



**HAL**  
open science

## Grammaires d'arbres polychromes

Marcel Cori, Jean-Marie Marandin

► **To cite this version:**

Marcel Cori, Jean-Marie Marandin. Grammaires d'arbres polychromes. Revue TAL : traitement automatique des langues, 1993, 1, pp.101-132. halshs-00784103

**HAL Id: halshs-00784103**

**<https://shs.hal.science/halshs-00784103>**

Submitted on 3 Feb 2013

**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.

## GRAMMAIRES D'ARBRES POLYCHROMES

Marcel CORI\* Jean-Marie MARANDIN\*\*

### Résumé - Abstract

Les grammaires d'arbres polychromes (GAP) constituent un formalisme syntaxique où les notions de position et d'occupabilité d'une position reçoivent une représentation distincte et univoque. Les arbres polychromes sont des arbres dont on peut différencier les branches. Une GAP construit des structures syntaxiques par composition à partir d'un ensemble de structures élémentaires. On enrichit le modèle des grammaires basées sur l'unification par une relation associée aux arbres afin de rendre compte des phénomènes d'interdépendance des valeurs de traits. La composition englobe une opération de compactage qui réduit le nombre de niveaux des arbres composés et a un effet sur l'ordre respectif des branches.

**Mots clés:** formalisme syntaxique, arbre syntagmatique, syntaxe positionnelle, unification.

Polychrome tree grammars (PTG) are a syntactic formalism in which the notions of position and occupability of a position are given distinct representations. Polychrome trees are trees whose branches can be differentiated. A PTG builds syntactic structures by combining elementary structures. A relation is associated with the trees to account for interdependency phenomena between feature values. Composition includes an operation of compacting which reduces the number of levels in resulting trees and which affects the relative order of the branches.

**Key words:** syntactic formalism, elementary tree, positional syntax, unification.

\* Université Paris 7, UFR d'Informatique et UFR de Linguistique, case 7003, 2 place Jussieu, 75251 Paris cedex 05.

\*\* CNRS, URA 1028, Université Paris 7, case 7003, 2 place Jussieu, 75251 Paris cedex 05.

## 1. INTRODUCTION

Le formalisme que nous présentons est un formalisme syntaxique. Son domaine est exclusivement constitué par les faits de syntaxe tels qu'ils sont délimités dans le cadre de l'hypothèse positionnelle (Milner 1989). Dans cette hypothèse relèvent de la syntaxe:

- (i) les organisations de positions qui rendent compte du placement des constituants les uns par rapport aux autres dans les énoncés,
- (ii) l'occupation des positions.

L'hypothèse positionnelle admet que la disposition des constituants ne peut être réduite à l'ordre linéaire que l'on constate dans les énoncés, ou dans chaque parenthésage de l'énoncé, et qu'elle constitue un système d'organisation indépendant des propriétés des unités lexicales.

La composante syntaxique que nous proposons constitue l'un des modules d'une grammaire que nous concevons comme une grammaire autolexicale (Sadock 1991). Dans un modèle non hiérarchique comme les grammaires autolexicales, les modules génèrent (ou analysent) un ensemble de représentations indépendamment de "ce qui se passe dans les autres modules". Chaque module est un système indépendant, définissant et "manipulant" ses propres unités. Comme le note Sadock (ibid., p. 21), l'objectif d'une telle approche est de définir le plus purement possible les différents ordres structuraux dans le langage.

Le problème que nous nous proposons de résoudre est le suivant: comment étendre minimalement les grammaires syntagmatiques pour prendre en compte l'hypothèse positionnelle? Plus précisément, comment distinguer (i) les différentes positions, (ii) une position de ce qui l'occupe.

Nous résolvons le problème en introduisant des arbres dont les branches sont différenciables: elles ont des "couleurs" qui permettent de les distinguer. C'est l'image d'un arbre aux branches de diverses couleurs qui est à l'origine de la dénomination *arbre polychrome*. Chaque couleur de branche représente une position de type différent. Nous obtenons ainsi des représentations syntaxiques qui autorisent une différenciation systématique des positions. Les positions sont données indépendamment des propriétés des unités lexicales, en particulier des catégories lexicales. Par ailleurs, l'opération de construction des arbres complexes à partir d'arbres élémentaires (qui définissent une grammaire) permet de traiter de l'ordre linéaire des constituants sans définir à part, comme dans GPSG, des règles de précédence linéaire.

Le formalisme des arbres polychromes constitue un tout. Toutefois, nous le décomposons dans cet article afin que ses différents éléments puissent être appréhendés progressivement. Au § 2 sont définis les arbres polychromes "nus", puis avec des sommets étiquetés (par des catégories et des traits). Nous introduisons au § 4 la relation d'interdépendance portant sur les valeurs des traits; elle se présente comme une augmentation de l'unification. L'opération de composition d'arbres qui est à la base des grammaires d'arbres polychromes (GAP) est d'abord définie au § 3 comme une opération de substitution, équivalente à la dérivation dans les grammaires syntagmatiques. Sur cette base est définie, au § 5, l'opération de compactage qui lui est obligatoirement associée et qui, elle, modifie fondamentalement la philosophie de construction des arbres: elle permet de dissocier dérivation et représentation (elle a pour effet d'"aplatir" les

arbres" aux seuls niveaux hiérarchiques pertinents pour la description syntaxique) et de rendre compte des différentes relations de proximité entre les constituants.

Cet article est essentiellement consacré à la présentation d'un formalisme. On ne cherche pas à corroborer ou à falsifier l'hypothèse positionnelle; de même, on ne présente pas une syntaxe du français: les exemples que nous introduisons ont pour seule valeur d'éclairer un point du formalisme<sup>1</sup>.

## 2. LES ARBRES POLYCHROMES

### 2.1. Motivation

#### 2.1.1. Les grammaires syntagmatiques

Les représentations arborescentes des grammaires syntagmatiques ordinaires ne permettent pas de distinguer entre les branches d'un arbre. Prenons les trois arbres ci-dessous:

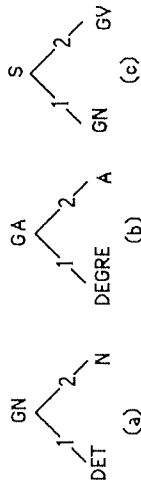


Figure 1

S domine GN de la même manière que GN domine DET ou que GA domine DEGRE, etc. Cette représentation a suscité deux types de critiques. Selon la première, les arbres de la figure 1 ne rendent pas compte des différences dans les "relations logico-sémantiques entre les termes" (Lyons 1970, Wehrli 1988, entre autres); cette critique trouve son écho dans les cadres lexicalistes qui distinguent entre opérateur et opérande (grammaires catégorielles). Selon la seconde, les arbres de la figure 1 ne captent pas les parallélismes configurationnels entre les constituants syntaxiques. Ce sont ces parallélismes que cherche à capter la convention (ou la notation) X-barre (Chomsky 1975, Jackendoff 1977, Emonds 1985 entre autres). X-barre fait une hypothèse forte sur l'organisation syntaxique des énoncés: les configurations dans lesquelles entrent les constituants peuvent être appréhendées indépendamment des unités lexicales et de leur appartenance catégorielle. C'est une première version de l'hypothèse positionnelle<sup>2</sup>.

1. Nous remercions, pour leur lecture critique d'une précédente version de ce travail, B. Fradin, D. Godard, B. Habert, J. Jayez, F. Kerleroux, R. Panckhurst et un lecteur anonyme de *T.A.L.*

2. Dans l'hypothèse positionnelle, les relations "logico-sémantiques" entre les termes sont distinguées des relations entre les positions (Milner 1989, p.292) et constituent un niveau d'organisation à part entière.

### 2.1.2. X-barre et la notion de position

La convention X-barre propose des schémas de règles comme (1):

- (1) a.  $X_{max} \rightarrow SP(X) X_{max-1}$   
 b.  $X_{max-1} \rightarrow X \text{ Complément}$  (Chomsky 1975)

Ces schémas peuvent être représentés informellement par les deux arbres suivants:

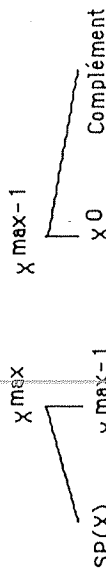


Figure 2

Il y a là tout à la fois une distinction entre branches d'un même arbre ( $X_{max}$  ne domine pas  $X_{max-1}$  de la même manière qu'il domine  $SP(X)$ ) et une parenté entre branches d'arbres différents ( $X_{max}$  domine  $X_{max-1}$  de la même manière que  $Y_{max}$  domine  $Y_{max-1}$ ). Plus exactement, selon la version la plus courante de X-barre, il y a trois manières de dominer: Spécifieur, Noyau, Complément. Nous dirons qu'il y a trois positions génériques différentes, ou trois couleurs de branches.

Construire une grammaire d'une langue donnée dans ce cadre revient à établir ce qui peut apparaître dans ces trois positions; autrement dit, la grammaire enregistrée, en fonction de la catégorie étiquetant le sommet père, quelles peuvent être les catégories étiquetant les sommets fils.

La distinction entre position et ce qui occupe une position est couramment admise pour certaines positions; c'est l'analyse donnée pour la position COMP (relative à la catégorie S) qui peut être occupée par des constituants de catégories différentes. Par exemple:

- (2) a. *La maison* [ $s'$  [ $comp$  ( $n$ " que) ] [ $s$  tu vois] ]  
 b. *La maison* [ $s'$  [ $comp$  ( $p$ " dans laquelle) ] [ $s$  tu vis] ]  
 c. *La maison* [ $s'$  [ $comp$  ( $p$ " sur le toit de laquelle) ] [ $s$  il y a un chat ] ]  
 d. *Je me demande* [ $s'$  [ $comp$  ( $a$ " quel) ] [ $s$  sera son destin] ]

La même position est occupée en (2) par des constituants de catégorie N", P" et A".

La même distinction doit être effectuée pour la position noyau. Les analyses rassemblées sous le chef de la distorsion catégorielle (Milner 1989) sont sur ce point déterminantes. Kerleroux (1990, 1991) montre que la position noyau de GN peut être occupée en français par un terme de catégorie V ou A:

- (3) a. *Le chanter juste est un don des dieux.*  
 b. *Il est d'un sentimental!*

Nous renvoyons à Kerleroux (ibid.) pour les arguments qui justifient l'analyse: il faut, en effet, démontrer que le verbe en (3.a) ou l'adjectif en (3.b) dans les GN n'ont pas subi de règle de dérivation non-affixe (conversion); il faut par ailleurs démontrer que le groupe formé à toutes les

propriétés (distributionnelles, fonctionnelles) d'un GN<sup>3</sup>. Pullum (1991) montre de façon parallèle que les gérondifs nominaux illustrés en (4) doivent être analysés comme des groupes nominaux comportant une tête de catégorie VP:

- (4) a. *Your having broken the record was a surprise.*  
 b. *Rejecting the offer, I guess we had no reason to anticipate, but that's what we ended up doing.*

Ces analyses faussent la proposition essentielle de la théorie X-barre sous deux de ses formulations:

- identité catégorielle: la catégorie de la tête lexicale est identique à la catégorie du constituant;
- univocité: un terme de catégorie X ne peut occuper qu'une position noyau dans une configuration étiquetée Y et la position noyau d'une configuration étiquetée Y ne peut être occupée que par un terme de catégorie X<sup>4</sup>.

En (3) et (4), la position noyau de GN (N" pour X-barre) est occupée par des termes de catégorie V ou A (ce qui contredit l'identité catégorielle); elle peut être occupée par des termes de catégorie N, V ou A (ce qui contredit l'univocité).

Nous tirons de ces analyses deux conclusions:

- les positions doivent être distinguées autrement que par les étiquettes catégorielles.
- l'endocentricité des structures syntaxiques doit être redéfinie en termes de rôle joué par une unité lexicale dans une structure et non plus en termes d'identité catégorielle.

### 2.1.3. "Généraliser" X-barre

Le formalisme que nous proposons autorise l'occupation multivoque de toutes les positions. Sur la base des données d'une langue particulière, on peut évidemment contraindre dans une grammaire l'occupation d'une position à être univoque: c'est le cas de la position noyau du GV en français par exemple<sup>5</sup>.

De même, c'est une question empirique que de déterminer le nombre de positions requises pour la description des structures d'une langue donnée. C'est pourquoi le formalisme autorise un nombre quelconque de positions génériques; ce nombre est fixé une fois pour toutes avant l'écriture d'une grammaire particulière.

3. La mention est un phénomène analogue: la position noyau de GN, ou la position GN toute entière, peut être occupée par un constituant sans considération de son appartenance catégorielle: (a.) *Le nous qui précède le verbe de la troisième ligne est un pronom.* (b.) *Taratata est une onomatopée.*

4. Kornat et Pullum (1990, p. 42) formalisent cette proposition en définissant la tête par une fonction univoque et inversible; ils la considèrent comme le contenu essentiel de X-barre.

5. Le français se distingue sur ce point du latin (par exemple) si la phrase dite nominale est analysée comme: [N" (N"omnis homo) ] [V" (A" mortalis) ] (Milner 1989, p. 367). Autrement dit, les possibilités de distorsion constituent une dimension typologique importante dans la comparaison des langues (ou des groupes de langues).

Les versions courantes de X-barre distinguent trois positions génériques (cf (1) ci-dessus). L'étude des tours incidents du français, par exemple, nous conduit à introduire deux positions supplémentaires. Ainsi, nous traitons les tours illustrés en (5) ci-dessous en introduisant une position à droite de la position noyau de GV:

- (5) a. *Pierre a, le con, vu le voleur sans rien faire.*  
 b. *Pierre a, le fait est scandaleux, vu le voleur sans rien faire.*

On se reportera à Cori et Marandin (1993) pour les arguments qui justifient le traitement positionnel de ce type de tour.

Le formalisme des arbres polychromes, en autorisant l'occupation multivoque de toutes les positions et en ne fixant pas a priori le nombre de positions génériques, peut être vu comme une généralisation de X-barre<sup>6</sup>.

### 2.1.4. Position et rôle fonctionnel

Le contenu propre d'une définition positionnelle de la syntaxe est d'établir comment les constituants se placent les uns par rapport aux autres dans une structure. L'hypothèse essentielle (que nous reprenons) est que les constituants d'une structure ne s'ordonnent pas les uns par rapport aux autres sur l'axe linéaire, mais se disposent relativement à une position distinguée: la position noyau. La position noyau joue, dans cette hypothèse, le rôle d'un pivot à partir duquel sont situés les constituants.

X-barre donne un deuxième contenu à la notion de position: les positions sont appariées à un rôle fonctionnel (on dit parfois que la position assigne aux termes qui l'occupent un rôle fonctionnel)<sup>7</sup>. Par *rôle fonctionnel*, on entend:

- les fonctions grammaticales au sens strict (sujet, objet, objet indirect, par exemple). On se reportera à la définition de sujet, objet, etc. reprise à la grammaire générative.
- des rôles qui ne sont pas distingués dans la tradition grammaticale. Par exemple, la position noyau assigne au terme qui l'occupe le rôle de tête lexicale: la tête impose ses propriétés de sous-catégorisation aux autres constituants de la structure; elle entre dans les relations de restriction avec les autres constituants de la phrase (Emonds 1985, p. 156).

Le formalisme que nous proposons n'inclut pas la notion de fonction. Il ne préjuge donc pas du type d'appariement entre positions et rôles fonctionnels que doit opérer la théorie syntaxique dans une grammaire particulière. Nous reviendrons sur ce point au § 5.3.2.

6. Le système de constituants marqués de Gladkii (1970) peut être considéré comme une des premières tentatives de distinction entre positions et constituants dans le cadre des grammaires syntagmatiques.

7. Les étiquettes Spécifieur, Noyau ou Complément sont explicitement fonctionnelles.

## 2.2. Définition des arbres polychromes

### 2.2.1. Approche intuitive

Prenons 1, 2 et 3 comme couleurs de branches, A, B, ..., G comme étiquettes catégorielles.

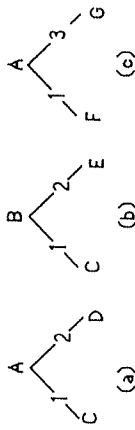


Figure 3

Dans la figure 3, sont présentés trois arbres dont les branches sont "coloriées" afin de les distinguer:

Une position est définie par une catégorie et une couleur:  $\langle A, 2 \rangle$  est la position 2 de A. Ainsi,  $\langle A, 2 \rangle$ ,  $\langle B, 2 \rangle$  et  $\langle A, 3 \rangle$  sont des positions différentes.  $\langle A, 1 \rangle$  est la même position dans les arbres (a) et (c), indépendamment de l'appartenance catégorielle des fils. Toutefois, il y a une parenté entre des positions qui mettent en jeu une même couleur. Par exemple, il y a une parenté entre  $\langle A, 2 \rangle$  et  $\langle B, 2 \rangle$ . Nous parlerons de position générique.

### 2.2.2. Définition formelle

On se fixe un nombre  $p$  (entier strictement positif) qui indique le nombre de couleurs possibles pour les branches des arbres. On définit un arbre polychrome (à  $p$  couleurs), ou  $p$ -arbre, par le couple  $\langle X, \delta \rangle$  où:

- (i)  $X$  est un ensemble de sommets, fini non vide.
- (ii)  $\delta = \langle \delta_1, \delta_2, \dots, \delta_p \rangle$  est une suite d'applications de  $X$  dans le monoïde libre  $X^*$ . Si, pour un  $x$  donné, un sommet  $y$  a une occurrence dans  $\delta_i(x)$ , il existe un arc  $\langle x, y \rangle$  qui reçoit la  $i$ -ème couleur.
- (iii) soit  $\delta_{1p}$  l'application de  $X$  dans  $X^*$  définie, pour tout  $x$  appartenant à  $X$ , par  $\delta_{1p}(x) = \delta_1(x) \delta_2(x) \dots \delta_p(x)$ ;  $\langle X, \delta_{1p} \rangle$ , la structure obtenue à partir de l'arbre polychrome en "oubliant" les couleurs, vérifie les propriétés caractéristiques des arborescences.

Remarque: Il y a un ordre (de précédence) entre les différentes couleurs: si  $i$  est inférieur à  $j$ , tout sommet  $y$ , fils de  $x$  de couleur  $i$  (c'est-à-dire appartenant à  $\delta_i(x)$ ), précède tout sommet  $z$  de  $x$  de couleur  $j$ . Il y a également un ordre de précédence entre arcs de la même couleur, puisque chaque  $\delta_i(x)$  est une suite d'éléments de  $X$ .

Rappelons que la racine d'une arborescence (et par conséquent d'un arbre polychrome) est le sommet unique qui n'appartient à aucun  $\delta_{1p}(x)$ , (et par conséquent à aucun  $\delta_i(x)$ ). Les feuilles sont les sommets  $z$  tels que  $\delta_{1p}(z)$  est vide, et par conséquent tels que:

$$\forall i \in \{1, \dots, p\} \delta_i(z) = \varepsilon$$

$\varepsilon$  désignant le mot vide du monoïde libre.

8. Voir par exemple Berge (1970).

### 2.2.3. Illustration

Nous illustrons cette première définition en représentant l'hypothèse configurationnelle défendue dans la théorie X-barre que nous représentons à la figure 4.

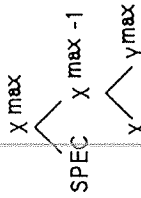


Figure 4

Plaçons-nous dans le formalisme des arbres polychromes. En fixant  $p = 3$ , on peut distinguer trois "couleurs" de branche:

- la branche 1 ou *spécifieur*,
  - la branche 2 ou *noyau*,
  - la branche 3 ou *position des compléments*.
- On décrit  $\langle X, \delta \rangle$  correspondant à (2) de la manière suivante:

$$\begin{aligned}
 X &= \{a, b, c, d, e\} \\
 \delta &= \langle \delta_1, \delta_2, \delta_3 \rangle \\
 \delta_1(a) &= b \quad \delta_2(a) = c \quad \delta_3(a) = \varepsilon \\
 \delta_1(b) &= \varepsilon \quad \delta_2(b) = \delta_3(b) = \varepsilon \\
 \delta_1(c) &= \varepsilon \quad \delta_2(c) = d \quad \delta_3(c) = e \\
 \delta_1(d) &= \delta_2(d) = \delta_3(d) = \varepsilon \\
 \delta_1(e) &= \delta_2(e) = \delta_3(e) = \varepsilon
 \end{aligned}$$

Ce qui se représente par la figure 5:

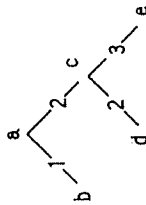


Figure 5

L'arbre n'exhibe que des positions, correspondant à celles qui sont données par les règles de (1). Dans chaque sous-arbre, il y a une relation de précédence entre arcs:  $\langle a, b \rangle$  est à gauche de  $\langle a, c \rangle$ ;  $\langle c, d \rangle$  est à gauche de  $\langle c, e \rangle$ . Les sommets sont "nus": on peut, de ce fait, spécialiser l'étiquetage des sommets pour l'expression d'un nouvel aspect de la structure: l'occupabilité des positions.

L'arbre de la figure 5 est un p-arbre où il n'y a qu'une branche par couleur sous un sommet donné. C'est un cas particulier: un p-arbre peut avoir plusieurs branches de même couleur. Par exemple :



Figure 6

On a dans ce cas:  $\delta_3(c) = efg$

Nous renvoyons au § 3.1. ci-dessous où des arbres tels que celui qui est figuré dans la figure 6 permettent de traiter les tours où une même position est occupée par plusieurs constituants sans former eux-mêmes un constituant (organisation paratactique).

## 2.3. L'étiquetage

### 2.3.1. Occupation et étiquetage des sommets

Reprenons l'arbre décrit par une règle comme (1.a) ci-dessus; il distingue deux positions. Il nous faut maintenant préciser quels sont les termes qui peuvent "s'insérer" sous les deux branches. Si l'on décrit un Groupe Nominal, quels sont les termes qui peuvent occuper la position spécifieur de GN? Est-ce que le français admet d'autres termes que des noms en position de tête? La position tête est-elle minimale? Etc. Ces questions constituent le problème de l'occupabilité des positions. Dans le formalisme des arbres polychromes, l'étiquetage des sommets peut être exclusivement consacré à sa spécification. De quel type doit être cet étiquetage?

Les modèles syntaxiques contemporains n'ont généralement considéré que l'occupabilité des positions minimales<sup>9</sup> sous le chef de l'insertion lexicale. Un consensus existe sur ce point: l'insertion lexicale n'est pas exhaustivement décrite par la seule mention de la catégorie syntaxique ("partie du discours"). Cette proposition est fondée sur une constatation: l'insertion est sujette à un certain nombre de restrictions contextuelles, au premier rang desquelles l'accord. C'est pour tenir compte de restrictions de ce type que les grammairiens ont introduit des structures de traits. La solution peut être généralisée à toutes les positions.

Nous reprenons la solution d'un étiquetage complexe étendu à toutes les positions: il doit nous permettre de rendre compte du détail des phénomènes d'occupation dans une langue donnée (on en trouvera un exemple au § 2.3.3 ci-dessous).

La version du formalisme que nous proposons ici traite différemment les catégories (lexicales ou syntagmatiques) et les traits. Ce choix est sous-tendu par un faisceau de motivations que nous signalons sans les développer:

- la version de l'endocentricité que nous adoptons vide de contenu l'analyse des catégories syntagmatiques par un couple de traits <catégorie lexicale, niveau hiérarchique><sup>10</sup>,
- catégorie et traits jouent des rôles distincts. C'est la catégorie globale qui est mise en jeu dans les phénomènes de distorsion ou, comme nous le verrons plus bas, dans l'opération de compactage; les traits n'y jouent aucun rôle. Les traits, par contre, jouent un rôle décisif dans les relations entre les constituants (l'accord et la sous-catégorisation étant les cas les plus notables).

9. Une position minimale est une position qui ne domine pas d'autres positions.

10. Par exemple, dans la notation reprise à Gazdar:  $N^i = \langle \text{cat: } N, \text{ barre: } 2 \rangle$ . Nous abandonnons de même l'analyse des catégories lexicales par les traits  $\langle +/- N \rangle$  et  $\langle +/- V \rangle$ .

De façon générale, nous considérons les traits comme des spécifications sur les catégories et non comme les éléments représentant les éléments constitutifs d'une catégorie (comme dans les analyses componentielles des structures et des constituants proposées dans les formalismes basés sur l'unification).

### 2.3.2 Définition des arbres polychromes étiquetés

On se donne un nombre entier strictement positif  $p$ , un ensemble de catégories  $Cat$ , un ensemble d'attributs  $T$ , un ensemble de valeurs (de traits)  $E$ . Un p-arbre étiqueté est défini comme étant un 5-uple  $A = \langle X, \delta, cat, \tau, t \rangle$  où:

- (i)  $\langle X, \delta \rangle$  est un p-arbre;
- (ii)  $cat$  est une application qui associe à chaque élément  $x$  de  $X$  une catégorie  $cat(x)$  appartenant à  $Cat$ ;
- (iii)  $\tau$  est une application qui associe à chaque  $x$  de  $X$  un sous-ensemble  $\tau(x)$  de  $T$ .  $\tau(x)$  est l'ensemble des attributs associés à  $x$ ;
- (iv)  $t$  est une application qui associe à chaque élément  $x$  de  $X$  une application  $t(x)$  de  $\tau(x)$  dans  $E$ .

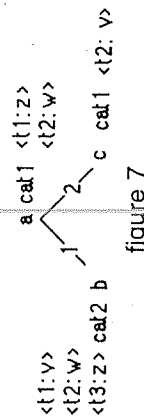
Si  $y$  appartient à  $\tau(x)$ , son image par  $t(x)$  sera notée  $t(x,y)$  plutôt que  $t(x)(y)$ .  $t(x,y)$  est la valeur associée à l'attribut  $y$  pour le sommet  $x$ .

Ainsi, à chaque sommet  $x$  sont associés une catégorie et un ensemble de traits: un trait est un couple attribut/valeur. L'ensemble des traits associés à  $x$  est l'ensemble qui suit:

$$\{ \langle y, z \rangle ; y \in \tau(x), z = t(x,y) \}$$

Nous noterons chaque couple attribut/valeur par  $\langle y:z \rangle$ .

Donnons un exemple d'arbre polychrome étiqueté abstrait (fig. 7):



On a:

$$cat(a) = cat(c) = cat1 \quad cat(b) = cat2$$

$$\tau(a) = \{t1, t2\} \quad \tau(b) = \{t1, t2, t3\} \quad \tau(c) = \{t2\}$$

$$t(a, t1) = z \quad t(a, t2) = w$$

$$t(b, t1) = v \quad t(b, t2) = w \quad t(b, t3) = z$$

$$t(c, t2) = v$$

Dans ce qui suit, quand nous parlerons de p-arbres, il sera toujours question de p-arbres étiquetés.

### 2.3.3. Illustration

Nous reprenons la description de l'infinitif en position noyau du GN donnée par Kerleroux (1990). La version que nous présentons ici est simplifiée pour l'illustration; on trouvera dans Cori et Marandin (soumis à publication) un traitement général du phénomène de distorsion (incluant en

particulier la notion d'occupation préférentielle qui n'est pas introduite ici). Le tour est illustré en (6) ci-dessous:

- (6) a. *Le manger cru des instinctivores n'a pas reçu de caution médicale.*
- b. *"Leur port de tête est un dire comme leur nuque voutée"* (cité dans Kerleroux, id).
- c. *"L'image dialectique [est un] un travail du voir"* (ibid).

La description du tour peut être résumée ainsi:

- (i) le verbe est à l'infinitif<sup>11</sup>; il ne peut pas être construit avec un complément réalisant un argument<sup>12</sup>; il peut être construit avec un adjectif adverbial (*le chanter juste, le parler vrai*),
- (ii) le groupe nominal est normal: il y a un déterminant; il peut y avoir un adjectif apposé et les compléments circonstanciels caractéristiques du GN apparaissent librement (en particulier le complément à interprétation possessive).

La position noyau de GN est occupable par un verbe à condition qu'il soit à l'infinitif. L'étiquetage de cette position requiert donc la mention d'une propriété en plus de la catégorie; on notera que ce trait est requis indépendamment de toute considération contextuelle.

En termes d'arbres polychromes, admettons d'attribuer à la position noyau la couleur 3 (la couleur 1 étant attribuée à la position du déterminant)<sup>13</sup>. On définit les deux arbres (a) et (b) ci-dessous; afin de traiter les tours comme *le manger cru*, on définit l'arbre (c):

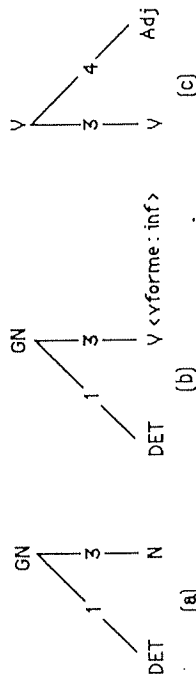


Figure 8

Nous admettons en définissant l'arbre (c) qu'une catégorie lexicale peut dominer une unité polylexicale; cela revient à étendre à la catégorie V la notion de synapsie proposée par Benveniste (1974) pour la catégorie N et d'analyser le complexe "V+Adj" comme une synapsie verbale.

Le phénomène de distorsion n'est pas limité à la position noyau des groupes nominaux, et plus généralement à la position noyau des syntagmes; il affecte d'autres positions, en particulier celles qui sont associées à une fonction (relativement à S). Nous renvoyons à Delaveau

11. Comme le rappelle Kerleroux, l'infinitif n'est pas la forme zéro du verbe, mais une de ses formes flexionnelles. L'auteur a également montré que l'on ne pouvait résoudre le problème par une règle morphologique de substantivation.

12. Ce trait illustré par les GN agrammaticaux ci-dessous distingue le français contemporain de l'ancien français: (a.) \* *Le gouverner les français*, (b.) \* *Le manger des frites fait popu.*

13. La grammaire qui nous sert à illustrer le formalisme dans cet article comporte cinq couleurs de branches, afin de permettre le traitement des tours incidents (cf § 5.1.2 ci-dessous).

(communication personnelle) pour l'analyse des tours illustrés en (7) où une relative apparaît dans une position occupable par un GN:

- (7) a. *Qui l'aime lui donne sa voix.*  
 b. *Je connais qui tu aimes.*  
 c. *Je me confierai à qui tu fais confiance.*

Ce dernier point est important en ce qu'il justifie que nous distinguions systématiquement les positions et non pas seulement la position noyau des groupes dont le pivot est occupé par une unité lexicale.

### 3. COMPOSITION DES ARBRES POLYCHROMES

#### 3.1. Composabilité et grammaire

Les arbres étiquetés étant introduits, on peut construire une grammaire en définissant:

- l'opération de composition des arbres polychromes,
- les caractéristiques des ensembles minimaux d'arbres polychromes susceptibles, par composition, d'engendrer tous les arbres syntaxiques. De tels ensembles minimaux seront appelés bases d'arbres polychromes.

Les arbres d'une base sont élémentaires (un seul niveau de profondeur) et pauvres (une seule branche d'une couleur donnée issue d'un sommet donné). On obtient par composition les arbres présentant plusieurs niveaux et ceux qui présentent plusieurs branches d'une même couleur. Illustrons le mécanisme par un exemple. Le groupe nominal français est susceptible de plusieurs expansions: groupes prépositionnels (GP), groupes adjectivaux (GA), phrases (GS). Ce qu'illustrent les énoncés ci-dessous:

- (8) a. *Le ski à Tignes en été à 3000m d'altitude.*  
 b. *Le projet que tu reviennes qui est cher à mon coeur.*  
 c. *Le livre de Zola tout à fait passionnant que tu m'as prêté.*

La représentation ci-dessous met en avant la similitude de statut des expansions:

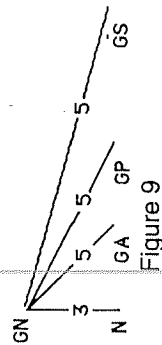


Figure 9

Toutefois, il n'est pas possible d'énumérer la combinatoire d'expansions permises: le nombre d'expansions n'est pas borné. Si on admet dans la base les arbres pauvres suivants:

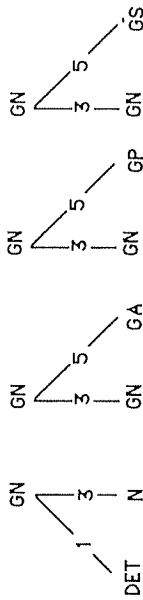


Figure 10

on pourra, par composition, obtenir tous les GN présentant autant d'expansions que l'on voudra. Par exemple:

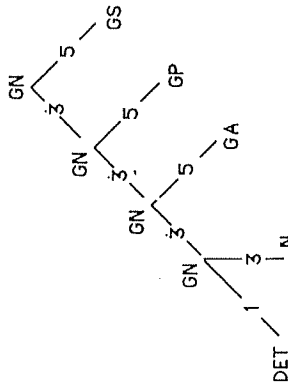


Figure 11

Nous reviendrons au paragraphe 5 sur "l'excès de structure" que manifeste la configuration ci-dessus (figure 11) par rapport à celle de la figure 9.

### 3.2. Définition de la composition

#### 3.2.1. Composabilité

On se donne deux p-arbres  $A_1 = \langle X_1, \delta_1, \text{cat}_1, \tau_1, t_1 \rangle$  et  $A_2 = \langle X_2, \delta_2, \text{cat}_2, \tau_2, t_2 \rangle$  tels que  $X_1 \cap X_2 = \emptyset$ ; désignons par  $b$  la racine  $A_1$  et  $A_2$  sont composables suivant  $a$  si et seulement si:

- (i)  $\text{cat}_1(a) = \text{cat}_2(b)$
- (ii)  $\forall w \in \tau_1(a) \cap \tau_2(b) \quad t_1(a, w) = t_2(b, w)$

La condition (i) signifie que les sommets  $a$  et  $b$  ont même catégorie, la condition (ii) que les traits associés à  $a$  et  $b$  sont compatibles: il n'existe pas un même attribut qui reçoit des valeurs différentes pour  $a$  et  $b$ . Ces deux conditions étant vérifiées, on peut composer les deux arbres en "accrochant" l'arbre  $A_2$  par sa racine à la feuille  $a$ . Les sommets  $a$  et  $b$  sont fusionnés, pour ne plus former qu'un seul sommet qu'on suppose (arbitrairement) être le sommet  $a$ . Ce sommet est étiqueté par la catégorie commune à  $a$  et  $b$  ainsi que par la réunion des traits associés à  $a$  et  $b$ , ce qui est possible d'après la condition (ii).



### 3.2.2. Arbre polychrome composé

Formellement, le p-arbre composé suivant a de  $A_1$  et  $A_2$  est le p-arbre  $A = \langle X, \delta, \text{cat}, \tau \rangle$ , noté  $[A_1, a, A_2]$ , et défini comme suit:

- (i)  $X = X_1 \cup X_2 - \{b\}$
- (ii)  $\forall x \in X_1 - \{a\} \quad \delta(x) = \delta_1(x)$   
 $\forall x \in X_2 - \{b\} \quad \delta(x) = \delta_2(x)$
- (iii)  $\delta(a) = \delta_2(b)$   
 $\forall x \in X_1 \quad \text{cat}(x) = \text{cat}_1(x)$   
 $\forall x \in X_2 - \{b\} \quad \text{cat}(x) = \text{cat}_2(x)$
- (iv)  $\forall x \in X_1 - \{a\} \quad \tau(x) = \tau_1(x)$   
 $\forall x \in X_2 - \{b\} \quad \tau(x) = \tau_2(x)$
- (v)  $\tau(a) = \tau_1(a) \cup \tau_2(b)$   
 $\forall x \in X_1 - \{a\} \quad \forall w \in \tau_1(x) \quad t(x, w) = t_1(x, w)$   
 $\forall x \in X_2 - \{b\} \quad \forall w \in \tau_2(x) \quad t(x, w) = t_2(x, w)$   
 $\forall w \in \tau_1(a) \quad t(a, w) = t_1(a, w)$   
 $\forall w \in \tau_2(b) \quad t(a, w) = t_2(b, w)$

Exemple: Les deux arbres qui suivent donnent, par composition suivant la feuille de (a) étiquetée par GN, l'arbre de la figure 13.



Figure 12

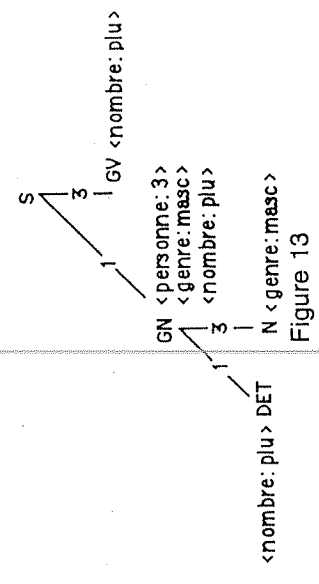


Figure 13

On montre aisément les deux propriétés suivantes:

Propriété 1: Si  $A_1, A_2$  et  $A_3$  sont trois p-arbres, si  $a$  est une feuille de  $A_1, a'$  une feuille de  $A_2$ , et si  $A_1$  et  $A_2$  sont composables suivant  $a$  et  $A_2$  et  $A_3$  sont composables suivant  $a'$ , alors:

$$[[A_1, a, A_2], a', A_3] = [A_1, a, [A_2, a', A_3]]$$

Propriété 2: Si  $A_1, A_2$  et  $A_3$  sont trois p-arbres, si  $a$  et  $a'$  sont des feuilles de  $A_1$ , et si  $A_1$  et  $A_2$  sont composables suivant  $a$  et  $A_1$  et  $A_3$  sont composables suivant  $a'$ , alors:

$$[[A_1, a, A_2], a', A_3] = [[A_1, a', A_3], a, A_2]$$

De ces deux propriétés on déduit que l'ordre dans lequel on compose les arbres, dans la mesure où la composition est possible, n'a pas d'importance.

### 3.3 Base de p-arbres élémentaires

Définition: Un p-arbre est pauvre si et seulement si pour tout  $x$  et pour tout  $i$   $\delta_i(x)$  a pour longueur 0 ou 1.  
 Propriété: Si  $A_1$  et  $A_2$  sont des p-arbres pauvres, alors tout composé de  $A_1$  et  $A_2$  est aussi un p-arbre pauvre.

Définition: Un p-arbre est élémentaire si et seulement si c'est un p-arbre pauvre dont un sommet et un seul (la racine) n'est pas une feuille.

Une base de p-arbres élémentaires est constituée par un ensemble fini  $B = \{A_1, A_2, \dots, A_m\}$  de p-arbres élémentaires.

La base engendre un ensemble  $L(B)$  de p-arbres. Cet ensemble est ainsi défini:

- (i) chacun des  $A_i$  appartient à  $L(B)$ ;
- (ii) si  $A$  et  $A'$  appartiennent à  $L(B)$ , alors tous les  $[A, a, A']$  appartiennent aussi à  $L(B)$ .

Propriété: Les arbres de  $L(B)$  sont des p-arbres pauvres.

### 3.4. Grammaires d'arbres polychromes

Une grammaire d'arbres polychromes est donnée par un quadruplet  $G = \langle \text{Cat}, T, E, B \rangle$  où  $\text{Cat}$  est l'ensemble des catégories,  $T$  est l'ensemble des attributs,  $E$  est l'ensemble des valeurs de traits et  $B$  est une base de p-arbres élémentaires.

Une telle grammaire engendre un ensemble de p-arbres: l'ensemble  $L(B)$ . Les éléments de cet ensemble sont des structures syntaxiques et non des phrases bien formées. En effet, nous n'avons pas distingué d'axiome fonctionnant comme la catégorie commune à toutes les suites bien formées d'une langue donnée, et les positions minimales ne sont pas nécessairement occupées par des unités du lexique. Ce sont les structures de  $L(B)$  qui sont soumises à l'opération de compactage définie au § 5. Compactées, elles contiendront le vecteur de l'information configurationnelle et positionnelle spécifique d'une langue donnée<sup>14</sup>.

14. En fait, une grammaire sera donnée par un 5-uplet  $G = \langle \text{Cat}, T, E, F, B \rangle$  après introduction, au § 4, des variables de traits.  $L(B)$  ne contient pas toute l'information réglant le placement des constituants dans un énoncé: l'ordre de placement linéaire des constituants dominés par une branche de même couleur obéit à des principes non syntaxiques.

#### 4. VARIABLES DE TRAITS ET RELATIONS D'INTERDEPENDANCE

##### 4.1. Unification et calcul des valeurs

Les deux structures ci-dessous (fig. 14) ne se distinguent pas structurellement; elles ne diffèrent que par la valeur du trait nombre:

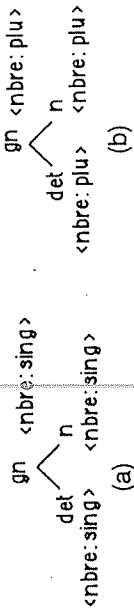


Figure 14

Si l'on cherche à rendre compte de l'identité structurale, il faut pouvoir identifier (a) et (b): c'est ce que permet le recours à des traits de la forme <attribut: valeur> et l'introduction de variables de trait qui apparaissent en position de valeur. Dès lors, on peut multiplier les attributs dans une structure, sans avoir à multiplier la représentation de cette structure pour rendre compte de la combinatoire des différentes valeurs des attributs:

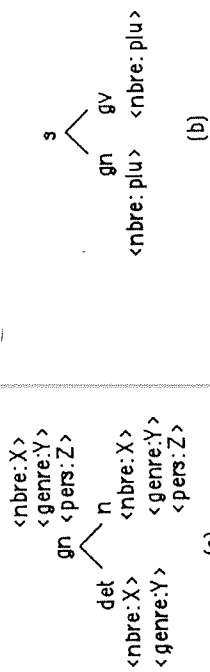


Figure 15

Enfin, le recours à des variables permet de transmettre l'information d'un arbre à un autre. Par exemple, on rendra compte d'un cas d'accord sujet/verbe en composant les arbres (a) et (b) (fig. 15) pour obtenir l'arbre suivant:

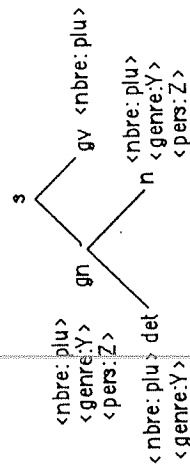


Figure 16

Dans la figure 16, il y a identification des sommets de catégorie GN des arbres (a) et (b) de la figure 15, et unification de leurs structures de traits. La valeur pluriel est ainsi transmise à tous les attributs <nombre>.

La transmission à l'identique, parfaitement illustrée par l'exemple précédent, est insuffisante pour traiter d'un certain nombre de phénomènes

dans les langues. C'est exemplairement le cas des types d'accord traités dans les études empiriques par des "règles de résolution". Nous illustrons ce cas par un exemple emprunté au latin:

(9)	<i>murus</i>	et	<i>porta</i>	<i>de caelo</i>	<i>tacta</i>	<i>erant.</i>
	masc sg		fem sg		neut plu	plu
	mur	et	porte	du ciel	touchés	étaient
	(Le mur et la porte furent touchés par la foudre)					

Le participe passé (*tacta*) s'accorde avec le sujet: on en déduit que le GN *murus* et *porta* est au neutre pluriel alors qu'il est composé d'un GN masculin et d'un GN féminin. Corbett (1983, p. 194) donne la règle de résolution (10) pour couvrir l'accord en genre impliquant des GN coordonnés en latin:

- (10) "1. si tous les conjoints sont masculin, alors on utilise le masculin,
2. si tous les conjoints sont féminin, alors on utilise le féminin,
3. si tous les conjoints réfèrent à des humains, alors on utilise le masculin,
4. autrement on utilise le neutre<sup>15</sup>."

Représentons le phénomène tel que le décrit la règle (10) par le schéma suivant:

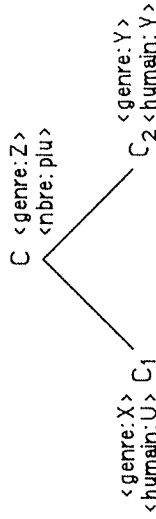


Figure 17

Les sous-règles (10.1) et (10.2) décrivent des groupes nominaux pour lesquels le genre est identique pour les conjoints et le groupe coordonné. Autrement dit, X, Y et Z prennent les mêmes valeurs. Ceci s'exprime naturellement par l'unification. Dans le cas (10.3), la valeur du genre dans le groupe coordonné est égale à la valeur du genre de l'un des conjoints: la valeur de Z est égale soit à X soit à Y. Mais cette valeur dépend de la propriété "humain/non humain" des conjoints, autrement dit de la valeur de U et V qui peut être *oui* ou *non* (ceci est un cas d'interdépendance des valeurs). La sous-règle (10.4) inclut le cas où Z peut prendre une valeur (*neutre*) distincte des valeurs de X et Y; on peut parler dans ce dernier cas d'émergence d'une valeur. Nous considérerons que l'émergence est un cas particulier de l'interdépendance.

On peut conserver l'unicité du phénomène en définissant une relation qui indique comment les valeurs de traits sont corrélées; nous appelons cette relation relation d'interdépendance.

15. "These rules are ordered; there is no need, therefore, to stipulate that the conjuncts in Rule 3 are of mixed gender. Similarly, Rule 4 will automatically cover cases of mixed gender and those where all the conjuncts are neuter". (Corbett, ibid.)

## 4.2. Nouvelle définition des arbres polychromes

### 4.2.1. Définition

On se donne un ensemble (dénombrable) de variables de traits  $\Phi$ .

Un p-arbre est à présent défini de la manière suivante:

$A = \langle X, \delta, \text{cat}, \tau, F, t, R \rangle$  où les conditions (i), (ii) et (iii) du paragraphe 2.2.2 sont vérifiées et où:

(iv)  $F = \{f_1, f_2, \dots, f_m\}$  est un sous-ensemble fini de  $\Phi$  (qui peut être vide).

(v)  $t$  est une application qui associe à chaque élément  $x$  de  $X$  une application  $t(x)$  de  $\tau(x)$  dans  $E \cup F$ . Chaque élément de  $F$  est image d'au moins un couple  $\langle x, y \rangle$ .

(vi)  $R$  est un sous-ensemble de  $E^m$ .

$R$  est la relation d'interdépendance qui est donnée avec l'arbre. Etant donné un  $m$ -uplet  $u = \langle u_1, u_2, \dots, u_m \rangle$  de  $R$ , on obtient un p-arbre sans traits variables  $A_u = \langle X, \delta, \text{cat}, \tau, t_u \rangle$  en définissant  $t_u$  de la manière suivante:

$$\forall x \in X \forall y \in \tau(x) \quad t(x,y) \in E \Rightarrow t_u(x,y) = t(x,y)$$

$$\forall x \in X \forall y \in \tau(x) \quad t(x,y) = f_i \Rightarrow t_u(x,y) = u_i$$

On a finalement deux définitions alternatives des p-arbres (§2.3.2. ou 4.2.1.) selon qu'elles autorisent ou pas des variables de traits. Quand on se placera dans le cadre de la première définition, on emploiera le terme "p-arbre à étiquetage constant". Par conséquent, en définissant un p-arbre, on définit une famille de p-arbres à étiquetage constant, qui ne diffèrent que par les valeurs associées aux traits.

Dans le cas particulier où  $F$  est vide (aucun trait variable), cette famille se réduit à un seul p-arbre.

Remarque: Considérons le cas où il existe un nombre  $k$  et une valeur  $v$  appartenant à  $E$  tels que tous les  $m$ -uplets  $\langle u_1, \dots, u_k, \dots, u_m \rangle$  de  $R$  vérifient  $u_k = v$ .

En ce cas, à l'arbre  $A' = \langle X, \delta, \text{cat}, \tau, F', t', R' \rangle$  obtenu en substituant à la variable  $f_k$  la constante  $v$ , est associée la même famille de p-arbres à étiquetage constant qu'à l'arbre  $A$  d'origine. Formellement:

$$(a) \quad F' = F - \{f_k\}$$

$$(b) \quad \forall x \in X \forall y \in \tau(x) \quad t(x,y) \neq f_k \Rightarrow t'(x,y) = t(x,y)$$

$$(c) \quad \forall x \in X \forall y \in \tau(x) \quad t(x,y) = f_k \Rightarrow t'(x,y) = v$$

(d)  $R'$  est le sous-ensemble de  $E^{m-1}$  défini comme suit:

$$\langle u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_m \rangle \in R' \text{ ssi } \langle u_1, \dots, u_{k-1}, v, u_{k+1}, \dots, u_m \rangle \in R$$

On suppose par conséquent que ce cas particulier ne se produira pas, puisqu'on peut toujours se ramener à une représentation plus simple (au sens qu'elle comporte moins de variables de trait).

### 4.2.2. Illustration

Admettons de traiter une coordination (binaire) de GN en latin par le p-arbre suivant:

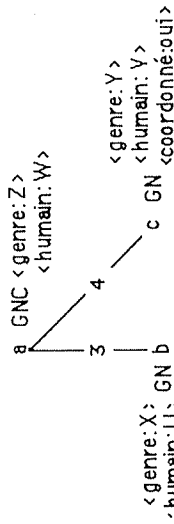


Figure 18

On définit une relation  $R(U, V, W, X, Y, Z)$  sur les variables de trait correspondant aux attributs genre et humain. On peut donner explicitement tous les  $n$ -uplets appartenant à la relation (il y en a 40). Une présentation plus transparente consiste à donner d'une part le sous-domaine de chaque variable de traits et, d'autre part, de "fragmenter" la relation d'interdépendance en relations partielles. La relation d'interdépendance peut, en effet, être vue comme la conjonction ou le produit (*produit* au sens de Codd (1970) et des bases de données relationnelles) de plusieurs relations partielles.

(i)  $U, V, W \in \{\text{oui}, \text{non}\}$

(ii)  $X, Y, Z \in \{\text{masc}, \text{fem}, \text{neutre}\}$

(iii)  $(U=\text{non}) \vee (V=\text{non}) \wedge (W=\text{non}) \vee (U=\text{oui} \wedge V=\text{oui} \wedge W=\text{oui})$

(iv)  $((X=Y) \wedge (Y=Z)) \vee ((X \neq Y) \wedge (V=\text{oui}) \wedge (Z=\text{masc})) \vee ((X \neq Y) \wedge (U=\text{non}) \wedge (Z=\text{neutre})) \vee ((X \neq Y) \wedge (V=\text{non}) \wedge (Z=\text{neutre}))$

La relation (iv) exprime, de manière déclarative, la règle de résolution de Corbett.

La relation d'interdépendance est par exemple satisfaite par le 6-uple  $\langle \text{non}, \text{non}, \text{non}, \text{masc}, \text{fem}, \text{neutre} \rangle$ . Ceci est un cas pour lequel une valeur (le neutre) émerge. Ce cas correspond à l'exemple (9) donné ci-dessus.

Le recours à une relation plutôt qu'à une règle procédurale comme en (10) permet de ne pas privilégier la racine par rapport aux feuilles. La valeur du trait en racine peut être calculée à partir des valeurs en feuilles, ou au contraire les valeurs des traits associés aux conjoints, ou du moins des restrictions quant à ces valeurs, peuvent être calculées à partir des valeurs des traits de la racine. Par exemple, dans la figure 18) ci-dessus: - si  $Z = \text{fem}$ , alors  $X = \text{fem}$  et  $Y = \text{fem}$ .

- si  $Z = \text{masc}$ , alors soit  $X = Y = \text{masc}$ , soit  $U = V = \text{oui}$  et  $X \neq Y$ .

### 4.3. Nouvelle définition de la composition

Désormais, avec chaque arbre polychrome est donnée une relation d'interdépendance. Par conséquent, si l'on compose deux arbres, il est nécessaire de pouvoir calculer la nouvelle relation d'interdépendance, associée à l'arbre composé. Il se peut que cette nouvelle relation soit vide. En ce cas, il y a échec de la composition.

### 4.3.1. Composabilité

Considérons les deux p-arbres  $A_1 = \langle X_1, \delta_1, \text{cat}_1, \tau_1, F_1, t_1, R_1 \rangle$  et  $A_2 = \langle X_2, \delta_2, \text{cat}_2, \tau_2, F_2, t_2, R_2 \rangle$  avec  $F_1 = \{f_1, f_2, \dots, f_n\}$  et  $F_2 = \{g_1, g_2, \dots, g_m\}$ . On suppose que  $F_1 \cap F_2 = \emptyset$ , cas auquel on peut toujours se ramener. Soient a une feuille de  $A_1$  et b la racine de  $A_2$ .

Pour que ces deux arbres soient composables suivant a il faut qu'il y ait identité des catégories étiquetant les sommets a et b (condition (i)), la même qu'au § 3.2.1), il faut que les traits associés à ces deux sommets puissent être unifiés (condition (ii)), différente de celle apparaissant en 3.2.1), il faut enfin que l'on puisse construire une nouvelle relation d'interdépendance sur l'arbre composé (iii). Formellement,  $A_1$  et  $A_2$  sont composables suivant a si et seulement si:

- (i)  $\text{cat}_1(a) = \text{cat}_2(b)$
- (ii) il existe une substitution  $\sigma$  qui, appliquée aux deux ensembles  $F_1$  et  $F_2$ , vérifie :  
Plus précisément, on pose que  $\sigma$  est une application de la forme  
$$\sigma : E \cup F_1 \cup F_2 \rightarrow E \cup F$$
 qui ne modifie pas les éléments de  $E$ ,  $F = \{h_1, h_2, \dots, h_m\}$  étant un sous-ensemble de  $\Phi$  dont chaque élément est image par  $\sigma$  d'au moins un élément de  $F_1$  ou  $F_2$ .
- (iii) la relation  $R$ , définie comme suit, n'est pas vide.  $R$  est le sous-ensemble de  $E^m$  tel que pour tout m-uplet  $\langle w_1, w_2, \dots, w_m \rangle$  de  $R$  il existe un r-uplet  $\langle u_1, u_2, \dots, u_r \rangle$  de  $R_1$  et un s-uplet  $\langle v_1, v_2, \dots, v_s \rangle$  de  $R_2$  vérifiant :  
(iii.1) pour tous les  $i$  tels que  $\sigma(f_i) \in E$   $u_i = \sigma(f_i)$   
(iii.2) pour tous les  $j$  tels que  $\sigma(g_j) \in E$   $v_j = \sigma(g_j)$   
(iii.3) s'il existe  $i$  et  $j$  tels que  $h_i = \sigma(f_j)$  alors  $w_i = u_j$   
(iii.4) s'il existe  $i$  et  $k$  tels que  $h_i = \sigma(g_k)$  alors  $w_i = v_k$

### 4.3.2. Arbres composés

Pour construire le p-arbre composé de  $A_1$  et  $A_2$ , on fait en sorte que  $\sigma$  soit la substitution la plus générale possible.

A ce moment-là, le p-arbre composé  $A = \langle X, \delta, \text{cat}, \tau, F, t, R \rangle$  de  $A_1$  et  $A_2$  est construit comme au § 3.2.2, à ceci près que la condition (v) est remplacée par la condition suivante:

- (v)  $\forall x \in X_1 - \{a\} \quad \forall w \in \tau_1(x) \quad t(x, w) = \sigma(t_1(x, w))$   
 $\forall x \in X_2 - \{b\} \quad \forall w \in \tau_2(x) \quad t(x, w) = \sigma(t_2(x, w))$   
 $\forall w \in \tau_1(a) \quad t(a, w) = \sigma(t_1(a, w))$   
 $\forall w \in \tau_2(b) \quad t(a, w) = \sigma(t_2(b, w))$

Remarquons que la relation  $R$  peut vérifier la propriété suivante: l'existence d'un nombre  $k$  et d'une valeur  $v$  appartenant à  $E$  tels que tous les m-uplets  $\langle w_1, \dots, w_k, \dots, w_m \rangle$  de  $R$  vérifient  $w_k = v$ . Auquel cas on remplacera l'arbre  $A$  par un arbre  $A'$  comme indiqué dans la remarque de § 4.2.1. Ceci revient à remplacer  $F$  par  $F'$  obtenu en retirant de  $F$  tous les  $h_k$  correspondants,  $\sigma$  par  $\sigma'$  obtenue en remplaçant tous les  $\sigma(f_i) = h_k$  par  $\sigma'(f_i) = w_k$  (de même pour  $\sigma(g_j) = h_k$ ) et  $R$  par  $R'$ , relation extraite de  $R$  et portant sur les variables de  $F'$ .

### 4.3.3. Première illustration: la coordination en latin

On peut composer l'arbre de la figure 18 avec l'arbre de la figure 19 ci-dessous afin d'obtenir une coordination multiple<sup>16</sup>:

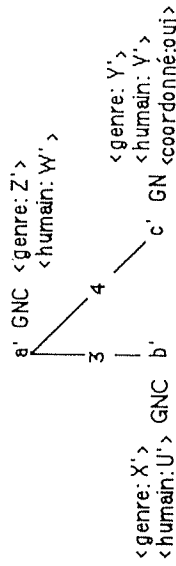


Figure 19

La substitution  $\sigma$  sera telle que  $\sigma(Y) = Z$  et  $\sigma(V) = W$ , les autres variables étant inchangées. On obtient alors:

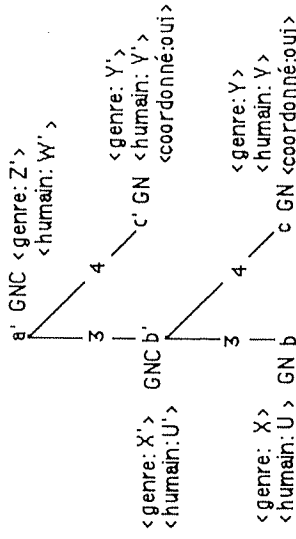


Figure 20

La nouvelle relation portant sur les variables peut être définie comme une conjonction de relations:

$$R(U, V, W, X, Y, Z, U', W', X', Z') = R(U, V, W, X, Y, Z) \wedge R(U', W', X', Z')$$

Cette relation est notamment satisfaite par le 10-uplet:

$\langle \text{non, non, non, masc, fem, neutre, non, non, masc, neutre} \rangle$ .

Ce qui correspond au GN coordonné (11):

(11) <i>Columnae</i>	<i>et</i>	<i>murus</i>	<i>et</i>	<i>porta</i>
<i>fem plu</i>		<i>masc sg</i>		<i>fem sg</i>
<i>colonnae</i>		<i>mur</i>		<i>porte</i>

Le phénomène d'émergence de valeur, et plus généralement d'interdépendance, n'est pas restreint aux cas qui impliquent un constituant coordonné; il est caractéristique de l'accord. L'accord par identité de valeur n'est de fait qu'un cas particulier. On se reportera sur ce point à Barlow et

16. La distinction entre coordination binaire et coordination multiple est analogue à celle que propose Gazdar et al. (1985, p. 171).

Ferguson (1987) qui illustrent le phénomène dans plusieurs langues ou à Pollard et Sag (1987) qui distinguent plusieurs types d'accord<sup>17</sup>.

#### 4.3.4. Deuxième illustration: les temps en français

Si le phénomène d'émergence et d'interdépendance caractérise l'accord, il ne lui est pas spécifique. La réalisation du temps en français, par exemple, présente des cas d'émergence d'une valeur.

La valeur de l'attribut temps associé au complexe "verbe auxiliaire + verbe principal" dans les temps dits composés ou surcomposés n'est pas identique à la valeur de l'attribut temps associé à l'auxiliaire; par contre, elle en dépend<sup>18</sup>. Ainsi la valeur présent composé (passé composé) dépend de la valeur présent associée à l'auxiliaire; la valeur imparfait composé (plus-que-parfait) de la valeur imparfait de l'auxiliaire; la valeur présent surcomposé ou imparfait surcomposé de la valeur présent composé ou imparfait composé respectivement de l'auxiliaire etc.

On notera qu'à la différence des tours coordonnés, la récursion est dans ce cas limitée; il n'y a pas de forme "sur-surcomposées" (des formes telles que: \**il a eu eu mangé*). Dans la solution que nous proposons, la récursivité est bloquée par la relation d'interdépendance: après un nombre fini de compositions, la relation d'interdépendance construite devient la relation vide.

On se donne le p-arbre élémentaire suivant pour traiter les formes composées<sup>19</sup>:

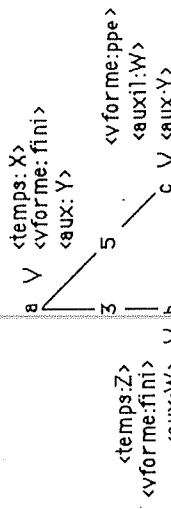


Figure 21

On définit la relation suivante entre W, Y, Z et X, qui est le produit d'une relation entre W et Y et d'une relation entre Z et X, toutes deux présentées sous la forme de tableaux:

(12)	W	Y	Z	X
	avoir	avoir	présent	présent-c
	avoir	être	imparfait	imparfait-c
	être	nil	prétérit	prétérit-c
	avoir	nil	futur	futur-c
			présent-c	présent-sc
			imparfait-c	imparfait-sc
			futur-c	futur-sc

17. Nous avons admis dans notre illustration la conception traditionnelle de l'accord comme une relation marquée sur les catégories. Pour une conception différente, voir Pollard et Sag (1987).

18. Nous reprenons la terminologie de Fradin (1989) pour nommer les temps; nous nous limitons pour cette illustration aux temps de l'indicatif.

19. Le trait auxil (dans la figure 21) permet de traiter la sélection de l'auxiliaire.

La relation établit toutes les formes composées en français. Les formes surcomposées sont obtenues en composant l'arbre de la figure 21 avec une copie de lui-même (une copie dans laquelle les symboles de variables sont changés). La composition donne l'arbre suivant:

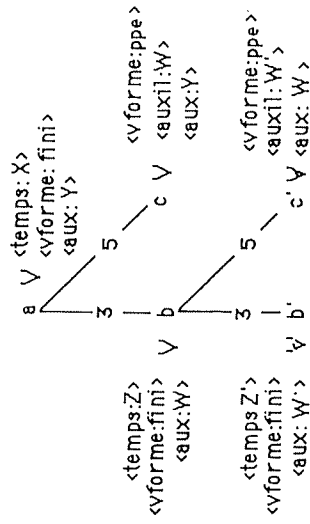


Figure 22

La nouvelle relation d'interdépendance, construite dans la composition, est donnée par les deux tableaux suivants:

(13)	W	Y	Z	X
	avoir	avoir	présent	présent-sc
	avoir	être	imparfait	imparfait-sc
	avoir	nil	futur	futur-sc
	avoir	être	nil	nil

Le 6-uplet <avoir, avoir, nil, futur, futur-c, futur-sc> appartient à cette relation; l'arbre de la figure 22 peut par exemple décrire la forme *[il] aura eu mangé*.

Si on essaie de réitérer la composition (autrement dit de composer l'arbre de la figure 22 avec une copie de l'arbre de la figure 21 repris ci-dessous (fig. 23), la relation résultante est la relation vide.

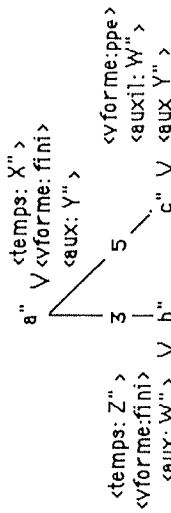


Figure 23

Il y a, en effet, deux possibilités:

- on fusionne les sommets b' et a": dans ce cas, les variables Z' et X' doivent être identifiées. Or, Z' a pour valeur un temps simple conformément à la relation d'interdépendance associée à l'arbre de la figure 22 et X' ne peut avoir pour valeur qu'un temps composé ou surcomposé selon la relation d'interdépendance associée à l'arbre de la figure 21,

- on fusionne les sommets a et b", dans ce cas, les variables X et Z" doivent être identifiées. Or, X prend pour valeur un temps surcomposé, ce qui est interdit pour Z".

#### 4.3.5. Commentaire

La relation d'interdépendance permet de capter des phénomènes qui ne se décrivent pas naturellement dans l'unification. La solution reste déclarative: elle n'oblige pas à recourir à des formalisations procédurales comme les règles de production ou à des extensions qui traitent de la détermination des valeurs dans les groupes coordonnés par généralisation (intersection). Cette dernière solution a été, par exemple, proposée par Karttunen (1986) pour la résolution de l'accord en personne. Outre le fait que cette solution ne s'étend pas naturellement au cas du genre, elle pose le problème de l'introduction et du déclenchement de cette opération.

On notera que le choix de représenter ces phénomènes par une relation est corrélé au choix de donner des valeurs atomiques et non structurées aux traits: la relation d'interdépendance ne pourrait pas être facilement définie sur des structures de traits.

### 5. LES ARBRES POLYCHROMES COMPACTS

#### 5.1. Motivation

##### 5.1.1. La profondeur des arbres

Certains arbres résultant de la composition d'arbres élémentaires telle qu'elle est définie au § 3 ci-dessus présentent une configuration particulière de branches et de sommets: un chemin de branche de même couleur relie des sommets étiquetés par la même catégorie. C'est le cas des arbres obtenus en ajoutant des expansions au groupe nominal (figure 11) ou celui des GN coordonnés (figure 20). Les différents niveaux reflètent le processus de composition de l'arbre alors même que l'on a une organisation paratactique: plusieurs groupes syntaxiques se succèdent dans la même position et dans un ordre quelconque (au regard de la syntaxe). Autrement dit, la hiérarchie correspondant à la composition des arbres élémentaires est dénuée de contenu descriptif ou structural.

On observe une situation similaire si on représente les configurations possibles de positions constitutives de GV par les arbres suivants (les arbres de la figure 24 sont simplifiés (les traits sont omis) et ne représentent que partiellement l'organisation de GV:

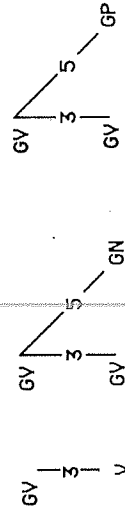


Figure 24

Par composition, on obtient les arbres (25.a) ou (25.b):

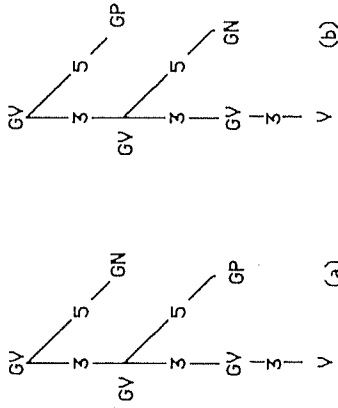


Figure 25

Les deux arbres (25.a) ou (25.b) correspondent aux énoncés *il a donné aux enfants du pain / il a donné du pain aux enfants*. Il est remarquable que la différence hiérarchique (entre 25.a et 25.b) n'est corrélée à aucune différence structurale ou interprétative; plus encore, la différence hiérarchique entre les compléments (dans chacun des arbres) n'est porteuse d'aucune signification<sup>20</sup>.

Dans les deux cas, la trace de la composition comporte plus de niveaux que la structure syntaxique n'en distingue<sup>21</sup>. L'opération de compactage que nous allons définir a pour effet de réduire le nombre de niveaux dans un arbre composé toutes les fois qu'un chemin de branches d'une couleur distinguée (la couleur 3 dans les exemples) relie des sommets portant la même étiquette catégorielle. Du coup, il faut admettre que l'arbre qui représente la structure n'est pas identique à l'arbre qui trace la composition (au sens du § 3) de la structure à partir des arbres élémentaires.

Notons que les tours que nous appelons paratactiques sont traités dans le cadre classique des grammaires syntagmatiques par des expressions régulières dans la partie droite des règles de réécriture. Une règle telle que  $GN \rightarrow N (GP + GA + GS)^*$  peut générer des arbres de la forme de celui qui est présenté en (fig 9). La solution que nous adoptons permet de maintenir les contraintes sur les traits sous la forme de relations d'interdépendance: il ne serait, en effet, pas évident de définir des relations sur un nombre illimité de variables de trait. Mais surtout, elle fournit un cadre pour le traitement de l'insertion intra-syntagmatique qui caractérise les tours incidents.

#### 5.1.2. Les tours incidents

Les tours incidents, que nous avons exemplifiés en (5), posent schématiquement le problème suivant:

- (i) le constituant incident est localisé à l'intérieur d'une structure,

20. La composition des arbres générant les positions de GV est contrainte par la sous-catégorisation des termes verbaux; nous laissons ce mécanisme de côté dans cet article.

21. Nous renvoyons à Milner (1984, 1991) pour une critique parallèle portant sur l'excessive richesse hiérarchique des représentations arborescentes décrites par les grammaires syntagmatiques.

## 5.2. Définitions

### 5.2.1. p-arbres à q-ième couleur principale

On se donne un nombre entier strictement positif  $p$  (nombre de couleurs de branches) et un nombre  $q$  strictement positif et inférieur ou égal à  $p$  (une couleur distinguée parmi les  $p$ ).

$A = \langle X, \delta, \text{cat}, \tau, t \rangle$  est un  $p$ -arbre à  $q$ -ième couleur principale, ou un  $\langle p, q \rangle$ -arbre, à condition que ce soit un  $p$ -arbre tel que<sup>23</sup> :

$$\forall x \in X \quad \delta_{1p}(x) \neq \varepsilon \Rightarrow |\delta_q(x)| = 1$$

(Si  $\alpha$  est un mot,  $|\alpha|$  est la longueur de  $\alpha$ .)

Propriété: Par composition de deux  $\langle p, q \rangle$ -arbres, on obtient un  $\langle p, q \rangle$ -arbre.

Par conséquent une base de  $\langle p, q \rangle$ -arbres élémentaires engendrera un ensemble de  $\langle p, q \rangle$ -arbres pauvres.

### 5.2.2. $\langle p, q \rangle$ -arbres compacts

Définition: Un  $\langle p, q \rangle$ -arbre  $A = \langle X, \delta, \text{cat}, \tau, t \rangle$  est compact si et seulement si il n'existe pas  $x$  et  $y$  appartenant à  $X$  tels que  $\delta_q(x) = y$ ,  $\delta_{1p}(y) \neq \varepsilon$  et  $\text{cat}(x) = \text{cat}(y)$ .

C'est-à-dire qu'il ne faut pas qu'il y ait un arc de couleur  $q$  dont l'origine et l'extrémité soient étiquetés par la même catégorie, sauf si cette extrémité est une feuille.

Définition: Un  $\langle p, q \rangle$ -arbre  $A' = \langle X', \delta', \text{cat}', \tau', t' \rangle$  est obtenu par compactage à partir d'un  $\langle p, q \rangle$ -arbre  $A = \langle X, \delta, \text{cat}, \tau, t \rangle$  si il existe  $x$  et  $y$  appartenant à  $X$  tels que  $\delta_q(x) = y$ ,  $\delta_{1p}(y) \neq \varepsilon$  et  $\text{cat}(x) = \text{cat}(y)$  (autrement dit  $A$  n'est pas compact) et si  $A'$  ne diffère de  $A$  que par :

- (i)  $X' = X - \{y\}$
- (ii)  $\delta'_q(x) = \delta_q(y)$
- (iii)  $\forall i < q \quad \delta'_i(x) = \delta_i(x) \delta_i(y)$
- (iv)  $\forall i > q \quad \delta'_i(x) = \delta_i(y) \delta_i(x)$

Remarque: Partant d'un  $\langle p, q \rangle$ -arbre pauvre, on peut de la sorte obtenir un arbre non pauvre.

Propriété: Par applications successives, dans un ordre quelconque, de l'opération de compactage à un  $\langle p, q \rangle$ -arbre  $A = \langle X, \delta, \text{cat}, \tau, t \rangle$  donné, on obtient un arbre compact  $A' = \langle X', \delta', \text{cat}', \tau', t' \rangle$  unique.

Preuve: Soit  $\langle x_0, x_1, \dots, x_m \rangle$  un chemin de  $A$  tel que :

- (i)  $\forall i \in \{0, 1, \dots, m-1\} \quad \delta_q(x_i) = x_{i+1}$
- (ii)  $\text{cat}(x_0) = \text{cat}(x_1) = \dots = \text{cat}(x_m)$
- (iii) si  $\delta_q(x_m) \neq \varepsilon$  alors  $\text{cat}(\delta_q(x_m)) \neq \text{cat}(x_m)$
- (iv) non  $\exists x_1 \quad x_0 = \delta_q(x_{-1})$  et  $\text{cat}(x_0) = \text{cat}(x_{-1})$

Il est clair que  $x_1, \dots, x_m$  n'appartiennent pas à  $X'$  et que  $x_0$  vérifie, quel que soit l'ordre d'application des opérations de compactage :

$$\begin{aligned} \delta'_q(x_0) &= \delta_q(x_m) \\ \forall i < q \quad \delta'_i(x_0) &= \delta_i(x_1) \delta_i(x_2) \dots \delta_i(x_m) \\ \forall j > q \quad \delta'_j(x_0) &= \delta_j(x_m) \delta_j(x_{m-1}) \dots \delta_j(x_1) \end{aligned}$$

- (ii) le constituant incident n'est pas en relation syntaxique ou sémantique avec les autres constituants de la structure,
- (iii) le nombre de constituants incidents n'est pas limité (c'est un trait caractéristique des tours incidents en général), ce qui est illustré en (14):

(14) *Pierre a, le con, tu le sais, vu le voleur sans rien faire.*

On peut décrire informellement ce tour comme un conflit de proximité:

- le constituant incident est spatialement plus proche de  $V$  qu'un tour sous-catégorisé,
- un tour sous-catégorisé est syntaxiquement et sémantiquement plus proche de  $V$  qu'un tour incident.

On a généralement admis que ces tours ne pouvaient pas être décrits par les règles qui génèrent les structures syntagmatiques<sup>22</sup>. Pourtant, ces tours présentent des propriétés qui les inscrivent minimalement dans l'ordre syntaxique: contrainte sur le placement, sur la catégorie des constituants, sur le rôle fonctionnel qu'ils jouent dans l'énoncé. Nous faisons l'hypothèse qu'ils sont susceptibles d'une analyse positionnelle: nous renvoyons à Cori et Marandin (1993) pour les arguments qui justifient cette analyse. Sur cette base, nous postulons des positions d'incidence et nous analysons les constituants incidents de (5) ou (14) comme occupant une telle position à droite de la position pivot de  $V$ . On se donne les deux arbres élémentaires (figure 26a et 26b) où la branche de couleur 4 est la position d'incidence; on obtient l'arbre (figure 26c) en composant (26a) et l'arbre (figure 22):

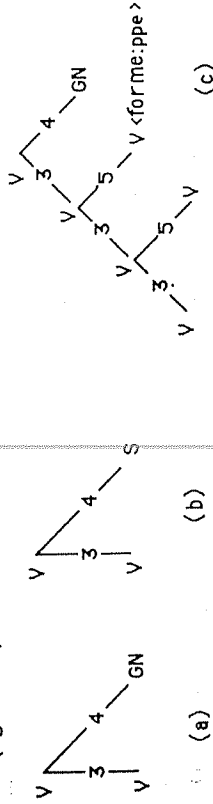


Figure 26

L'opération de compactage a pour effet de réordonner les branches; l'arbre (c) ci-dessus correspond à l'énoncé: *[lorsqu'il] a, le con, eu payé*; on obtiendrait avec les mêmes arbres élémentaires: *il avait, le con, payé* ou *il paie, le con*.

22. On a proposé le recours à des transformations spécifiques (Emonds), à une composante ad hoc (la composante 'stylistique'), à une redéfinition des arbres syntaxiques (cf les arbres 'discontinus' de McCawley 1982, 1987), à une différence de géométries (Milner 1989).

23. On se place dans le cas où il n'y a pas de relation d'interdépendance, afin d'alléger l'écriture.

TAG). Enfin, l'opération de compactage met cruciallement en jeu la couleur de la branche, ce qui permet de restreindre l'opération à un seul type de position.

Par ailleurs, l'opération de compactage ne présuppose aucune partition sur la base d'arbres élémentaires: la base d'arbres élémentaires d'une GAP est unique. La partition entre arbres élémentaires et initiaux auxiliaires dans une TAG renvoie à la distinction ajout/non-ajout; cette distinction, traitée dans une TAG comme une différence dans les primitifs syntaxiques, relève pour une GAP de l'appariement position/rôle<sup>24</sup>.

### 5.3.2. La nature différente des branches

Le compactage repose sur les caractéristiques définitoires des arbres polychromes:

- la différenciation des branches,
  - la distinction entre catégories et traits.
- Il requiert une identité de catégories et de branches sans rien imposer sur les structures de traits. L'hypothèse (conforme à l'approche non lexicale adoptée ici) est en effet la suivante: les propriétés des termes n'interviennent pas dans l'aspect purement géométrique des configurations syntaxiques. Par contre, il implique cruciallement que l'on puisse distinguer une position (la branche q).

La position distinguée (q) prend un caractère qui l'oppose à toutes les autres. Son interprétation naturelle est celle de pivot dans la structure syntaxique. Dans un groupe syntagmatique (groupe nominal, verbal,...), ce sera la position occupée par la tête lexicale. Cela revient à donner une définition purement formelle de la nucléarité syntagmatique. Cette conception s'oppose très directement à la version lexicale (par exemple de X-barre) où le caractère nucléaire des groupes syntagmatiques est rapporté à un trait substantiel (le trait lexical nom, verbe, etc. identique dans la catégorie lexicale et la catégorie syntagmatique).

Dans un GN coordonné analysé par l'arbre de la figure 19, la position pivot est occupée par le premier constituant du groupe<sup>25</sup>. Cette analyse implique une dissymétrie entre les constituants et, du point de vue de l'appariement position/rôle fonctionnel, que le premier constituant a un statut distingué. Plusieurs indices rendent l'analyse plausible: (entre autres) le premier constituant peut être le seul impliqué par l'accord (14.a) et, dans les coordinations qui le permettent, il ne peut être l'objet d'ellipse (14b vs 14c).

24. Les p-arbres élémentaires sont de profondeur 1 alors que les arbres élémentaires dans une TAG sont de profondeur arbitraire. On notera sur ce point que l'une des justifications avancées pour l'introduction d'arbres de profondeur supérieure à 1 est le traitement des expressions dites figées (Abeillé, 1989); cet argument n'a pas de force dans un cadre auto-structural où le phénomène est réputé relever du module interprétatif. Les expressions dites figées sont, en effet, syntaxiquement régulières (autrement dit, il n'y a pas de caractérisation syntaxique du "figement") et se spécifient au niveau interprétatif par un défaut de compositionnalité (Abeillé, ibid.) ou plutôt par des mécanismes de construction référentielle qui leur sont spécifiques.

25. La position pivot est la branche gauche en français, il n'est pas évident qu'il en soit de même en latin. Nous laissons la question ouverte dans cet article.

On voit que les chemins de ce type sont "contractés" et remplacés par un arc unique. Dans ce processus, les traits associés aux sommets intermédiaires (c'est-à-dire  $x_1, \dots, x_{m-1}$ ) disparaissent. Il en découle que ces traits n'ont pas de valeur descriptive, mais une valeur constructive.

Exemple: Partant des arbres (b) et (c) de la figure 25, et considérant que la couleur principale est 3, on obtient les p-arbres respectifs suivants, qui effectivement ne sont plus pauvres.



Figure 27

### 5.2.3. Nouvelle définition de la composition

La présentation que nous avons adoptée dans cet article semble indiquer que la construction des arbres syntaxiques s'effectue en deux phases: d'abord la composition, puis le compactage. Or, il revient au même (en raison de la propriété énoncée en 5.2.2.) d'effectuer simultanément ces deux opérations: la composition inclut alors le compactage. Nous employons dorénavant *composition* avec ce sens.

L'opération présente deux cas. Le premier est équivalent à une substitution. Le second correspond à une composition qui s'effectue selon un sommet (une feuille) qui occupe la position pivot et qui est étiquetée par la même catégorie que son père.

En définitive, une GAP  $G = \langle \text{Cat}, T, E, \Phi, B \rangle$  engendre un ensemble d'arbres  $Lc(B)$ , obtenus par composition avec effet de compactage: on n'admet pas dans  $Lc(B)$  les arbres non compactés. La composition fait sortir les grammaires d'arbres polychromes de la classe des grammaires indépendantes du contexte sans qu'il y ait recours à des règles contextuelles ou que cela soit dû aux structures de traits. (cf. Cori et Marandin, 1994).

## 5.3. Commentaires

### 5.3.1. Composition et adjonction

Le compactage ne fait pas de la composition une opération analogue à l'opération d'adjonction qui est à la base des opérations d'arbres adjoints (TAG, Joshi, 1985), même si ces deux opérations mettent en jeu des arbres qui présentent une configuration similaire: une même catégorie étiquette la racine et un des fils.

L'opération de compactage réduit le nombre de niveaux hiérarchiques alors que l'adjonction l'augmente; le compactage intervient sur les feuilles et non à l'intérieur des arbres comme dans l'adjonction. On notera, sur ce dernier point, que cela permet de conserver une structure de traits et une opération d'unification la plus simple possible (au contraire de l'organisation bi-partite (top-bottom) requise par l'adjonction dans une



- (14) a. *Jean et ses manies m'a toujours agacé.* (ex. de M. Gross).  
 b. *Pierre boit de l'eau et Marie du vin.*  
 c. \**Pierre de l'eau et Marie boit du vin.*

A la lumière de ces éléments d'analyse, on peut reprendre la question de l'appariement position-rôle fonctionnel que nous concevons comme la mise en relation de deux niveaux distincts de représentation. Nous avons associé à la branche (q) un rôle générique: distinguer un des termes; nous lui avons aussi associé des rôles fonctionnels distincts selon les structures: distinguer la tête lexicale dans un syntagme, distinguer un terme dominant dans un tour coordonné. L'appariement est moins simple pour les autres couleurs de branche: il ne paraît pas possible de leur associer un rôle générique. On doit alors procéder position par position, c'est-à-dire considérer chacune des positions définies par le couple <catégorie de la racine, couleur> (cf. § 2.2.1). La question est la suivante: étant donné une position <X,n>, peut-on lui associer un rôle fonctionnel qui ne soit pas induit par la catégorie ou une propriété lexicale du terme qui l'occupe? Les résultats de ce programme d'observation ne peuvent pas être exposés ici faute de place. Notons qu'il requiert que toutes les positions, et pas seulement la seule position noyau, soient distinguées; c'est ce que nous donne le formalisme que nous proposons.

Si l'opération de compactage nécessite que soit distinguée une position, elle crée en retour de l'indistinction. Reprenons un arbre compacté comme celui de la figure 9.

Dans cet arbre, une même position (la position 5) se trouve "démultipliée": la position est occupée par plusieurs unités contiguës ne formant pas entre elles une unité. C'est le cas, par exemple, de la position droite dans le groupe nominal. On admettra le fait d'être plurioccupable au nombre des propriétés d'une position. Les positions plurioccupables constituent des structures syntaxiquement sous-spécifiées.

Les éléments paratactiques, en effet, ne peuvent pas être distingués par la position qu'ils occupent puisqu'ils occupent la même position. Ils ne se distinguent donc que par leur ordre dans la chaîne. Dans ce cas, la syntaxe (en tant qu'organisation de positions) ne fournit pas de principe de linéarisation ou d'interprétation des constituants.

Si on appelle *géométrie* les organisations de positions, les arbres compactés présentent une des caractéristiques mise en avant par Milner (1991): ce sont des géométries pauvres. Elles ne distinguent que peu de niveaux hiérarchiques et elles y ménagent des zones d'indistinction.

## 6. CONCLUSION

Les grammaires d'arbres polychromes se caractérisent par:  
 - (i) l'introduction des couleurs de branches dans la définition des arbres syntaxiques,

- (ii) l'opération de composition avec effet de compactage.  
 Les couleurs offrent un principe de distinction systématique des positions entre elles, des positions et de ce qui les occupe. Le principe de distinction est strictement positionnel: les positions sont repérées par rapport à la position pivot et déterminent le placement des constituants sur

l'axe linéaire<sup>26</sup>. L'opération de composition rend compte des relations de proximité; elle permet, en particulier, de traiter le conflit de proximité caractéristique des tours incidents de manière monostratale. Par ailleurs, elle a un effet d'indistinction positionnelle.

Couleurs et composition sont solidaires. C'est un trait important au regard de la définition de la syntaxe que nous avons admise.

Si l'étiquetage des sommets met en jeu des traits, il est nécessaire, pour rendre compte du phénomène d'interdépendance (émergence) de valeur, d'introduire:

- (iii) une augmentation de l'opération d'unification.

Les GAP nous paraissent un cadre formel explicite pour l'hypothèse positionnelle et, donc, un des vecteurs de sa fausification empirique. L'intégration d'une GAP dans un cadre grammatical autolexical permettra d'apparier les représentations des organisations syntaxiques avec leurs représentations sémantiques.

## REFERENCES

- ABEILLÉ, A. (1989) : "L'unification dans une grammaire d'arbres adjoints: quelques exemples en syntaxe française", *T.A. Informations*, vol. 30, n° 1-2 Klincksieck, pp. 69-112.
- BARLOW, M. & FERGUSON, C., (eds), (1987) : *Agreement in Natural Language*, Stanford, CSLI, University of Chicago Press.
- BENVENISTE, E. (1974) : "Formes nouvelles de la composition nominale", *Problèmes de linguistique générale*, Gallimard, pp. 163-176.
- BERGE, C. (1970) : *Graphes et hypergraphes*, Paris, Dunod.
- CHOMSKY, N. (1975) : "Remarque sur les nominalisations", *Questions de sémantique*, Paris, Le Seuil, pp. 73-131 (trad. fr.).
- CODD, E.F. (1970) : "A relational Model of Data for Large Shared Data Banks", *Communications of the ACM*, Vol. 13, n° 6.
- CORBETT, G. (1983) : "Resolution rules: agreement in person, number and gender", *Order, Concord and Constituency*, Gazdar G. et al. (eds.), Dordrecht, Foris Publications, pp. 175-206.
- CORI, M. & MARANDIN, J.-M. (1993) : "Discontinuity without Discontinuous Trees", ms. Université de Paris 7.
- CORI, M. & MARANDIN, J.-M. (1994) : "Polychrome Tree Grammars (PTG): a formal approach", *Current Issues in Mathematical Linguistics*, C. Martin-Vide (ed.), Amsterdam, Elsevier.
- CORI, M. & MARANDIN, J.-M. (soumis à publication) : "Préférence in NL Syntax".
- EMONDS, J. (1985) : *A unified Theory of Syntactic Categories*, Dordrecht, Foris Publications.

26. On comparera ce choix, qui est lié à la définition restrictive de la syntaxe que nous mettons en oeuvre, avec le choix opéré dans d'autres cadres grammaticaux, HPSG, par exemple, où le typage des branches dans la définition des signes syntagmatiques réfère à un principe fonctionnel assez proche de celui qui est sous-jacent à l'étiquetage de X-barre: branche-Tête, branche-Complément, branche-Ajout, branche-Marqueur (Pollard & Sag, 1987, 1994).

## LECTURES

rubrique préparée par Jacques ANIS\*

## OUVRAGES REÇUS

- DE RIJKE Maarten (Ed.) (1993), *Diamonds and Defaults*, Studies in Pure and Applied Intensional Logic, Kluwer Academic Publishers, Boston/Dordrecht/Londan.
- FUCHS C. (1993), *Linguistique et Traitements/automatiques des Langues*, avec la collaboration de DANLOS L., LACHERET-DUJOUR A., LUZZATID. et VICTORRI B., Hachette, Paris.
- HUTCHINS W. J. & SOMERS H. L. (1992), *An Introduction to Machine Translation*, Academic Press.
- KÖHLER R. & RIEGER B. (Eds) (1993), *Contributions To Quantitative Linguistics*, Proceedings of the First International Conference On Quantitative Linguistics, Qualico, Trier, 1991, Kluwer Academic Publishers, Boston/Dordrecht/London.
- LEERMAKERS R. (1993), *The Functional Treatment of Parsing*, avant-propos de F. Pereira, Kluwer Academic Publishers, Boston/Dordrecht/London.

## NOTES

- \* NIRENBURG S., CARBONNEL J., TOMITA M. & GOODMAN (1992), *Machine Translatic. A Knowledge-Based Approach*, Morgan Kaufmann, San Mateo, Ca. Un récent numéro de notre revue rendait compte de GOODMAN & NIRENBURG, *The KBMT Project: A Case Study in Knowledge-Based Machine Translation*. L'ouvrage de Nirenburg et al. s'inscrit bien sûr dans le même cadre théorique mais a un caractère moins technique. Le premier chapitre, "MT in a Nutshell", est consacré à un rappel historique, à une évocation des problèmes généraux de la T.A. et à une présentation générale de l'approche KBMT. Celle-ci est définie par l'imperatif de la compréhension complète du texte source, le choix de l'interlingua ou langage-pivot pour la représentation du sens - le contraste transfert-interlingua est néanmoins relativisé par l'évolution vers le "deep transfer" - et l'appel à des bases de connaissances intégrant un modèle du monde. Ces aspects programmiques et leur discussion à l'intérieur de la communauté de la T.A. sont développés dans le deuxième chapitre "Treatment of Meaning in MT Systems". Le chapitre suivant décrit les diverses composantes de l'interlingua, notamment à travers le langage Tamerlan, qui utilise un formalisme de frames. Le chapitre 4 traite de la

\* Université de Paris-X Nanterre, Département de Linguistique et Centre de Recherches Linguistiques, 200 avenue de la République, 92001 NANTERRE Cedex.

- FRADIN, B. (1989) : "Sur la nécessité d'un constituant AUX dans la grammaire du français", *Actes du 19ème Congrès International de Philologie et Linguistique Romane*, Saint-Jacques de Compostelle.
- GAZDAR, G., KLEIN, E., PULLUM, G., SAG, I. (1985) : *Generalized Phrase Structure Grammar*, Oxford, Basil Blackwell Publisher.
- GLADKIL, A. (1970) : *Leçons de linguistique mathématique*, Dunod.
- JACKENDOFF, R. (1977) : *X-bar Syntax: a Study of Phrase Structure*, Cambridge, Cambridge UP.
- JOSHI, A. (1985) : "Tree adjoining grammars: How much context-sensitivity is required to provide reasonable structural descriptions?", *Natural Language Parsing*, Dowty D., Karttunen L., Zwicky A. (eds.), Cambridge, Cambridge UP, pp. 206-250.
- KARTTUNEN, L. (1986) : "Features and Values", Actes 10e COLING, Stanford, 1984. Repris dans *A Compilation of Papers on Unification Based Formalisms*, Stanford, CSLI.
- KERLEROUX, F. (1990) : "Du mode d'existence de l'infinifit substantivé en français contemporain", *Cahiers de grammaire*, Université de Toulouse-Le Mirail, n°15, pp. 57-99.
- KERLEROUX, F. (1991) : "Il est d'un calme et d'un élégant! Un phénomène de distorsion", *BULAG*, Université de Franche-Comté, n°17, pp. 79-116.
- KORNAL, A. & PULLUM G. (1990) : "The X-bar theory of phrase structure", *Language*, n° 66, pp. 24-50.
- LYONS, J. (1970) : *Linguistique générale* [trad. fr.], Paris, Larousse.
- McCRAWLEY, J. (1982) : "Parentheticals and Discontinuous Constituent Structure", *Linguistic Inquiry*, n°13, pp. 91-106.
- McCRAWLEY, J. (1989) : "Individuation in and of Syntactic Structures", *Alternative Conceptions of Phrase Structure*, Baitin, M. & Kroch, A. (eds.), Chicago, University of Chicago Press, pp. 117-138.
- MILNER, J-C. (1984) : "De l'inutilité des arbres en linguistique", *TA Informations*, vol 25, n° 1, Klincksieck, pp. 5-12.
- MILNER, J-C. (1989) : *Introduction à une science du langage*, Le Seuil.
- MILNER, J-C. (1991) : "Géométries", *Le gré des langues*, vol 2, L'Hamattan, pp. 8- 26.
- POLLARD, C & SAG, I. (1987) : *Information-based Syntax and Semantics*, vol. 1, Stanford, CSLI.
- POLLARD, C & SAG, I. (1994) : *Head-driven Phrase Structure Grammar*, Chicago, University of Chicago Press.
- PULLUM, G. (1991) : "English nominal gerund phrases as noun phrases with verb-phrase heads", *Linguistics*, n° 29, pp. 763-799.
- SADOCK, J. (1991) : *Autolexical Syntax*, Chicago: University of Chicago Press.
- WEHRLI, E. (1988) : "Parsing with a GB-Grammar", Reyle U. et Rohrer C., (eds.), *Natural Language Parsing and Linguistic Theory*, Dordrecht, D. Reidel, pp. 177-201.