épithètes est représentée sur la figure 6.

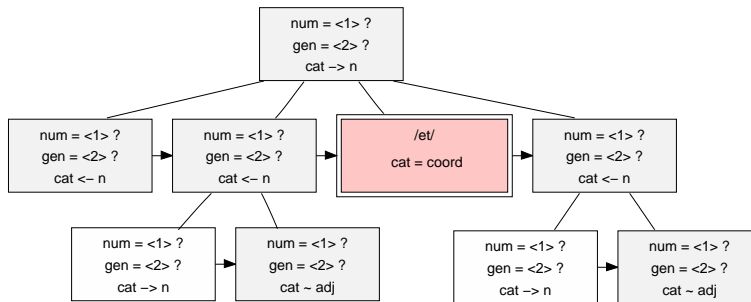


Figure 5. DAS pour la conjonction et dans le cas de la coordination d'adjectifs épithètes droites

4.4. La coordination de non-constituants

De nombreuses langues autorisent la coordination de segments qui ne sont pas des constituants, qui ne forment pas d'unité syntaxique. Nous allons montrer que la superposition partielle de descriptions d'arbres propre aux GI joue dans la modélisation de ce type de coordination un grand rôle.

Dans cette partie nous traitons d'abord du cas de montée de nœuds simples. Puis, nous étendrons la modélisation à la coordination de séquences de compléments.

4.4.1. La montée de nœuds

On peut distinguer les montées de nœuds gauches et les montées de nœuds droits, selon que la partie *factorisée* se trouve à gauche ou à droite du segment coordonné.

- (7) (a) Jean [boit du vin et mange]. (montée de nœud gauche)
- (b) [Marie aime mais Pierre déteste] les glaces. (montée de nœud droit)

Nous nous plaçons toujours dans le cadre d’une modélisation symétrique du phénomène. La seule différence avec la coordination de constituants complets est qu’ici les conjoints ont des dépendances non satisfaites. Leur interface a donc une structure plus complexe. La figure 7 représente la DAS associée à la conjonction dans l’exemple (7a). La partie haute comporte un nœud $cat \rightarrow s$, la coordination fournit une proposition au reste de la phrase et un autre nœud $cat \leftarrow np, funct \rightarrow subj$, la coordination requiert un syntagme nominal sujet. Dans les parties basses au contraire la conjonction demande une proposition à chaque conjoint et leur fournit un syntagme nominal sujet à chacun.

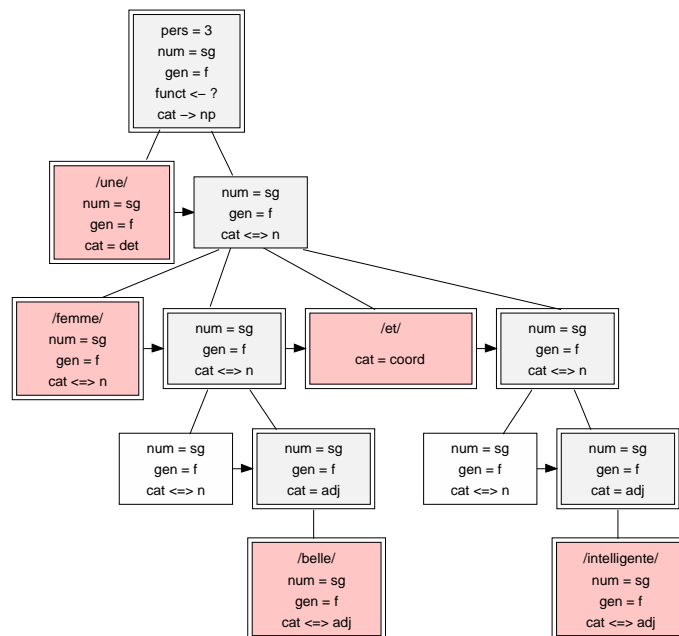


Figure 6. Coordination de modificateurs : DAS associée à une femme belle et intelligente

C'est la conjonction qui va leur fournir ces dépendances et présenter au reste de la phrase un seul bloc avec les mêmes demandes et les mêmes ressources que chacun des conjoints. On peut voir l'analyse de la phrase (7a) sur la figure 8.

La DAS associée à la conjonction ici n'autorise pas à aller chercher la dépendance profondément dans le conjoint⁸. C'est tout à fait envisageable en divisant le nœud associé en deux nœuds, le premier neutre dominant largement le second polarisé.

4.4.2. *Les séquences de compléments*

Dans la section 4.3, nous avons vu qu'il était envisageable de *décorer* les différentes parties d'une DAS associée à une conjonction de nœuds ne faisant pas partie de l'interface des conjoints dans le but de placer correctement ces conjoints. Ceci est une utilisation assez classique des traits neutres dans les GI : ils fournissent un contexte pour la superposition des traits polarisés. Nous utilisons cette technique pour modéliser la coordination de séquences de compléments. Les trois parties de la DAS vont contenir de nombreux nœuds neutres qui vont permettre de reconstituer une structure en constituants pour chaque conjoint. La DAS associée à la conjonction pour ce type de coordination possède toujours ces trois parties caractéristiques. Une DAS pour la conjonction de coordination en cas de coordination de séquences de compléments similaires à celui de l'exemple (8) est représentée sur la figure 9.

(8) L'entreprise offre [un emploi à l'ingénieur et un stage à l'étudiant].

Pour les phénomènes précédents, nous avons choisi une analyse symétrique mais nous aurions très bien pu choisir une analyse asymétrique, comme dans la section suivante pour les coordinations avec ellipse.

4.5. *La coordination avec ellipse*

Pour des phrases comme (3), nous modélisons un effacement – du verbe dans cet exemple. C'est à première vue la seule *entorse* à l'analyse symétrique que nous développons. Pourtant, c'est justement pour retrouver une certaine symétrie entre les conjoints que nous avons besoin de cette analyse.

Ici la DAS associée à la conjonction (cf. figure 10) vient neutraliser le premier conjoint, grâce aux polarités, fournir la partie manquante au second en la dupliquant depuis le premier, grâce aux traits neutres et à la coréférence, et présenter comme précédemment une interface unifiée au reste de la phrase. Le résultat est exposé sur la figure 11.

Bien sûr, pour des cas plus complexes les deux approches peuvent être utilisées conjointement. L'exemple (9) fait intervenir ellipse et montée de nœud droit.

8. Ceci n'a pas lieu d'être pour le sujet mais les dépendances non bornées sont autorisées pour l'objet.

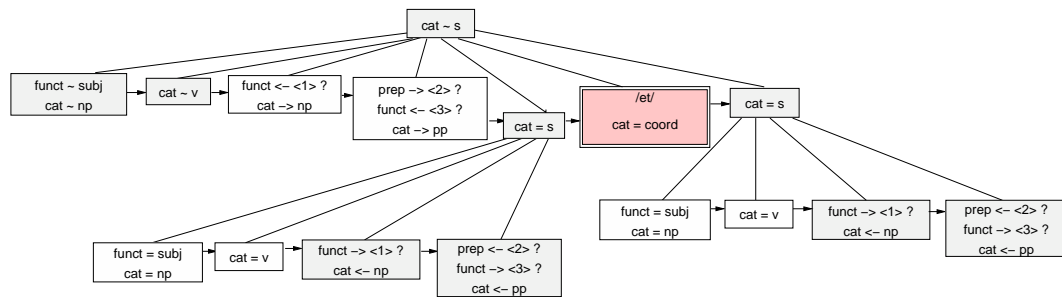


Figure 9. DAS de la conjonction pour une coordination de séquences de compléments

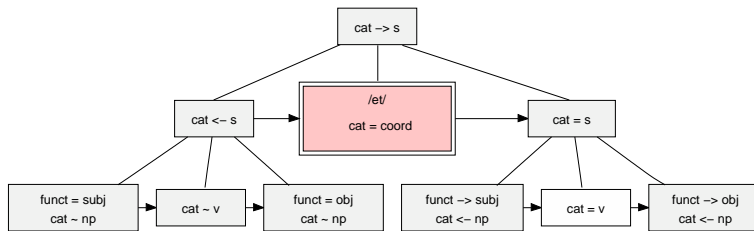


Figure 10. La DAS pour la coordination de propositions avec ellipse du second verbe

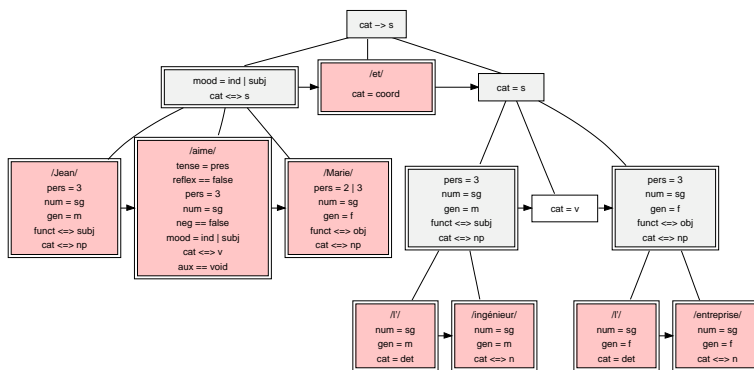


Figure 11. Coordination avec ellipse : DAS correspondant à Jean aime Marie et l'ingénieur l'entreprise

- (9) L'entreprise persuade le conseil et le directeur l'ingénieur que le spécialiste doit venir.

Ceci nous semble être un point clé de notre approche. Les GI sont suffisamment souples pour exprimer à la fois les aspects symétriques et les aspects asymétriques de la coordination dans une même description.

5. Comparaison avec d'autres formalismes

5.1. Les grammaires catégorielles

Notre analyse reprend celle développée dans les GC quelle qu'en soit la version, logique (Moortgat, 1996) ou combinatoire (Steedman, 1985; Steedman, 1990; Steedman, 2000), en remplaçant la montée de type par des constructions supplémentaires avec montée de nœud. Mais l'objection à cette modélisation, formulée dans (Beavers et Sag, 2004), ne tient pas pour notre proposition. Certaines langues présentent des dépendances non bornées mais pas de coordination de non-constituants. Or les GC utilisent les mêmes mécanismes, montée de type et raisonnement hypothétique ou leur contrepartie combinatoire, pour analyser les deux phénomènes. Ces grammaires peuvent difficilement interdire l'un sans l'autre. Les GI n'ont pas ce problème puisque nous modélisons la coordination et les dépendances à longue distance différemment.

Indéniablement, l'analyse avec utilisation de la montée de type dans les GC est séduisante par son élégance. Cependant, il faut nuancer l'approche catégorielle. Tout d'abord la rigidité de son noyau (les grammaires de Lambek) entre souvent en contradiction avec la plasticité des langues naturelles. Ainsi le changement de position des adverbes dans (10a) et (10b) correspond uniquement à un changement de la structure communicative. Or les GC modélisent difficilement (10b) car les adverbes n'ont pas une position périphérique dans la proposition qu'ils modifient. Avec les GI, on peut utiliser la *sister adjunction* pour modéliser le placement libre de l'adverbe.

- (10) a) Jean présente Marie à Pierre aujourd'hui et Claire à François demain
b) Jean présente aujourd'hui Marie à Pierre et demain Claire à François

D'autre part, les GC ont tendance à surgénérer parce que l'information donnée par le type d'une expression est trop pauvre. On peut en effet accepter les phrases fausses (11a) ou (11b) si l'on n'utilise pas une version des GC enrichie de modalités, comme par exemple celle présentée par (Morrill, 2002). Dans les GI, le problème n'existe pas grâce à la structure arborescente qui interdirait une telle analyse. En outre, les contraintes sur les dominations larges interdisent l'extraction à partir d'un seul conjoint, comme dans (11b) mais autorisent (11c) où l'extraction se fait dans la partie factorisée.

- (11) a) * [Jean et Marie sait que Pierre] dorment.

- b) * Jean que [Marie aime Pierre et Nathalie déteste _] dort.
 c) Jean que [Marie aime et Nathalie déteste] _ dort.

Les GC multimodales évitent certainement ces écueils en permettant de contrôler de manière très fine l'associativité et la commutativité de la composition syntaxique et théoriquement les phénomènes étudiés sont modélisables dans ce formalisme. En pratique, les GC multimodales à large couverture (Hockenmaier, 2006; Moot, 2007) sont extraites automatiquement et les techniques d'apprentissage utilisées peinent à modéliser les coordinations les plus difficiles sans surgénérer.

5.2. HPSG

Il existe de nombreuses propositions de modélisation de la coordination en HPSG. Ce formalisme a toutefois quelques problèmes à donner un statut aux non-constituants coordonnés dans le cadre d'une approche symétrique. C'est pourquoi l'approche asymétrique avec effacement d'éléments est souvent préférée.

(Beavers et Sag, 2004) cherchent à modéliser les coordinations de non-constituants différemment des constructions à dépendances non bornées. La modélisation proposée n'est pas symétrique, puisqu'elle cherche plutôt à rétablir une symétrie qui n'existe pas (plus) entre les conjoints dans la forme de surface. Elle suppose que l'un des conjoints est un constituant *défectif* et qu'il est possible de lui rajouter la partie manquante par copie d'éléments depuis l'autre conjoint.

Les auteurs présentent une série de règles pour proposer une approche unifiée de la montée de nœud droit et la coordination de séquences de compléments. Intuitivement, les parties à gauche du premier conjoint et à droite du dernier conjoint peuvent être réutilisées par les conjoints intermédiaires.

- (12) Jean est allé à Paris lundi, ~~Jean est allé~~ à Londres mardi et ~~Jean~~ ira à Shangai dimanche.

Cette modélisation analyse des phénomènes qui sont hors du champ du présent article comme l'accord partiel d'un des conjoints avec le verbe ou la semi-grammaticalité de certaines constructions. Néanmoins, l'approche par effacement souffre d'un défaut : puisqu'elle présuppose que la forme étendue (sans ellipse) est canonique, elle ne permet pas d'expliquer pourquoi (13a) est correct et (13b) ne l'est pas, comme le fait remarquer (Mouret, 2006).

- (13) a) Jean défend le droit à la différence et (non) pas la différence des droits.
 b) *Jean défend le droit à la différence et (non) pas Jean défend la différence des droits.

Plus récente, l'approche de (Mouret, 2006) pour la coordination de séquences de compléments est symétrique. Il ne semble pas que ces approches soient compatibles

et qu'une même grammaire HPSG puisse traiter certains aspects de la coordination symétriquement et d'autres asymétriquement, ce que font les GI très naturellement.

6. Implantation de la grammaire

Dans cette partie, nous décrivons l'organisation de la grammaire de la coordination que nous proposons. Cette grammaire est implantée et utilisée par notre analyseur *Leopar*⁹. C'est aussi un exemple d'implantation d'une grammaire de la coordination construite manuellement visant une couverture large. Cette implantation s'appuie sur l'outil XMG et la méthodologie décrite dans (Crabbé, 2005).

XMG, (Duchier *et al.*, 2005; Le Roux *et al.*, 2006), est un compilateur de métagrammaire, c'est à dire un outil destiné aux grammairiens. Il leur permet de se contenter d'une description concise de la grammaire que l'on appelle métagrammaire, qu'il va traduire vers une grammaire cible utilisable dans un système de TAL. Cette description se fait *via* un langage déclaratif de type PROLOG.

Selon la méthodologie définie par (Crabbé, 2005) pour XMG, le grammairien commence par définir des fragments de DAS réutilisables qui se comportent comme des modules que d'autres fragments peuvent importer de manière à construire des fragments plus importants qui finalement forment des DAS complètes. La réutilisation des fragments s'apparente à la programmation orientée objet avec une hiérarchie d'héritage multiple.

Nous ne détaillons que la partie de la métagrammaire portant sur la coordination de syntagmes nominaux avec éventuellement montée de nœud. La méthodologie est la même pour les autres types de coordination. L'organisation des fragments de la métagrammaire est représentée schématiquement sur la figure 12.

La classe initiale *Proto_Coordination* définit la forme en 3 parties d'une DAS associée à une conjonction de coordination. Ce fragment est réutilisé par tous les types de coordination. Ensuite sont définies des sous-hiérarchies pour chaque catégorie de conjoint : propositions, groupes nominaux, prépositionnels et adjectivaux.

Une hiérarchie pour un type de conjoint T commence toujours par un fragment *Pre_T_Coordination* qui fixe les traits et les valeurs communs aux DAS de ce type. Pour les syntagmes nominaux, *Pre_NP_Coordination* impose les traits *cat* \rightarrow *np* et *funct* \leftarrow ? signifiant ainsi que les coordinations sont des groupes nominaux qui doivent recevoir une fonction grammaticale (quelconque). Ce dernier trait est coindexé entre les conjoints et leur conjonction.

Après cette étape, les hiérarchies de coordination diffèrent sensiblement. Pour les syntagmes nominaux, l'étape suivante, décrite pas la classe *Agr_NP_Coordination*,

9. <http://www.loria.fr/equipes/calligramme/leopar>

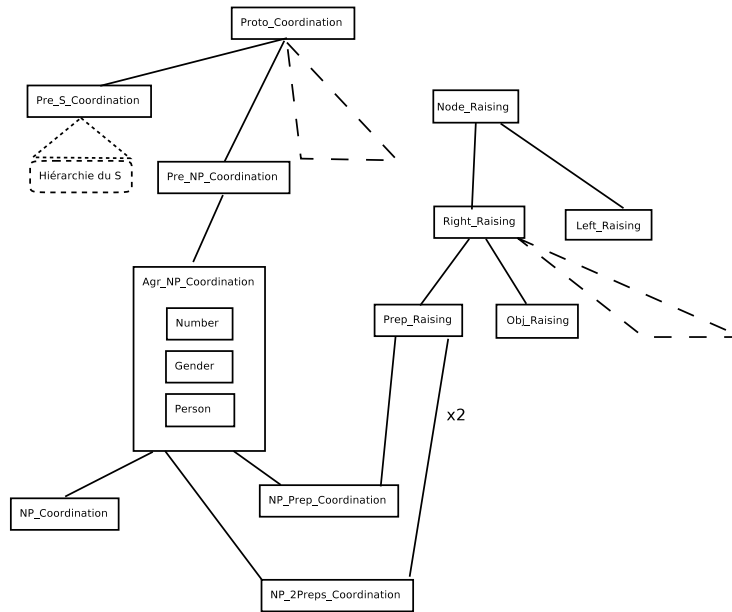


Figure 12. Organisation d'un fragment de la grammaire de la coordination

consiste à croiser les informations d'accord, principalement pour la personne, le genre et le nombre des conjoints et de la coordination. Cette étape est optionnelle¹⁰.

Ensuite, soit on évalue la classe `NP_Coordination` qui génère les DAS de syntagmes nominaux complets simples, soit on va croiser la classe `Agr_NP_Coordination` avec des classes modélisant des dépendances non satisfaites (pour les cas de montées de nœuds). Ces classes héritent toutes de `Right_Raising` qui indique l'existence de nœuds supplémentaires à droite dans les DAS associées et spécifient le type de dépendance (préposition, complémenteur...). Par exemple, pour la coordination (*la femme et la fille*) de *Jean*, la DAS associée à la conjonction est issue de la classe `NP_Prep_Coordination` qui croise les classes `Agr_NP_Coordination` et `Prep_Raising`. Pour la coordination (*l'acceptation ou le refus*) de *l'entrée aux mineurs*, `X_NP_Coordination` est croisé avec deux instances de `Prep_Raising`.

La classe `Right_Raising` hérite quant à elle de `Node_Raising` la classe abstraite qui déclare l'existence des nœuds porteurs des dépendances supplémentaires

10. C'est cette optionnalité qui permet de générer deux grammaires.

non satisfaites. Les montées de nœuds dans le cas des coordinations de propositions en héritent également.

XMG permet de créer rapidement une grammaire assez conséquente de la coordination en gardant une vue globale du phénomène. L'approche métagrammaticale s'inscrit bien dans notre modélisation puisque les DAS que nous voulons obtenir respectent toutes un patron général commun, défini dans les classes *hautes* telles que *Proto_Coordination* ou *Pre_NP_Coordination*, qui est ensuite spécifié et instancié vers des DAS concrètes.

Nous pouvons produire deux grammaires, selon que l'on tient compte des traits d'accord (450 DAS) ou pas (45 DAS). Cette différence de taille est un artefact du formalisme même, puisque nous ne pouvons pas exprimer dans la version actuelle des GI le fait que, par exemple, le genre du syntagme nominal coordonné est fonction du genre des conjoints. Nous ne pouvons qu'indiquer les corrélations entre les genres, c'est-à-dire lister les cas. Les GI doivent donc incorporer à l'avenir les fonctions de valeurs de trait, ce qui ne doit pas poser de problème majeur.

Ces deux grammaires sont intégrées à une grammaire plus générale du français et utilisées dans l'analyseur *Leopar*. Nous avons évalué notre implantation sur la TSNLP¹¹ et d'ores et déjà nous modélisons 80 % des phrases grammaticales contenant des coordinations. Les phrases agrammaticales sont rejetées intégralement.

7. Conclusion

Nous avons montré que les GI permettent de représenter les principaux aspects de la coordination. Il reste évidemment du travail pour arriver à les couvrir tous. Nous n'avons pas pu, par manque de place, détailler notre modélisation de la coordination disparate. Les valeurs de traits seraient alors construites à partir de constantes et de deux opérateurs représentant le syncrétisme et l'ambiguïté, comparables à la conjonction et à la disjonction additives de la logique linéaire. En cela, nous nous inspirons du travail fait dans les GC (Bayer et Johnson, 1995). L'opération de superposition ne se contente plus d'unifier les valeurs de traits, un calcul logique est nécessaire pour vérifier la superposabilité de deux traits et pour donner la valeur du trait dans le modèle.

Nous n'avons pas non plus évoqué les nombreux aspects informatiques du travail présenté. Parmi ceux-ci, nous en citerons deux :

1) Même sans tenir compte des traits d'accord, notre lexique associe 45 DAS aux conjonctions de coordination. Ceci est dû au polymorphisme intrinsèque à la coordination et tous les formalismes y sont confrontés. À l'heure actuelle, pour filtrer les entrées lexicales des conjonctions de coordinations, nous avons proposé une méthode qui étend celle présentée dans (Bonfante *et al.*, 2004) basée sur les polarités et la

11. La TSNLP est une suite de phrases tests pour les analyseurs syntaxiques. Elle n'est plus maintenue mais reste accessible à <http://cl-www.dfki.uni-sb.de/tsnlp/>

neutralité globale des sélections valides en cherchant l'existence d'une symétrie de polarités localisée autour des conjonctions de coordination (Bonfante *et al.*, 2006).

2) Bien que les DAS présentées semblent complexes, les polarités guident l'analyse syntaxique dans les GI. Cependant, les phrases contenant des coordinations se révèlent en général plus complexes à analyser que la moyenne – ne serait-ce que par leur longueur. Nous avons donc été amenés à développer de nouvelles stratégies d'analyse, notamment deux algorithmes l'un fondé sur CKY et l'autre sur Earley, moins sensibles à la longueur des syntagmes que l'actuel *shift/reduce*.

8. Bibliographie

- Abeillé A., « A Lexicon- and Construction-Based Approach to Coordinations », *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar*, p. 5-25, 2003.
- Abeillé A., « In defense of lexical coordination », in O. Bonami, P. Cabredo Hofherr (eds), *Empirical Issues in Syntax and Semantics 6*, p. 7-36, 2006.
- Bayer S., Johnson M., « Features and agreement », *33rd Meeting of the Association for Computational Linguistics, San Francisco*, p. 70-76, 1995.
- Beavers J., Sag I., « Coordinate Ellipsis and Apparent Non-Constituent Coordination », *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, p. 48-69, 2004.
- Bonfante G., Guillaume B., Perrier G., « Polarization and abstraction of grammatical formalisms as methods for lexical disambiguation », *20th Conference on Computational Linguistics, CoLing'2004, Genève, Switzerland*, p. 303-309, 2004.
- Bonfante G., Le Roux J., Perrier G., « Lexical disambiguation with Polarities and Automata », in O. H. Ibarra, H.-C. Yen (eds), *The 11th International Conference on Implementation and Application of Automata (CIAA 2006)*, 2006.
- Borsley R., « Against ConjP », *Lingua*, vol. 115, n° 4, p. 461-482, 2005.
- Carston R., Blakemore D., « Introduction to Coordination : Syntax, Semantics and Pragmatics », *Lingua*, vol. 115, n° 4, p. 353-358, 2003.
- Crabbé B., Représentation informatique de grammaires fortement lexicalisées, PhD thesis, Université Nancy 2, 2005.
- Dowty D., « Type raising, functional composition and non constituent conjunction », in R. Oehrle, E. Bach, D. Wheeler (eds), *Categorial Grammars and Natural Language Structures*, D. Reidel Publishing Company, p. 153-197, 1988.
- Duchier D., Le Roux J., Parmentier Y., « XMG : Un Compilateur de Méta-Grammaires Extensible », *Actes de TALN05*, 2005.
- Emms M., « Parsing with polymorphism », *EACL*, p. 120-129, 1993.
- Godard D., « Problèmes syntaxiques de la coordination et propositions récentes dans les grammaires syntagmatiques », *Langages*, vol. 160, p. 3-24, 2005.
- Hockenmaier J., « Creating a CCGbank and a wide-coverage CCG lexicon for German », *Proceedings of COLING/ACL*, 2006.

- Kahane S., « Polarized Unification Grammars », *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, p. 137-144, 2006.
- Le Roux J., Crabbé B., Parmentier Y., « A Constraint Driven Metagrammar », *Proceedings of TAG+8*, 2006.
- Lecomte A., Retoré C., « Extending Lambek Grammars : a Logical Account of Minimalist Grammars », *ACL*, p. 354-361, 2001.
- Moortgat M., « Categorical Type Logics », in J. van Benthem, A. ter Meulen (eds), *Handbook of Logic and Language*, Elsevier, chapter 2, 1996.
- Moot R., *The Complexity of Lexical Descriptions and its Relevance to Natural Language Processing : A Supertagging Approach*, MIT Press, chapter Automated Extraction of Type-Logical Supertags from the Dutch Spoken Corpus, 2007.
- Morrill G., *Type Logical Grammar : Categorical Logic of Signs*, Kluwer, Dordrecht, 1994.
- Morrill G., Islands, coordination and parasitic gaps, Technical report, Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, 2002.
- Mouret F., « A phrase structure approach to argument cluster coordination in French », *13th International Conference on Head-Driven Phrase Structure Grammar*, 2006.
- Nasr A., « A Formalism and a Parser for Lexicalised Dependency Grammars », *4th International Workshop on Parsing Technologies (IWPT)*, 1995.
- Perrier G., « La sémantique dans les grammaires d'interaction », *Traitement Automatique des Langues*, vol. 45, n° 3, p. 123-144, 2004.
- Rambow O., Vijay-Shanker K., Weir D., « D-Tree Substitution Grammars », *Computational Linguistics*, vol. 27, n° 1, p. 87-121, 2001.
- Rogers J., Vijay-Shanker K., « Reasoning with descriptions of trees », *30th Annual Meeting of the Association for Computational Linguistics*, p. 72-80, 1992.
- Sag I., « Coordination and Underspecification », *The Proceedings of the 9th International Conference on Head-Driven Phrase Structure Grammar*, p. 267-291, 2002.
- Sarkar A., Joshi A. K., « Coordination in Tree Adjoining Grammars : Formalization and Implementation », *COLING*, p. 610-615, 1996.
- Seddah D., Sagot B., « Modélisation et analyse des coordinations elliptiques par l'exploitation dynamique des forêts d'analyse », *Actes de TALN 2006*, p. 609-618, 2006.
- Steedman M. J., « Dependency and Coordination in the Grammar of Dutch and English », *Language*, vol. 61, n° 3, p. 525-568, Sept., 1985.
- Steedman M. J., « Gapping as Constituent Coordination », *Linguistics and Philosophy*, vol. 13, n° 2, p. 207-263, 1990.
- Steedman M. J., *The syntactic process*, MIT Press, Cambridge, Massachusetts, 2000.