

## Descriptions d'arbres avec polarités : les Grammaires d'Interaction

Guy Perrier  
LORIA, CNRS  
BP 239

54506 Vandœuvre-lès-Nancy Cedex France  
perrier@loria.fr

### Mots-clefs – Keywords

Grammaire formelle, Grammaire Catégorielle, Description d'arbres  
Formal Grammar, Categorical Grammar, Tree description

### Résumé - Abstract

Nous présentons un nouveau formalisme linguistique, les Grammaires d'Interaction, dont les objets syntaxiques de base sont des descriptions d'arbres, c'est-à-dire des formules logiques spécifiant partiellement des arbres syntaxiques. Dans ce contexte, l'analyse syntaxique se traduit par la construction de modèles de descriptions sous la forme d'arbres syntaxiques complètement spécifiés. L'opération de composition syntaxique qui permet cette construction pas à pas est contrôlée par un système de traits polarisés agissant comme des charges électrostatiques.

We present a new linguistic formalism, Interaction Grammars. The basic syntactic objects are tree descriptions, that is logical formulas which specify syntactic trees partially. In this context, parsing consists in building models of descriptions which are completely specified syntactic trees. The operation of syntactic composition which allows this construction to go step by step is controlled by a system of polarized features acting as electrostatic charges.

### Introduction

La notion de *description d'arbre* a été introduite par (Vijay-Shanker, 1992) pour rendre l'opération d'adjonction des Grammaires d'Arbres Adjoints (TAG) monotone. Elle a ensuite été formalisée et étudiée d'un point de vue mathématique par (Rogers & Vijay-Shanker, 1992). Elle a enfin été à la base de nouveaux formalismes linguistiques (Kallmeyer, 1999; Rambow *et al.*, 2001). L'intérêt est de remplacer la manipulation d'arbres syntaxiques complètement spécifiés par la manipulation de spécifications partielles de ces arbres. Ainsi ces propriétés peuvent être

combinées librement pour exprimer toute la complexité et la souplesse d'une langue. Cela permet entre autres une organisation modulaire et hiérarchique des grammaires dans laquelle l'information est factorisée au maximum (Candito, 1999). Ceci est décisif pour la gestion et la maintenance des grammaires à large couverture. Les descriptions d'arbres permettent aussi de représenter des structures sous-spécifiées et d'exprimer simplement l'ambiguïté d'une langue (Egg *et al.*, 1998; Muskens, 2001). Enfin, cette approche s'inscrit dans une remise en cause plus générale de la syntaxe générative (Blache, 1999).

Dans cette approche, l'analyse syntaxique consiste à chercher des modèles de descriptions d'arbres sous forme d'arbres syntaxiques complètement spécifiés. Dans toute sa généralité, ce processus est hautement indéterministe. Pour être réaliste, tout formalisme linguistique basé sur les descriptions d'arbres restreint cet indéterminisme en contraignant la syntaxe des descriptions et en définissant à partir de là une opération de composition syntaxique des descriptions qui permette de construire pas à pas des arbres syntaxiques complètement spécifiés.

Les Grammaires d'Interaction, si elles s'inscrivent dans ce cadre, trouvent leur originalité dans le mécanisme qui fonde leur opération de composition syntaxique et qui est régi par un système de *polarités*. Cette notion de polarité est liée à la dualité besoins-ressources qui est à la base des Grammaires Catégorielles (Retoré, 2001). Les syntagmes y sont considérés comme des ressources consommables. Selon que ces ressources sont disponibles ou qu'elles représentent un besoin, elles sont positives ou négatives d'où le lien avec la notion de polarité. Dans le processus de composition syntaxique, les ressources complémentaires vont chercher à se neutraliser. Par exemple, un verbe transitif se cherche un sujet et un objet, ce qui peut se traduire par le fait qu'il est porteur de deux besoins ou deux polarités négatives. Un syntagme nominal se présente lui comme une ressource disponible, c'est-à-dire porteur d'une polarité positive. Il va donc pouvoir venir satisfaire un des besoins porté par le verbe transitif ou, en d'autres termes, une polarité positive va pouvoir venir neutraliser une polarité négative.

La logique linéaire (Retoré, 2001) étant caractérisée comme une logique sensible aux ressources, c'est vers elle que nous nous sommes tournés pour ébaucher une première formalisation des Grammaires d'Interaction (Perrier, 2000a; Perrier, 2000b). Dans ce papier, pour définir formellement les Grammaires d'Interaction, nous adoptons un cadre plus familier puisqu'il s'agit de la théorie des modèles de la logique classique.

## 1 La forme des descriptions syntaxiques

Les objets de base des Grammaires d'Interaction sont les *descriptions syntaxiques*. Celles-ci sont construites à partir d'une signature composée d'un ensemble dénombrable  $\mathcal{N}$  de nœuds syntaxiques et d'une base de traits  $\mathcal{F}$ . Les nœuds syntaxiques représentent des syntagmes et les traits permettent d'exprimer les propriétés linguistiques de ces syntagmes. La base de traits  $\mathcal{F}$  est un ensemble fini de couples  $(F, D_F)$  formés d'un nom de trait  $F$  auquel est associé l'ensemble  $D_F$  de toutes les valeurs atomiques possibles qu'il peut prendre. Par exemple, si  $F = \text{genre}$ ,  $D_F = \{m, f\}$ .

L'originalité des Grammaires d'Interaction est que les traits y sont *polarisés*. Ils sont positifs, négatifs ou neutres selon qu'ils représentent une ressource disponible, une ressource attendue ou un élément d'information qui n'est pas une ressource consommable. Par exemple, un nom propre est porteur d'un trait positif *cat* avec la valeur *sn* qui exprime qu'il est une ressource disponible en tant que syntagme nominal. En même temps, il est porteur d'un trait négatif *fonct* avec une valeur indéterminée pour exprimer qu'il attend de recevoir une fonction syntaxique dans la phrase. Par contre, un verbe transitif possède dans sa représentation syntaxique un

## Descriptions d'arbres avec polarités : les Grammaires d'Interaction

nœud représentant son objet porteur d'un trait négatif *cat* avec la valeur *sn* pour indiquer qu'il attend un syntagme nominal. Ce nœud est pourvu en même temps d'un trait positif *fonct* avec la valeur *obj* pour indiquer que le verbe va fournir la fonction syntaxique *objet* à ce syntagme attendu. Les traits d'accord sont, quant à eux, toujours des traits neutres.

Définissons formellement les notions de *valeur de trait* et de *trait polarisé*.

**Définition 1.1** Si  $F$  est un nom de trait tiré de  $\mathcal{F}$ , une valeur du trait  $F$  est une partie non vide de  $D_F$ , éventuellement accompagnée d'un indice entier.

Un trait polarisé sur  $\mathcal{F}$  est un triplet formé d'un nom de trait  $F$ , d'une polarité qui est l'une des valeurs  $-1, 0, +1$ , et d'une valeur du trait  $F$ .

Selon la polarité  $-1, 0, +1$  d'un trait dont le nom est  $F$  et la valeur  $V$ , ce trait sera noté  $F \leftarrow V$ ,  $F = V$  ou  $F \rightarrow V$ . Si  $V$  est égale à  $D_F$ , elle est omise sinon elle est notée de façon extensive en indiquant ses éléments séparés par l'opérateur de disjonction  $\vee$ .  $V$  comporte un indice  $I$  quand on veut que cette valeur soit partagée par un autre trait comportant le même indice  $I$ .  $V$  est alors notée par  $\langle I \rangle$  suivie de la valeur effective.

Voici quelques exemples de traits polarisés:

$$cat \rightarrow sn, \quad fonct \leftarrow, \quad cat \leftarrow \langle 5 \rangle (sn \vee comp), \quad genre = m, \quad genre = \langle 2 \rangle$$

Venons en maintenant à la définition d'une *description syntaxique*.

**Définition 1.2** Une description syntaxique sur une signature  $(\mathcal{N}, \mathcal{F})$  est une formule logique classique  $D$  définie par la grammaire suivante:

$$D ::= \top \mid \perp \mid D \wedge D \mid N > N \mid N >^* N \mid N < N \mid N = N \mid N : T$$

où  $N$  représente un nœud quelconque de  $\mathcal{N}$  et  $T$  un trait polarisé quelconque sur  $\mathcal{F}$ .

$\top$  et  $\perp$  sont les constantes logiques *vrai* et *faux*. Les relations  $>$ ,  $>^*$  et  $<$  représentent des relations binaires de parenté immédiate, domination et précéence sur  $\mathcal{N}$ . Pour composer des descriptions, nous nous contentons d'utiliser la conjonction mais on peut enrichir la composition avec la disjonction et même la négation.

La figure 1 donne un exemple de description syntaxique présentée sous une forme logique à

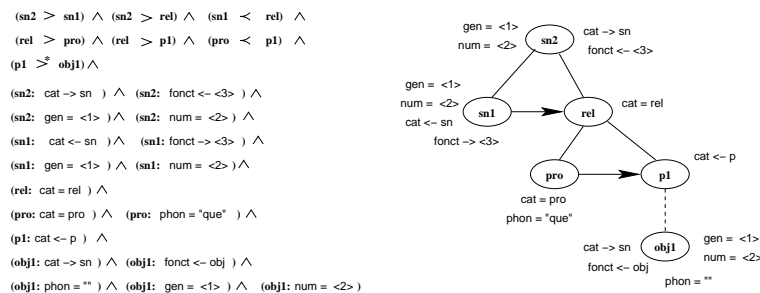


Figure 1: Description syntaxique représentant une entrée lexicale possible pour le pronom relatif *que*

gauche et graphique à droite. Les relations de parenté immédiate, domination et précéence sont représentées respectivement par des traits continus, discontinus et par des flèches.

Cette description exprime la configuration syntaxique dans laquelle doit s'inscrire le pronom relatif *que*. Les nœuds *sn1* et *p1* représentent le syntagme nominal et la proposition relative qui sont attendus respectivement à gauche et à droite du pronom. Le nœud *obj1* représente l'objet

qui est fourni par le pronom à la relative. Cet objet peut être complément du verbe qui est la tête de la relative mais il peut aussi être complément du verbe tête d'une complétive enchâssée plus ou moins profondément dans la relative comme dans l'expression *Marie que Jean pense que Pierre souhaite rencontrer*. C'est pourquoi la relation de domination entre *pl* et *objl* est non spécifiée.

## 2 Les arbres syntaxiques comme modèles minimaux et neutres de descriptions syntaxiques

Maintenant que l'on a campé les objets syntaxiques de base, les descriptions syntaxiques, on peut définir formellement une grammaire d'interaction<sup>1</sup>.

**Définition 2.1** Une grammaire d'interaction sur une signature  $(\mathcal{N}, \mathcal{F})$  est définie par son lexique, c'est-à-dire une fonction qui à tout mot d'un vocabulaire  $\mathcal{V}$  associe un ensemble fini de couples  $(D, A)$  où  $D$  est une description syntaxique sur  $(\mathcal{N}, \mathcal{F})$  et  $A$  est un nœud de  $D$  distingué comme ancre.

Pour une grammaire d'interaction donnée, il s'agit maintenant de définir le langage qu'elle engendre. Cette définition s'appuie sur la notion de modèle d'une description d'arbre, telle quelle est établie par (Rogers & Vijay-Shanker, 1992). Cette notion est particularisée par une condition de neutralité qui donne une sémantique à la notion de polarité.

**Définition 2.2** Un modèle minimal et neutre d'une description syntaxique  $D$  est un couple formé d'un arbre  $A$  et d'une interprétation  $I$ .

$A$  est ordonné localement<sup>2</sup> et ses nœuds sont étiquetés par des structures de traits de la forme  $(F : V)$  où  $F$  est un nom de trait issu de  $\mathcal{F}$  et  $V$  est un élément de  $D_F$ .

$I$  est une fonction de l'ensemble  $|D|$  des nœuds de  $D$  dans l'ensemble  $|A|$  des nœuds de  $A$  qui vérifie les propriétés suivantes:

1. Pour tout prédicat  $(N : T)$  qui compose  $D$ , la structure de traits de  $I(N)$  comprend un trait de même nom que  $T$  avec une valeur figurant dans  $T$ .
2. Si  $D$  comporte deux prédicats  $(N_1 : T_1)$  et  $(N_2 : T_2)$  où  $T_1$  et  $T_2$  sont des traits indexés par le même entier,  $I(N_1)$  et  $I(N_2)$  doivent avoir dans leurs structures de traits la même valeur pour les traits correspondants.
3. (condition de neutralité) Pour tout nœud  $N$  de  $A$  et pour tout nom de trait  $F$  présent dans l'étiquette de  $N$ , deux cas seulement sont possibles: soit il n'existe pas dans  $D$  de prédicat de la forme  $(N' : F \rightarrow V')$  ou  $(N' : F \leftarrow V')$  tel que  $I(N') = N$  soit il existe exactement un prédicat de la forme  $(N' : F \rightarrow V')$  tel que  $I(N') = N$  et un prédicat de la forme  $(N'' : F \leftarrow V'')$  tel que  $I(N'') = N$ .
4. Si  $D$  comporte une relation  $N_1 > N_2$ ,  $I(N_1)$  est le père de  $I(N_2)$  dans  $A$ .

<sup>1</sup>Nous avons choisi une présentation lexicalisée des Grammaires d'Interaction mais on peut très bien les définir de façon non lexicalisée en se donnant un ensemble de descriptions élémentaires.

<sup>2</sup>Un arbre localement ordonné est un arbre dans lequel chaque nœud a l'ensemble de ses fils qui est totalement ordonné. Cette relation d'ordre est étendue par héritage.

5. Si  $D$  comporte une relation  $N_1 >^* N_2$ ,  $I(N_1)$  est un ancêtre de  $I(N_2)$  ou égal à  $I(N_2)$  dans  $A$ .
6. Si  $D$  comporte une relation  $N_1 \prec N_2$ ,  $I(N_1)$  précède  $I(N_2)$  dans  $A$  pour l'ordre local qui lui est attaché.
7. Si  $D$  comporte une relation  $N_1 = N_2$ ,  $I(N_1)$  est égal à  $I(N_2)$ .
8. (condition de minimalité) Si  $N_1$  est le père d'un nœud  $N_2$  dans  $A$ , il existe au moins une relation  $N'_1 > N'_2$  dans  $D$  telle que  $N_1 = I(N'_1)$  et  $N_2 = I(N'_2)$

Une description syntaxique qui possède un modèle minimal et neutre est dite valide.

La notion de modèle minimal et neutre va nous permettre maintenant de caractériser les structures syntaxiques acceptables d'une grammaire d'interaction ainsi que les phrases grammaticales.

**Définition 2.3** Soit une grammaire d'interaction  $G$  définie sur un vocabulaire  $\mathcal{V}$  par son lexique  $L$ . Une phrase  $w_1 \dots w_n$  de  $\mathcal{V}^*$  est grammaticale si on peut trouver dans chaque entrée lexicale  $L(w_i)$  un élément  $(D_i, A_i)$  tel que la description  $D_1 \wedge \dots \wedge D_n \wedge (A_1 \prec A_2) \wedge \dots \wedge (A_{n-1} \prec A_n) \wedge (N >^* \text{Root}(D_1)) \wedge \dots \wedge (N >^* \text{Root}(D_n)) \wedge (N : \text{cat} \leftarrow p)$ , où  $N$  est un nom de nœud ne figurant pas dans  $D_1, \dots, D_n$ , soit valide<sup>3</sup>.

Si c'est le cas, tout modèle de cette description représente un arbre syntaxique de la phrase  $w_1 \dots w_n$ .

La description qui est le centre de cette définition peut être découpée en 3 parties:

- $D_1 \wedge \dots \wedge D_n$  représente les ressources lexicales utilisées pour l'analyse;
- $(A_1 \prec A_2) \wedge \dots \wedge (A_{n-1} \prec A_n)$  représente l'ordre de précédence entre les ancres qui correspond à l'ordre des mots dans la phrase.
- Enfin  $(N >^* \text{Root}(D_1)) \wedge \dots \wedge (N >^* \text{Root}(D_n)) \wedge (N : \text{cat} \leftarrow p)$  permet de caractériser le syntagme  $N$  qui représente la phrase à analyser. Il est la racine de la description générale, ce que nous exprimons en introduisant pour chaque description  $D_i$  une racine  $\text{Root}(D_i)$ . Par définition,  $\text{Root}(D_i)$  est un nœud de  $D_i$  qui n'est pas dominé et qui n'a pas de père. Nous supposons l'unicité d'un tel nœud mais cette hypothèse n'est pas vraiment indispensable à la définition 2.3. Le prédicat  $(N : \text{cat} \leftarrow p)$  indique que le nœud  $N$  attend un syntagme de type *proposition*.

Illustrons la définition 2.3 par l'analyse de la phrase *Marie que Jean a vue dort*. La figure 2 présente la description syntaxique  $D_0$  fournie par un lexique pour analyser cette phrase. On y retrouve l'entrée lexicale du pronom relatif *que* présentée dans la figure 1. Pour simplifier la description, nous supposons que *a vue* forme un seul mot. Dans l'entrée lexicale de celui-ci, on peut prévoir que l'objet du verbe est placé avant étant donné le genre du participe passé. Cela se traduit dans la description par le fait qu'on trouve à la place canonique de l'objet une trace (sa forme phonologique est vide) représentée par le nœud *obj2*. La figure 3 représente l'unique modèle de la description  $D_0$  qui constitue l'unique arbre syntaxique de la phrase *Marie que Jean a vue dort* et qui démontre sa grammaticalité.

<sup>3</sup>On suppose que les noms de nœuds et les indices de traits présents dans les  $D_i$  sont locaux à ces descriptions, sinon on procède à un renommage.

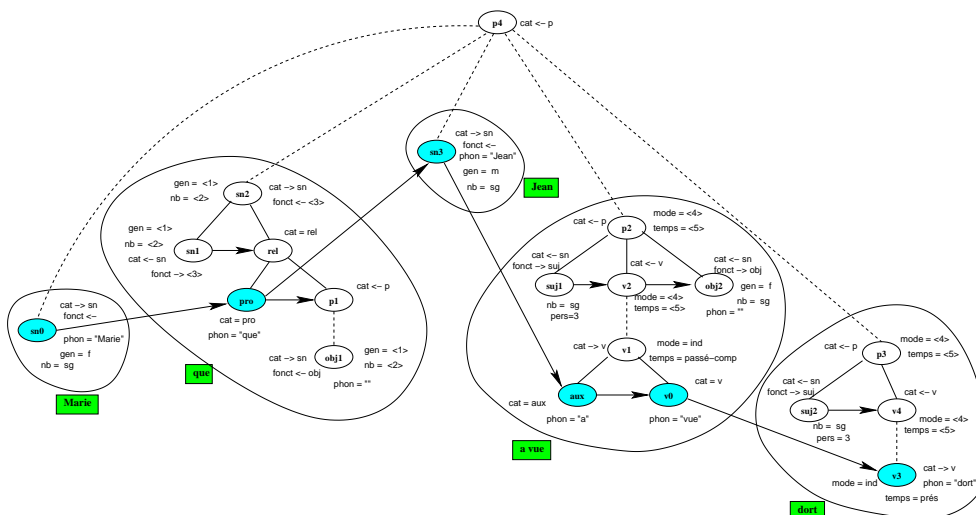


Figure 2: Description syntaxique  $D_0$  utilisée pour l'analyse de la phrase *Marie que Jean a vue dort*

### 3 La composition syntaxique comme mécanisme de neutralisation des traits polarisés

La définition 2.2, si elle fournit un moyen de reconnaître un modèle minimal et neutre d'une description syntaxique, ne décrit pas de mécanisme qui permette de construire un tel modèle. Le lexique fournit un ensemble de descriptions syntaxiques  $D_1, \dots, D_n$  correspondant aux différents mots d'une phrase à analyser. Celles-ci sont rassemblées en une seule description  $D$  dont il s'agit de trouver un modèle. De fait, les descriptions  $D_1, \dots, D_n$  ne sont que juxtaposées au sein de  $D$ , ce qui laisse beaucoup d'indéterminisme quant à la façon de trouver un modèle de  $D$ . Un moyen de réduire cet indéterminisme est de définir un mécanisme de composition des descriptions  $D_1, \dots, D_n$  qui permette de préciser pas à pas les contours d'un éventuel modèle pour aboutir à une description complètement spécifiée d'un arbre syntaxique telle que ce dernier s'en déduise immédiatement. Il faut bien entendu que ce mécanisme garantisse les propriétés suivantes : si  $D$  n'est pas valide, toute description qui est obtenue à partir de  $D$  par ce mécanisme doit être non valide et si  $D$  est valide, il doit exister une façon d'appliquer ce mécanisme qui aboutisse à une description valide.

Tout formalisme qui s'exprime à l'aide de descriptions d'arbres fournit en même temps un tel mécanisme de composition syntaxique. Dans les TAG par exemple, il est constitué des deux opérations de substitution et d'adjonction, la première pouvant s'exprimer à l'aide de la seconde.

Pour les Grammaires d'Interaction, nous allons nous appuyer sur l'exigence de neutralité des modèles pour définir l'unique opération qui va servir à réaliser la composition syntaxique et que nous appellerons *neutralisation élémentaire*.

**Définition 3.1** Si une description syntaxique  $D_1$  comporte deux prédicats  $(N_1 : F \rightarrow V_1)$  et  $(N_2 : F \leftarrow V_2)$ , la description syntaxique  $D_2$  qui résulte de la neutralisation élémentaire de ces deux traits est définie de la façon suivante: si  $D_1$  comporte deux prédicats  $(N_1 : F' \text{ pol}_1 V_1)$  et  $(N_2 : F' \text{ pol}_2 V_2)$  et si  $|\text{pol}_1 + \text{pol}_2| > 1$  ou si  $V_1 \cap V_2 = \emptyset$ , alors  $D_2 = \perp$  sinon  $D_2$  s'obtient en substituant  $N_1$  à  $N_2$  dans  $D_1$  et en remplaçant tous les prédicats de la forme  $(N_1 : F' \text{ pol}_1 V_1)$  et

## Descriptions d'arbres avec polarités : les Grammaires d'Interaction

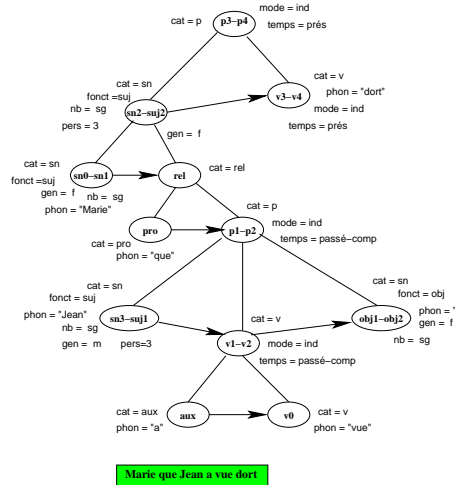


Figure 3: Arbresyntaxique de la phrase *Marie que Jean a vue dort* modèle de la description  $D_0$

( $N_2 : F'pol_2V_2$ ) par un unique prédicat ( $N_1 : F'(pol_1 + pol_2)V_1 \cap V_2$ ) et si  $V_1$  ou  $V_2$  comporte des indices, on les conserve en les identifiant.

On note alors:  $D_1 \xrightarrow{N_1 \stackrel{F}{=} N_2} D_2$ .

Une neutralisation élémentaire est donc une opération qui consiste à identifier deux nœuds porteurs de traits duaux. Donnons en un exemple en reprenant la description  $D_0$  de la figure 2. En effectuant une neutralisation élémentaire du trait  $cat \rightarrow sn$  du nœud  $sn0$  par le trait  $cat \leftarrow sn$  du nœud  $sn1$ , on obtient la description  $D_1$  dont la figure 4 présente la partie qui a été modifiée<sup>4</sup>.

A la suite d'une neutralisation élémentaire, il est utile de simplifier la description obtenue,

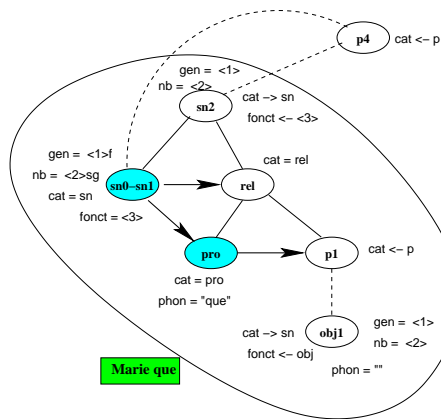


Figure 4: Vue partielle de la description syntaxique  $D_1$  résultant d'une opération de neutralisation élémentaire sur  $D_0$

c'est-à-dire de la remplacer par une description plus simple équivalente logiquement dans la théorie des arbres localement ordonnés. Ainsi dans la description  $D_1$  on peut supprimer toutes les relations entre des nœuds qui sont redondantes. Cela s'applique à la relation de précédence entre le nœud  $sn0$ - $sn1$  et le nœud  $pro$  et à la relation de domination entre le nœud  $p4$  et le nœud

<sup>4</sup>Pour des raisons mnémotechniques, on concatène le nom des deux nœuds qu'on identifie même si on s'écarte légèrement du cadre syntaxique fixé par la définition 3.1.

*sn0-sn1*. Simplifier une description, c'est aussi mettre à jour les traits indexés. Ainsi dans  $D_1$ , on instancie les traits *gen* et *nb* des nœuds *sn2* et *obj1* par les valeurs *f* et *sg*.

Parfois, la simplification est plus subtile. Si par exemple, deux nœuds ont un fils commun, on peut les identifier car dans un arbre, un nœud a au plus un père. Par contre, si deux nœuds en domine un troisième, on peut dire qu'ils sont en position de domination l'un par rapport à l'autre mais savoir dans quel sens.

L'opération de neutralisation élémentaire comme opération de base du processus de composition syntaxique trouve sa justification dans la proposition suivante dont la démonstration est immédiate:

**Proposition 3.1** *Si une description  $D_1$  qui comporte au moins un trait non neutre est valide, il existe au moins une neutralisation élémentaire sur  $D_1$  qui produise une description valide. Si une description  $D_1$  n'est pas valide, toute neutralisation élémentaire sur  $D_1$  produit une description non valide.*

Cette proposition fonde un mécanisme de construction d'un arbre syntaxique modèle d'une description donnée qui consiste à itérer l'opération de neutralisation élémentaire. Cette itération est nécessairement finie et peut s'arrêter de 3 façons différentes:

1. Il reste encore des traits à neutraliser, ce qui signifie que les traits positifs n'équilibrent pas les traits négatifs et la description de départ n'est pas valide.
2. On aboutit à une description porteuse d'une contradiction et on ne peut rien conclure. On peut avoir fait un mauvais choix quant aux neutralisations précédentes qui ont été effectuées et il faut faire un retour arrière pour essayer de nouvelles neutralisations.
3. On aboutit à une description neutre. S'il existe un modèle d'une telle description, la description de départ est valide sinon il faut faire un retour arrière pour essayer de nouvelles neutralisations.

Dans le dernier cas, on se trouve dans la situation de chercher un modèle minimal d'une description d'arbre d'où les polarités ont été éliminées. Si on impose la contrainte supplémentaire que les relations de domination doivent être réalisées directement ou indirectement par des neutralisations, les seules descriptions auxquelles on aboutit dans ce cas sont celles d'arbres syntaxiques où l'indétermination qui subsiste est celle concernant l'ordre local entre les fils des différents nœuds.

Le calcul qui vient d'être défini n'est bien entendu pas confluent. L'ambiguïté syntaxique fait qu'une même description peut mener à plusieurs arbres syntaxiques différents. Ce n'est pas le cas de la description  $D_0$  qui se réduit de la façon suivante:

$$D_0 \xrightarrow{sn0 \stackrel{cat}{\equiv} sn1} D_1 \xrightarrow{sn3 \stackrel{cat}{\equiv} suj1} D_2 \xrightarrow{p1 \stackrel{cat}{\equiv} p2} D_3 \xrightarrow{obj1 \stackrel{fnct}{\equiv} obj2} D_4 \xrightarrow{sn2 \stackrel{cat}{\equiv} suj2} D_5 \xrightarrow{p4 \stackrel{cat}{\equiv} p3} D_6 \xrightarrow{v1 \stackrel{cat}{\equiv} v2} D_7 \xrightarrow{v3 \stackrel{cat}{\equiv} v4} D_8$$

L'ordre des neutralisations est libre mais on aboutit à un seul résultat, la description  $D_8$  qui est une spécification complète de l'arbre syntaxique présentée dans la figure 3.

## 4 Puissance d'expression de la composition syntaxique

La plupart des formalismes grammaticaux basés sur les arbres utilisent la substitution comme opération élémentaire dans la composition syntaxique. C'est le cas des grammaires algébriques



lexicalisées, des grammaires de dépendances lexicalisées (Nasr, 1995) mais c'est aussi le cas de formalismes utilisant des mécanismes plus complexes comme les TAG ou les Grammaires de Descriptions d'Arbres (Rambow *et al.*, 2001). Ainsi l'adjonction des TAG peut être présentée comme une double substitution.

La substitution est une opération qui présente des limites dans le pouvoir d'expression qu'elle permet. Du fait qu'elle consiste à accrocher un arbre à un nœud particulier d'un autre arbre, elle ne permet d'imposer des contraintes que sur les propriétés du point de jonction attendu mais en aucun cas sur le contexte syntaxique de ce dernier.

La linguistique nous montre que ceci est parfois nécessaire. De ce point de vue, l'opération de neutralisation élémentaire des Grammaires d'Interaction permet d'aller au-delà de l'action purement localisée de substitution. Les deux nœuds qui sont identifiés dans une neutralisation élémentaire ne sont pas forcément une racine et une feuille mais ils peuvent occuper n'importe quelle position dans leurs arbres respectifs et leur identification peut s'accompagner d'une *superposition partielle de ces arbres* permettant d'exprimer des contraintes sur les contextes syntaxiques des nœuds qui sont identifiés.

Prenons un exemple qui se situe dans le prolongement de l'analyse de la phrase *Marie que Jean a vue dort*. Un des effets du pronom relatif *que* est qu'en réalisant l'objet du verbe de la relative, il permet de relâcher l'ordre canonique sujet-verbe et on peut dire *Marie qu'a vue Jean dort*. Il est possible d'exprimer cela en considérant les nouvelles entrées lexicales pour *que* et *a vue* qui sont présentées par la figure 5. Le nœud *obj1* apparaît maintenant comme complément d'objet

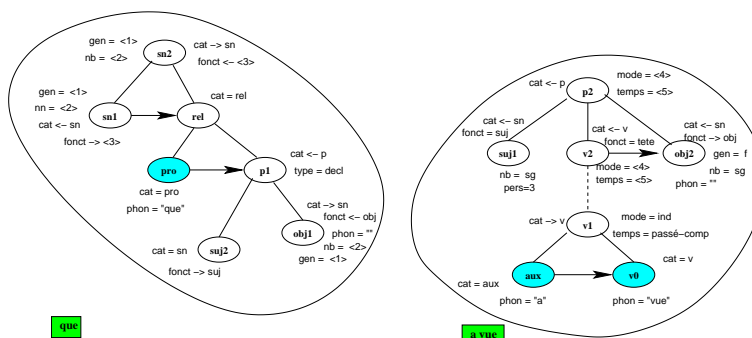


Figure 5: Nouvelles entrées lexicales pour *que* et *a vue*

direct du verbe de la relative car le pronom *que* commande uniquement l'inversion sujet verbe dans la relative et pas dans une complétive enchâssée. Le sujet de la relative est maintenant représenté explicitement par le nœud *suj2*. Le trait *fonct* → *su*j qu'il porte indique que cette fonction est réalisée quelle que soit la position de *suj2* par rapport aux autres syntagmes qui composent la relative. Dans l'entrée lexicale de *a vue*, il n'y a pas plus de relation de précedence entre les nœuds *suj1* et *v2* et le trait *fonct* avec la valeur *su*j est maintenant neutre ce qui signifie que ce n'est plus l'ordre des syntagmes qui est en mesure de réaliser la fonction sujet; celle-ci doit être réalisée de façon externe. Ces entrées vont permettre d'analyser comme correctes les deux phrases *Marie que Jean a vue dort* et *Marie qu'a vue Jean dort*.

Dans un formalisme qui repose sur la substitution d'arbres comme opération de base de composition syntaxique, le phénomène d'inversion du sujet dans une relative ne peut pas être exprimé de cette façon car la substitution ne permet la superposition d'arbres. Il ne peut pas être exprimé dans l'entrée lexicale du pronom relatif mais il doit être rattaché à l'entrée lexicale du verbe de la relative (c'est ce qui se passe par exemple en TAG), ce qui a tendance à alourdir le lexique.

## Conclusion

Dans ce papier, nous nous sommes attachés à définir formellement les Grammaires d'Interaction et nous avons donné une idée du pouvoir d'expression du formalisme. Quand on sait que pouvoir d'expression ne fait pas bon ménage avec efficacité computationnelle, on peut se demander s'il est possible de concevoir des algorithmes d'analyse efficaces pour les Grammaires d'Interaction. Cette question ne peut pas être traitée sérieusement dans le cadre de ce papier mais le système de polarités qui fait l'originalité des Grammaires d'Interaction est un atout pour cela. Ainsi, par exemple, on peut concevoir une analyse incrémentale guidée par une règle de maximisation des neutralisations élémentaires et de limitation des polarités non neutralisées. Un dernier atout des Grammaires d'Interaction que nous n'avons pas eu la place d'évoquer est celui de leur organisation modulaire et hiérarchisée qui doit permettre de construire de telles grammaires avec une large couverture.

## References

- BLACHE P. (1999). *"Contraintes et théories linguistiques : des Grammaires d'Unification aux Grammaires de Propriétés"*. Habilitation à diriger des recherches, Université Paris 7.
- CANDITO M.-H. (1999). *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien*. Thèse de doctorat, Université Paris 7.
- EGG M., NIEHREN J., RUHRBERG P. & XU F. (1998). Constraints over lambda structures in semantic underspecification. In *COLING/ACL'98, Aug 98, Montreal, Quebec, Canada*.
- KALLMEYER L. (1999). *Tree Description Grammars and Underspecified Representations*. PhD thesis, Universität Tübingen.
- MUSKENS R. (2001). Talking about trees and truth-conditions. *Journal of Logic, Language, and Information*, **10**(4), 417–455.
- NASR A. (1995). A formalism and a parser for Lexicalised Dependency Grammars. In *4th Int. Workshop on Parsing Technologies*: State University of NY Press.
- PERRIER G. (2000a). From intuitionistic proof nets to interaction grammars. In *TAG+5, Paris*, p. 177–184.
- PERRIER G. (2000b). Interaction grammars. In *CoLing '2000, Sarrebrücken*, p. 600–606.
- RAMBOW O., VIJAY-SHANKER K. & WEIR D. (2001). D-tree Substitution Grammars. *Computational Linguistics*, **27**(1), 87–121.
- RETORÉ C. (2001). Systèmes déductifs et traitement des langues: un panorama des grammaires catégorielles. *Technique et Science Informatiques*, **20**(3), 301–336.
- ROGERS J. & VIJAY-SHANKER K. (1992). Reasoning with descriptions of trees. In *30th Annual Meeting of the Association for Computational Linguistics*, p. 72–80.
- VIJAY-SHANKER K. (1992). Using Description of Trees in a Tree Adjoining Grammar. *Computational Linguistics*, **18**(4), 481–517.