

# Pour une représentation décentralisée de l'information syntaxique

Philippe Blache  
Aix-Marseille Universités & CNRS  
Laboratoire Parole et Langage  
pb@lpl.univ-aix.fr

**Résumé.** L'évolution récente des théories linguistiques conduit à une représentation différenciée des types d'information syntaxiques (par exemple la dominance et la précédence). Nous proposons dans cet article de systématiser cette approche pour conduire à une véritable représentation *décentralisée* de la syntaxe, nous permettant d'identifier un certain nombre de propriétés de base, devant être d'une façon ou d'une autre présente dans tout formalisme syntaxique.

**Abstract.** Modern linguistic theories tends to represent separately different kinds of syntactic information, (such as dominance or precedence). We propose in this paper to systematize this approach such as to obtain a *decentralized* representation, making it possible to identify a set of fundamental syntactic properties, to be represented in all syntactic formalism.

**Mots-clés :** Formalismes syntaxiques, propriétés syntaxiques, Model-Theoretic Syntax.

**Keywords:** Syntactic formalisms, basic syntactic properties, Model-Theoretic Syntax.

## 1 Introduction

Une des grandes idées proposées par GPSG ((Gazdar & al. 85)) consistait, dans le cadre des grammaires syntagmatiques, à représenter de façon distincte les informations hiérarchiques des informations de précédence linéaire. Derrière ce qui à première vue représente un intérêt en termes de représentation de l'information réside en fait une avancée majeure pour les théories linguistiques. Nous proposons dans cet article de pousser cette logique de différenciation de l'information syntaxique. Cette démarche nous permet de proposer une représentation beaucoup plus précise de ces informations et des interactions existant entre elles. Elle permet de plus d'ouvrir une alternative à une vision purement générative de la grammaire, sans en rejeter ce qui en fait l'intérêt (par exemple en termes de définition de classes d'objets). Cette vision, que nous appelons *décentralisée*, de l'information syntaxique repose sur la remise en question du statut central occupé dans tous les formalismes par la notion de dominance. Nous en faisons une information au même niveau que les autres, permettant ainsi la représentation de phénomènes complexes.

Cet article, plutôt que de présenter un formalisme particulier, précise les différents types d'information syntaxique qu'un formalisme doit représenter, si possible de façon explicite. Après avoir situé les besoins auxquels doivent répondre les formalismes de haut niveau, nous proposons la description d'un ensemble de phénomènes illustrant les limites rencontrés par une repré-

sensation reposant prioritairement sur la dominance. Nous proposons donc une représentation totalement décentralisée, sous la forme d'un ensemble de propriétés devant être représentées en syntaxe.

## 2 La représentation de l'information syntaxique

### 2.1 L'évolution des théories

Le formalisme DI/PL propose d'explicitier le statut d'une information jusque là implicite : l'ordre linéaire<sup>1</sup>. Une représentation arborescente simple ne permet en effet pas de rendre compte de ce type d'information. Par ailleurs, GPSG propose de systématiser la représentation d'une partie des informations syntaxiques sous forme de traits associés aux catégories. Cette représentation de l'information permet l'introduction d'un nouveau type de relations comme les restrictions de cooccurrence de traits ou encore l'introduction de valeurs de traits par défaut. Il s'agit de contraintes portant directement sur la forme voire le contexte de réalisation d'une catégorie. Les informations concernant la structure hiérarchique sont ainsi complétées par d'autres types de contraintes permettant d'en contrôler la construction tout en précisant son contenu.

Les informations ainsi représentées sous forme de contraintes (cf. (Rogers97)) permettent de poser différemment la question du statut de la grammaire et sa relation par rapport au langage. Les approches génératives classiques proposent en effet de définir cette relation en termes de dérivation. La grammaire constitue ainsi un mécanisme d'énumération permettant de générer le langage. C'est ce que Pullum appelle la syntaxe générative énumérative (ou GES, cf. (Pullum & Scholz 01), (Huddleston & Pullum 02)). Une conception différente consiste à décrire la syntaxe dans la perspective de la théorie des modèles (on parle ainsi de *Model Theoretic Syntax*, ou MTS). Cette approche, présentée notamment dans (Blackburn & al. 93) puis développée dans (Backofen & al. 95), ou (Cornell & Rogers 00), consiste à concevoir l'analyse syntaxique comme une recherche de modèle dans un domaine spécifiée. Il s'agit d'une vision beaucoup plus souple du processus d'analyse, consistant, pour une affectation donnée (i.e. un ensemble de catégories) à vérifier les propriétés qu'elle satisfait. Une telle approche de l'analyse reposant sur les contraintes est présente de façon plus ou moins directe dans des théories comme HPSG (cf. (Sag al. 03)). Les théories basées sur les modèles n'ont donc pas pour objectif de construire une structure syntaxique en même temps qu'on vérifie la grammaticalité de l'input, mais plutôt de rechercher les propriétés qui sont vérifiées par cet énoncé. Les questions de la construction de la structure et de sa forme deviennent du même coup secondaire. En revanche, celles concernant l'identification des contraintes, et donc du type d'information à représenter, en plus de la dominance, devient centrale.

### 2.2 Les besoins

Les théories linguistiques, et donc les formalismes qui les représentent, doivent permettre rendre compte d'un certain nombre de phénomènes.

- **Non canonicité de l'entrée** : l'analyse de la langue parlée, mais également de matériel "tout

---

<sup>1</sup>Cette information n'est souvent prise en compte de façon marginale, y compris dans les théories modernes (par exemple HPSG, cf. (Sag al. 03))

venant”, nécessite de rendre compte de phénomènes particuliers que l’on peut ranger globalement sous la rubrique “entrées non canoniques”. Il s’agit d’entrées traditionnellement considérées comme mal formées pour différentes raisons : disfluences, structure incomplète, violation de règles, etc. Cet objectif est fondamental dans la plupart des approches modernes, il est même explicite dans certaines théories comme les grammaires de construction (cf. (Fillmore98)) ou l’optimalité (cf. (Prince93)).

- **Domaines** : l’information linguistique est répartie dans plusieurs domaines. Le traitement d’un message (sa compréhension) est le résultat de la convergence des parties d’information provenant des différents domaines (cf. (Blache02)). Il y a deux conséquences importantes à ce phénomène. Tout d’abord, les informations présentes dans chacun des domaines peuvent donc être partielles. Concrètement, chaque domaine n’est donc pas nécessairement porteur d’information pertinente pour la totalité de l’entrée. Pour ce qui concerne la syntaxe en particulier, cela signifie que toutes les unités identifiées n’entrent pas nécessairement dans une relation syntaxique. En d’autres termes, il n’est pas toujours possible (ni indispensable) de construire une structure syntaxique couvrant totalement l’entrée.

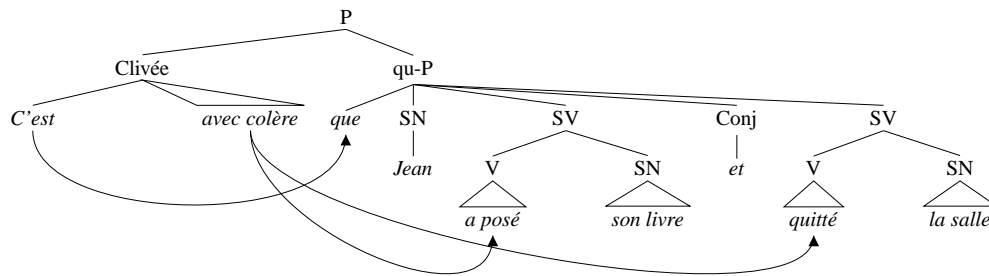
Par ailleurs, il est nécessaire de préciser comment les différents domaines interagissent. Par exemple, la vision compositionnelle classique de l’interaction syntaxe-sémantique consiste à calculer la représentation sémantique à partir de la structure syntaxique. De même, les relations prosodie/syntaxe sont généralement décrite en termes de superposition de structures (constituants prosodiques matchés aux constituants syntaxiques). Ces mécanisme posent donc problème dans le cas de structures partielles.

- **Motivation linguistique** : les informations portées dans un formalisme de haut niveau doivent être linguistiquement motivées et distinctes des éventuelles informations opérationnelles relevant de la mise en œuvre du formalisme plus que de la représentation de l’information linguistique. Par exemple, les mécanismes décrivant la propagation des traits ou contrôlant leur instanciation doivent être représentés séparément des propriétés linguistiques à proprement parler.

### 3 Les limites d’une représentation hiérarchique

Les approches syntagmatiques reposent toutes sur une hiérarchisation stricte de l’information. Le domaine décrit par les grammaires est donc celui des arbres. Les évolutions récentes, y compris celles remettant en cause les approches génératives strictes, consistent à adapter ce mode de représentation en le complétant de différentes manières et notamment en associant aux arbres des contraintes plus ou moins complexes, ce qui permet aux arbres d’avoir un degré de généralité plus ou moins grand : ensembles d’arbres locaux associés à des contraintes en DCG, ou à l’opposé schémas d’arbres soumis à des principes en HPSG. La notion d’arbre reste donc au cœur des représentations. Cela signifie (y compris pour les approches génériques comme HPSG) qu’il est nécessaire de construire un arbre pour en vérifier les contraintes. Les contraintes correspondant aux propriétés syntaxiques des objets analysés, cela signifie que la description d’une entrée repose sur la capacité de construire un arbre. Or, de nombreux exemples montrent les limites de ce type de représentation. Nous en relevons quelques uns.

- Les **dépendances à distance** : les constructions clivées, comme d’une façon plus générale les phénomènes d’extraposition, sont analysées en deux parties d’une même structure générale comme décrit dans l’exemple suivant :

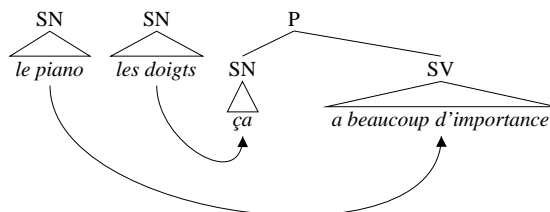


Cette représentation présente plusieurs lacunes, à commencer par la factorisation de l'auxiliaire dans la coordination, mais également concernant les relations entre l'élément clivé, son verbe recteur, ainsi que le pronom introduisant la *qu-phrase*. Ces relations sont selon les formalismes représentés sous la forme de traits se propageant à travers la structure (par exemple les traits *slash* en GPSG ou HPSG). Cela signifie que plusieurs types d'informations doivent être représentés : le découpage en unités, leur composition (informations représentées par l'arbre), mais également les dépendances entre ces unités (représentées dans la figure précédente par des flèches).

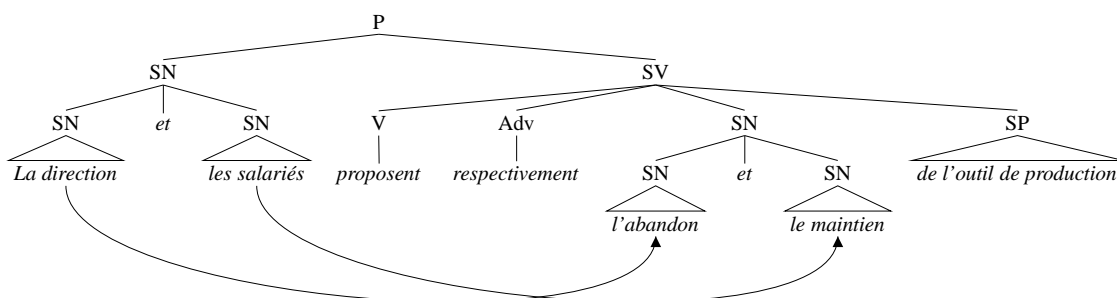
- **Extrapositions complexes** : plusieurs extrapositions peuvent se cumuler. L'exemple suivant<sup>2</sup> présente le cas d'une topicalisation doublée d'une dislocation :

(1) *le piano les doigts ça a beaucoup d'importance*

Il devient ici difficile (en tous cas peu pertinent) de proposer une structure arborescente unique. Il est en effet préférable de considérer qu'au niveau syntaxique, nous avons trois éléments dont les relations ne sont pas exprimées en termes de constituance, mais de dépendance.



- **Dépendances croisées** : ce type de phénomène intervient de façon systématique dans certaines langues. On le retrouve également souvent dans des constructions complexes telle la distribution sur des facteurs conjoints comme dans l'exemple suivant :



Cette construction illustre la nécessité d'indiquer des dépendances croisées construites par un verbe unique. Il est ici nécessaire de représenter à la fois les fonctions syntaxiques, en même temps que les relations existants entre les sujets et les objets.

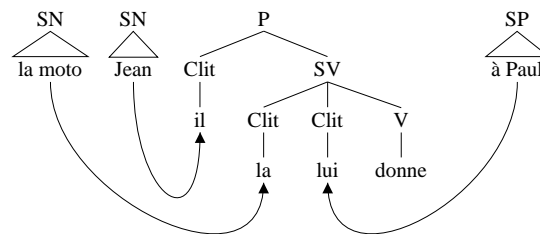
<sup>2</sup>Cet exemple, donné par José Deulofeu, est construit, mais censé relever de la langue parlée.

- **Phénomènes paradigmatiques** : ces phénomènes fréquents à l'oral, mais également présents à l'écrit, consistent à réaliser plusieurs fois une même position de la structure syntaxique (cf. (Blanche-Benveniste84)). L'exemple suivant illustre à l'aide d'une grille un phénomène de ce type, appelé "entassement paradigmatique" (détaillé dans le cadre des Grammaires de Propriétés dans (Guénot06)) :

	les Anglais	qui ont quand même beaucoup d'humour	
euh	les journaux anglais		
enfin	les médias britanniques		ont fini par me répondre

Dans le cas, le noyau du SN sujet est réalisée plusieurs fois, il s'agit d'un phénomène de reprise dans lequel chaque nouvel élément remplace sémantiquement le précédent. Cet exemple présente en outre la possibilité dans ce type de construction de factoriser ce noyau qui sera modifié par la même relative (qui elle n'est réalisée qu'une fois). Il est clair qu'une structure arborescente n'est pas adaptée à la représentation de ce type de phénomène.

Un autre cas de ce type de phénomène concerne ce que (Blanche-Benveniste84) nomme le double marquage. Cela concerne typiquement les constructions disloquées dans lesquelles un élément référentiel est repris par un clitique :



En conclusion, les approches génératives classiques, en s'appuyant sur une représentation arborescente, accordent un statut particulier aux informations de dominance. Celles-ci doivent être vérifiées en préalable et les conditions de bonne formation sont exprimées en termes de contraintes sur les arbres. Il s'agit d'une limitation importante pour plusieurs raisons. Tout d'abord, nous avons vu qu'il n'est pas toujours possible ni souhaitable de représenter une structure syntaxique sous forme d'arbre. Par ailleurs, les exemples précédents montrent que les relations de dominance constituent une (petite) partie de l'information syntaxique. Celle-ci est complexe, composée d'informations différentes (que nous décrivons dans la section suivante) qui doivent être considérées comme également importantes, en tous cas de même niveau. Une telle conception décentralisée de l'information permet de rendre compte de constructions particulières du type de celles décrites ici. Notre proposition revient à dire qu'une structure n'est pas nécessairement *connexe* ou plus précisément, qu'elle n'est pas nécessairement connexe du point de vue de la dominance. En revanche, elle l'est si toutes les informations sont considérées comme des relations, au même titre et au même niveau que la dominance : les différentes unités de la structure décrivant un énoncé sont ainsi connectées par des relations variées.

## 4 Une représentation décentralisée

Plusieurs relations peuvent être proposées pour décrire l'information syntaxique. Nous cherchons ici à étendre la démarche initiée en GPSG consistant à représenter de façon séparée les

différents types d'information. Cette section recense un ensemble qui nous paraît minimal d'informations qui doivent être représentées, quelque soit le formalisme. L'objectif est de rendre explicite ces informations comme GPSG a rendu explicites les relations d'ordre. Notre position est que chacune de ces propriétés est porteuse d'une partie de l'information, quelquefois de façon redondante. C'est l'interaction (ou la simple addition) de ces propriétés qui permet de fournir une description syntaxique précise d'un énoncé.

Les deux premières informations identifiées portent donc l'une sur la constituance, l'autre sur l'ordre.

- **Constituance** : une représentation en termes de constituants (correspondant donc dans une représentation arborescente à une information de dominance) est indispensable pour la définition des unités en termes syntagmatiques. Ce type de représentation reste aujourd'hui largement dominant dans les formalismes linguistiques. Elle permet tout d'abord d'exprimer des régularités en termes de substitution : des classes d'objets peuvent ainsi être décrites, partageant des propriétés communes. De plus, ce type de représentation permet également de donner des indications sur la fonction syntaxique en termes topologiques, par rapport à la position de l'élément relativement à d'autres.

- **Précédence linéaire** : GPSG a proposé de décrire le contrôle de l'ordre linéaire sous forme de déclarations. Il s'agit de contraintes contrôlant l'ordre entre éléments constituants d'une même unité, s'ils sont réalisés. Il est nécessaire de décrire de façon explicite ces paramètres. Certaines approches proposent de généraliser la précédence en l'étendant à des ensembles de catégories, permettant ainsi d'implanter des informations de type contextuel. Par ailleurs, il est utile d'étendre cette relation aux classes abstraites de catégories.

Nous trouvons également en GPSG d'autres types d'informations, qui ne sont pas représentées au même niveau que les deux précédentes. Il s'agit notamment de la restriction de cooccurrence de traits, permettant d'exclure la réalisation conjointe de certains types ou sous-types de catégories. Nous proposons de systématiser ce type d'information en représentant explicitement les relations de nécessité et d'impossibilité de cooccurrence.

- **Cooccurrence** : la relation de complémentation est typique de ce type de contrainte. La réalisation d'une catégorie particulière (souvent la tête) entraîne l'obligation de réaliser conjointement d'autres unités (les compléments). Le même type de relation existe également entre un spécifieur et la tête. Mais, au moins théoriquement, rien ne devrait empêcher la possibilité d'imposer la cooccurrence entre deux éléments, sans relation de réaction, indépendamment de la tête.

- (2) a. *Je me suis dit.*  
 b. *Je me le suis dit.*  
 c. *\*Je le suis dit.*

Il existe dans ce cas une contrainte de cooccurrence directement entre les compléments, sans intervention de la tête verbale.

- **Exclusion** : A l'inverse de la cooccurrence, la réalisation de certaines catégories peut entraîner l'impossibilité d'en réaliser d'autres. Comme pour la cooccurrence, ce type de contrainte peut exister entre la tête et d'autres constituants (par exemple, le nom propre ne permet pas la construction d'un déterminant). Mais ici aussi ce type de relation peut exister en dehors de la tête. Ce cas est illustré dans l'exemple suivant :

- (3) a. *Les trois livres.*  
 b. *Plusieurs livres.*  
 c. *\*Plusieurs trois livres.*

Il existe donc une contrainte particulière, indépendante de la tête nominale, entre le déterminant et l'adjectif. Celle-ci doit donc être représentée directement entre les constituants concernés.

- **Tête** : il est nécessaire dans une représentation syntagmatique d'identifier la tête d'une unité parmi l'ensemble de ses constituants<sup>3</sup>. Les informations de sous-catégorisation sont bien entendu importantes. Dans le type de représentation proposé ici, elles sont cependant portées par les contraintes de cooccurrence et instanciées au niveau du lexique. L'identification de la tête est cependant importante pour plusieurs raisons. Il s'agit en effet d'une catégorie obligatoirement réalisée dans une unité (éventuellement de façon isolée). Par ailleurs, elle permet de transmettre un certain nombre de caractéristiques à l'unité dont elle est la tête.

- **Unicité** : dans de nombreux cas, certaines catégories dans une unité ne peuvent pas être répétées. Cela concerne non seulement la tête (de façon systématique), mais également en fonction des unités d'autres constituants. Par exemple, un déterminant ne peut être répété dans un syntagme nominal. Il n'y a pas de régularité ni de prédictibilité particulière pour ce phénomène. Il est donc nécessaire de le représenter de façon explicite.

- **Dépendance** : Dans les grammaires de dépendance, cette relation est à la fois porteuse d'informations syntaxique et sémantique. Elle est de ce point de vue à la fois complémentaire et pour partie redondante avec certaines des informations exprimées précédemment. Nous proposons de représenter cette information, en particulier pour le rôle qu'elle joue dans la construction de la représentation sémantique, mais également car elle peut être porteuse de phénomènes de contrôle particuliers comme l'accord. Par ailleurs, rien n'impose que les relations de dépendances convergent toutes vers la tête. Plusieurs constructions, par exemple les effets de liste, illustrent ce type de phénomène :

- (4) *J'en veux beaucoup beaucoup*

Dans cet exemple, la répétition de l'adverbe provoque un effet d'intensification. Les deux adverbes ne dépendent pas directement du verbe, nous préférons représenter le second adverbe comme dépendant du premier, cette relation ne passant pas par la tête verbale.

Ces différentes informations forment en quelque sorte le cœur de la structure syntaxique. Elles ne permettent cependant pas de décrire tous les phénomènes. Il est important de poursuivre l'examen de données variées de façon à exhiber de nouvelles propriétés devant être décrites de façon explicite. Nous en proposons deux nouvelles : l'une précisant l'ordre linéaire, la seconde exprimant des relations plus générales sur la distribution des unités.

- **Adjacence** : cette relation permet d'exprimer le fait que deux catégories doivent être juxtaposées. Ce type d'information est relativement peu fréquent et donc moins visible que les autres contraintes. Pour autant, il est nécessaire de le représenter directement, plutôt que par une multiplication de contraintes de précédence. L'exemple suivant illustre ce phénomène :

- (5) a. *C'est un homme aux doigts d'or.*

<sup>3</sup>Certaines approches font jouer un rôle central à la tête (typiquement HPSG) en faisant transiter toutes les informations et tous les contrôles par cette catégorie. Cependant, même si d'un point de vue opérationnel ce choix peut être intéressant, rien ne l'impose.

- b. *C'est un homme tranquille aux doigts d'or.*
- c. *\*C'est un homme aux doigts d'or tranquille.*

Dans ce cas, l'adjectif ne peut être séparé du nom par un *SP*, il doit lui être adjacent.

Les informations syntaxiques sont classiquement exprimées en termes d'unités et de relation entre ces unités. Il est cependant nécessaire d'ajouter une dimension supplémentaire permettant de définir des classes d'unités et d'exprimer des relations sur celles-ci. L'analyse de la langue parlée montre en effet la nécessité de dépasser le cadre de la simple distribution. Il convient en effet de décrire des phénomènes très fréquents, en particulier à l'oral, consistant à entasser des éléments sur un même paradigme de réalisation. Il est pour cela nécessaire d'identifier explicitement dans la structure syntaxique le paradigme de réalisation et d'en indiquer les contraintes.

- **Paradigme** : L'approche pronominale (voir (Blanche-Benveniste84)) a fourni une définition et une description de la notion de paradigmes. Elle repose sur la possibilité d'identifier des positions dans la structure syntaxiques en précisant la notion de substitution, lui substituant celle de proportionnalité (voir (Eynde & Mertens 03) pour une application au lexique). Cette relation permet d'identifier des classes d'arguments substituables (ou proportionnels) à un même pronom. Une structure prédicative est ainsi définie par la séquence des pronoms que le prédicat construit :

- (6) a. *Je le leur dit.*
- b. *Je leur en propose.*

Nous ne revenons pas ici sur la définition des paradigmes. Les trois premiers paradigmes (notés *P0*, *P1*, *P2*) correspondent aux fonctions sujet, objet direct, objet indirect en précisant des sous-classes. Ce type de représentation d'information permet de préciser la structure argumentale des prédicats et donc de proposer un cadre pour la représentation des informations de sous-catégorisation. Celui-ci est ainsi décrit de façon très fine, en offrant également la possibilité de stipuler des relations de cooccurrence ou d'exclusion directement entre les pronoms identifiants des sous-classes.

Mais les paradigmes peuvent également jouer un rôle plus général. Il est en effet possible de proposer des contraintes telles que définies précédemment directement sur les paradigmes. Il est possible d'exprimer des contraintes générales sur l'unicité ou l'ordre des paradigmes, par exemple  $P0 \prec P1 \prec P2$ , cette contrainte pouvant être précisée voire remplacée par des contraintes locales spécifiques.

Par ailleurs, et il s'agit là d'une information essentielle, la connaissance du paradigme de réalisation permet de représenter de façon explicite ses réalisations pour un énoncé donné. Ainsi, dans le cas d'un entassement paradigmatique, il suffit d'indiquer l'appartenance de chaque unité de la liste à son paradigme, plusieurs unités pouvant apparaître dans le même paradigme. De même, les phénomènes de double marquage sont également représentés par l'indication des deux réalisations du même paradigme.

## 5 Représentation formelle

On peut préciser plus formellement les propriétés décrites précédemment. On utilise dans la suite les notations suivantes :  $x, y$  (minuscules) pour représenter les variables individuelles ;

$X, Y$  (majuscules) les variables ensemble. On note  $C(x)$  l'ensemble des variables individuelles dans le domaine de la catégorie  $C$  (voir (Backofen & al. 95) pour plus de précisions). On utilise les prédicats binaires représentant la constituance ( $\triangleleft$ ), la précédence linéaire ( $\prec$ ) et l'égalité ( $\approx$ ). Dans ce qui suit, on considère que le domaine est celui des graphes, plutôt que des arbres. Les variables individuelles représentent des nœuds de ces graphes. La définition des informations de constituance consiste en une déclaration d'ensemble, de même que la notion de paradigme. Les définitions suivantes concernent les autres informations.

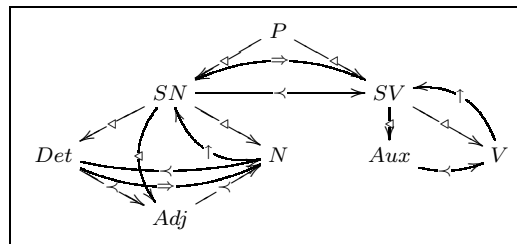
- *Unicité*( $A$ ) :  $(\forall x, y)[A(x) \wedge A(y) \rightarrow x \approx y]$   
Si un élément de catégorie  $A$  est réalisé, il ne peut exister d'autre nœud de même catégorie. L'unicité précise donc les constituants ne pouvant être répétés dans une construction.
- *Précédence*( $A, B$ ) :  $(\forall x, y)[(A(x) \wedge B(y) \rightarrow y \not\prec x)]$   
Cette définition indique que si les nœuds  $x$  et  $y$  sont réalisés, alors  $y$  ne peut pas précéder  $x$ .
- *Tête*( $A$ ) :  $(\exists x)(\forall y)[A(x) \wedge A(y) \rightarrow x \approx y]$   
Il existe un nœud  $x$  de catégorie  $A$  et il n'existe pas d'autre nœud  $y$  de catégorie indentique. Une tête n'est réalisée qu'une fois dans une construction.
- *Cooccurrence*( $A, B$ ) :  $(\forall x, y)[A(x) \rightarrow B(y)]$   
Si un nœud  $x$  de catégorie  $A$  est réalisé, un nœud  $y$  de catégorie  $B$  doit l'être également.
- *Exclusion*( $A, B$ ) :  $(\forall x)(\neg \exists y)[A(x) \wedge B(y)]$   
S'il existe un nœud  $x$ , il ne peut exister un nœud de même niveau  $y$ .
- *Adjacence*( $A, B$ ) :  $(\forall C)(\forall x, y, z)[A(x) \wedge B(y) \wedge C(z) \wedge z \triangleleft x \wedge z \triangleleft y \rightarrow (\forall t)[y \prec t \vee t \prec x]]$   
Si deux nœud  $x$  et  $y$  sont réalisés comme constituants d'une même catégorie  $C$ , alors tout autre nœud  $t$  devra soit les précéder, soit les suivre.

Au total, on peut donc définir une grammaire comme un n-uplet de la forme :

$$G = \langle W, \Rightarrow, \otimes, \circ, \triangleleft, \uparrow, \prec, \theta, \mathcal{P}, \frown \rangle$$

dans lequel les symboles représentent les relations suivantes :  $\Rightarrow$  (cooccurrence),  $\otimes$  (exclusion),  $\circ$  (unicité),  $\triangleleft$  (constituance),  $\uparrow$  (tête),  $\prec$  (précédence),  $\frown$  (adjacence) et  $\mathcal{P}$  (paradigmes).

La figure suivante illustre un ensemble de relations pouvant être spécifiées entre différentes unités. Chaque relation est typée par le symbole de la propriété qu'elle porte.



## 6 Conclusion

Les formalismes syntaxiques de haut niveau doivent permettre la représentation de l'information syntaxique sans dépendre de contraintes opérationnelles spécifiques. En particulier, l'évolution récente des théories linguistiques a montré la nécessité d'une part de distinguer les différents types d'information et d'autre part de les traiter au même niveau. Il devient alors possible de

se libérer de la nécessité de construire une structure syntaxique sous la forme d'un arbre avant de pouvoir vérifier les autres propriétés décrivant l'énoncé analysé. La décentralisation de l'information permet de plus d'identifier explicitement l'ensemble des informations syntaxiques qui entrent en jeu dans une description. Nous avons proposé dans cet article un ensemble de relations décrivant les informations syntaxiques. Ces relations ont vocation à faire partie de tout formalisme de haut niveau.

## Références

- Blackburn P., C. Gardent & W. Meyer-Viol (1993) "Talking About Trees" in proceedings of *EACL*
- Backofen R., J. Rogers, and K. Vijay-Shanker (1995) "A First-Order Axiomatization of the Theory of Finite Trees" in *Journal of Logic, Language, and Information*, 4 :1
- Blache P. & A. Di Cristo (2002), "Variabilité et dépendances des composants linguistiques", in actes de *TALN-2002*.
- Blache P. (2005) "Property Grammars : A Fully Constraint-Based Theory", in H. Christiansen & al. (eds), *Constraint Solving and NLP*, Lecture Notes in Computer Science, Springer.
- Blanche-Benveniste C., J. Deulofeu, J. Stefanini & K. van den Eynde (1984), *Pronom et syntaxe. L'approche pronominale et son application au français*, Selaf.
- Cornell T. & J. Rogers (2000) "Model Theoretic Syntax", in L. Lai-Shen Cheng & R. Sybesma (eds), *The Glot International State of the Article Book I*, Holland Academic Graphics
- Eynde K. van den & P. Mertens (2003) "La valence : l'approche pronominale et son application au lexique verbal", *Journal of French Language Studies* 13.
- Fillmore C. (1998) "Inversion and Constructional Inheritance", in *Lexical and Constructional Aspects of Linguistic Explanation*, Stanford University.
- Gazdar G., E. Klein, G. Pullum & I. Sag (1985) *Generalized Phrase Structure Grammars*, Blackwell
- Guénot M.-L. (2006) *Éléments de grammaire du français : pour une théorie descriptive et formelle de la langue*, Thèse de doctorat, Université de Provence
- Huddleston R. G. Pullum (2002) *The Cambridge Grammar of the English Language*, Cambridge University Press.
- Prince A. & Smolensky P. (1993) *Optimality Theory : Constraint Interaction in Generative Grammars*, Technical Report RUCCS TR-2, Rutgers Center for Cognitive Science.
- Pullum G. & B. Scholz (2001) "On the distinction between model-theoretic and generative-enumerative syntactic frameworks", in proceedings of the conference on *Logical Aspects of Computational Linguistics*, Springer
- Rogers J. (1997) "Grammarless Phrase Structure Grammar", in *Linguistics and Philosophy*, 20
- Sag I., T. Wasow & E. Bender (2003) *Syntactic Theory. A Formal Introduction*, CSLI.