

Université Paris 7
UFR de Linguistique
Laboratoire d'Automatique Documentaire et Linguistique

Une grammaire lexicalisée d'arbres adjoints pour le français

Application à l'analyse automatique

Anne ABEILLÉ

Thèse de Doctorat de linguistique

Directeur: Maurice Gross

Janvier 1991

Jury

G. Bès
L. Danlos
M. Gross
~~A.K. Joshi~~
J-C. Milner
G. Sabah

Remerciements

Je remercie Gabriel Bès, Jean-Claude Milner et Gérard Sabah d'avoir bien voulu faire partie du jury.

Je tiens à remercier Aravind Joshi qui m'a accueilli dans son équipe à Philadelphie en 1987-88, m'a fait découvrir les grammaires d'arbres adjoints, et qui a entretenu depuis une collaboration régulière entre Penn et le LADL.

Je suis également très reconnaissante à Laurence Danlos de l'intérêt qu'elle a porté à mon travail depuis le début et de la confiance qu'elle m'a témoignée en me confiant un cours de maîtrise sur les grammaires d'unification et les grammaires d'arbres adjoints. Sa relecture attentive m'a été précieuse.

Je remercie les membres du LADL qui m'ont aidée à différentes étapes de mon travail, notamment A. Guillet, E. Laporte, C. Leclère et M. Silberstein, et particulièrement B. Habert qui a bien voulu relire cette thèse. Je remercie également pour leur accueil mes collègues de l'UFRL, et particulièrement Danièle Godard pour sa relecture critique et ses nombreux conseils.

Que les membres des départements d'informatique et de linguistique de l'université de Pennsylvanie avec qui j'ai pu avoir des échanges fructueux soient ici remerciés, en particulier Anthony Kroch, Sharon Cote, Robert Frank, Mitch Marcus, Jack Hoeksema, M. Steedman et K. Vijay-Shanker.

J'ai une dette particulière envers Yves Schabes. Cette thèse doit beaucoup au travail effectué en commun à partir de 1987.

Enfin, j'ai grand plaisir à exprimer ici ma reconnaissance à mon directeur Maurice Gross qui m'a formée, guidée et m'a soutenue à toutes les étapes de mon travail. C'est sur ses conseils que je suis allée passer un an à Philadelphie.

Je remercie également C. Marchello-Nizia pour m'avoir encouragée au départ à me spécialiser en linguistique informatique, ainsi que François-Xavier Testard-Vaillant pour m'avoir donné accès à une machine SUN et pour son aide en Common Lisp.

Enfin, sans la fondation Jean Walter Zellidja dont la bourse a permis de financer une partie de mon séjour aux Etats Unis, cette thèse n'aurait pu voir le jour.

24. presence of in fiction.

NOTATIONS UTILISEES

Notations générales :

- linguistiques

- * phrase inacceptable
- # phrase inacceptable dans l'interprétation idiomatique
(*Jean + Marie*) *part* : notation factorisée pour les deux phrases *Jean part* et *Marie part*
- E élément vide
Jean mange (E + une pomme) : notation factorisée pour les deux phrases *Jean mange* et *Jean mange une pomme*
- N₀ premier argument (sujet)
- N_i argument nominal
- P_i argument phrastique

- formelles

- U unification
- * étoile de Kleene
- / disjonction
- => implication
- ~ négation
- ° noeud de furcation (chapitre 2)

Notations pour les grammaires d'arbres adjoints :

- dans les arbres :

- ↓ marque les noeuds à substituer
- * marque les noeuds pieds
- (X) marque que le noeud X (et le sous-arbre qu'il domine) est optionnel .
- (*) noeud pied avec variation d'ordre des mots (chapitre 6)
- ◇ marque la tête de l'arbre
- e chaîne vide

- dans le lexique :

- [xx] représente n'importe quelle forme fléchié (ou élidée) du lemme xx
- / marque la disjonction (sur des traits ou sur des formes lexicales)
- => implication (sur des traits)
- ~ négation (sur des traits)

TABLE DES MATIERES

Introduction.....	3
Chapitre I. Le cadre théorique.....	12
1. Présentation des grammaires d'arbres adjoints.....	12
2. Pour une lexicalisation intégrale de la grammaire.....	29
3. Sur la notion d' arbre.....	45
Chapitre II. Justification des syntagmes employés	58
1. La notion de phrase.....	60
2. Pourquoi nous n'avons pas de syntagme verbal.....	63
3. La notion de syntagme nominal.....	87
4. Les syntagmes prépositionnels.....	99
5. Autres syntagmes.....	102
Chapitre III. Les structures de base de la grammaire d'arbres adjoints du français.	104
1. Les différents types d'arbres élémentaires.....	104
1.1. Les verbes.....	105
1.2. Les clitiques.....	120
1.3. Les prépositions.....	123
1.4. Les noms.....	126
1.5. Les déterminants.....	132
1.6. Les adjectifs.....	138
1.7. Les adverbes.....	140
2. Dictionnaire morphologique et dictionnaire syntaxique.....	142
3. Les "familles" d'arbres.....	144
3.1. Les TAGs "auto-synchrones".....	144
3.2. Métarègles et règles lexicales.....	148
3.3. Les classes du LADL.....	151
3.4. Familles d'arbres pour les phrases simples.....	154
3.5. Familles d'arbres pour les ajouts.....	167
Chapitre IV. les phénomènes de semi- (ou non-) compositionnalité.....	170
1. Les phrases à verbe support.....	170
1.1. Discussion théorique.....	170
1.2. représentation en TAG.....	173
2. Les expressions figées.....	200
2.1. Discussion théorique.....	200
2.2. représentation en TAG.....	205

Chapitre V. Les phénomènes d'extraction.....	228
1. Les contraintes d'extraction.....	228
1.1. Discussion théorique.....	228
1.2. Représentation en TAG.....	232
2. Les "dépendances non bornées".....	247
2.1. Discussion théorique.....	247
2.2. Représentation en TAG.....	252
Chapitre VI. Recherches en cours.....	274
1. Les "ajouts-arguments"	274
2. Phénomènes de contrôle.....	279
3. Variations d'ordre des mots.....	285
4. Cas de coordination.....	289
Chapitre VII. L'analyse automatique.....	294
1. Quelques analyseurs existants.....	296
2. La stratégie d'analyse adoptée.....	299
3. Le traitement des ambiguïtés.....	312
4. L'implémentation actuelle.....	316
5. Autres applications.....	318
Conclusion.....	324
Bibliographie.....	325
Annexes.....	336
Annexe 1. Traits et catégories utilisés.....	336
Annexe 2. Liste des familles d'arbres.....	338
Annexe 3. Extraits des fichiers dictionnaires.....	377

Introduction

Le traitement automatique des langues connaît aujourd'hui un grand développement et ses applications vont se diversifiant. La plupart de ces applications, des plus simples (correction d'orthographe, indexation automatique) aux plus ambitieuses (résumé, traduction automatique), ont besoin d'un module d'analyse syntaxique.

L'analyse syntaxique automatique nécessite des dictionnaires et des grammaires d'une précision et d'une exhaustivité non imaginées jusque là. Ainsi, les dictionnaires du commerce sont inadéquats pour constituer des dictionnaires électroniques exploitables de façon automatique (Cf M. Gross 1989). Le lien entre grammaire et dictionnaire doit aussi être rendu complètement explicite.

Des dictionnaires électroniques spécifiques sont en cours de constitution pour le français au LADL, mais les connaissances syntaxiques correspondantes (les lexiques-grammaires) n'ont pas donné lieu à une grammaire entièrement formalisée et directement exploitable sur machine.

Une grammaire informatique doit être formalisée et plus elle se rapproche d'un système formel connu et restrictif (grammaires régulières, grammaires hors contexte), plus le programme d'analyse sera aisé à écrire. Seuls des systèmes formels à la capacité générative bien définie et suffisamment contrainte peuvent permettre la réalisation d'algorithmes d'analyse décidables et de programmes efficaces. Mais une telle grammaire doit également avoir la couverture linguistique la plus étendue possible compte tenu de la grande variété lexicale et syntaxique rencontrée dans les textes¹.

Il existe de nombreux systèmes d'analyse, pour le français, dont aucun n'est entièrement satisfaisant. La plupart privilégient l'objectif d'efficacité en s'en remettant à des formalismes rudimentaires mais éprouvés, qui rendent difficiles ensuite les descriptions de langue fines et empêche donc l'obtention d'une bonne couverture syntaxique. La plupart ont des grammaires de couverture restreinte, toutes ont un petit dictionnaire.

Nous concevons l'analyse syntaxique comme un module commun pour des applications diverses, qui peut être constitué de façon indépendante du domaine d'application, ou de la nature de l'application.

Nous avons écrit une grammaire du français en nous fondant sur un formalisme particulier : les grammaires d'arbres adjoints (*Tree adjoining grammars* ou TAG). Il s'est agi pour nous d'explorer les potentialités de ce formalisme, défini à l'origine par Joshi et al. en 1975 et basé sur des combinaisons de structures arborescentes, pour la description

¹ Les grammaires génératives ne répondent ni au premier objectif ni au second: elles ne correspondent pas à un système formel aux propriétés informatiques intéressantes et ne décrivent qu'un faible nombre d'exemples dont la représentativité n'est pas assurée.

et l'analyse syntaxique du français. Nous nous sommes parallèlement appuyée sur la base de données linguistiques extensive que constituent les lexiques-grammaires du LADL pour le français et qui sont reliés à un dictionnaire de grande envergure. Le lexique-grammaire ne constitue pas une grammaire en tant que telle, mais organise les données qu'il recense selon des notations dont la consistance et la cohérence sont assurées. L'intérêt théorique est double :

- d'un point de vue linguistique, les grammaires TAG redéfinissent la notion de localité et permettent de reconsidérer les phénomènes d'interaction entre lexique, syntaxe et sémantique. Elles sont particulièrement bien adaptées à la formalisation des hypothèses qui sous-tendent les lexiques-grammaires définis au LADL et peuvent contribuer à les préciser.

- d'un point de vue informatique, les grammaires TAG, tout en étant plus puissantes que les grammaires hors contexte, ont cependant un pouvoir génératif suffisamment limité pour garder des propriétés computationnelles intéressantes et permettre des algorithmes de traitement efficaces.

Les grammaires TAG se situent dans la lignée des nouveaux formalismes grammaticaux qui ont été conçus à la suite de critiques théoriques contre la grammaire générative transformationnelle et qui témoignent d'un souci plus marqué pour les applications informatiques. Ces formalismes se sont développés depuis le début des années 80 d'abord aux Etats-Unis puis en Europe. Ces systèmes sont beaucoup moins puissants que le modèle chomskyen, et contraignent davantage les descriptions linguistiques (offrant ainsi moins de champ à l'arbitraire du linguiste). Ils permettent également une meilleure intégration des aspects lexicaux, syntaxiques et sémantiques grâce à une structure de données beaucoup plus souple.

Nous allons présenter brièvement les caractéristiques essentielles de ces nouveaux formalismes avant de préciser ce qui fait l'originalité de celui que nous utilisons.

On regroupe souvent ces formalismes sous le nom de grammaires d'unification (*unification-based formalisms*). Ce terme recouvre en fait des situations hétérogènes : certains formalismes sont seulement des métalangages commodes pour la programmation informatique, parfois directement exprimés dans les termes d'un langage de programmation (*Grammaires de clauses définies*). D'autres sont liés à une véritable théorie linguistique, accompagnée d'études sur des phénomènes et des langues variés (*Grammaire lexicale fonctionnelle*). Et il existe nombre de situations intermédiaires (ex: *Grammaires syntagmatiques généralisées*). Certains points communs néanmoins conduisent à les regrouper.

Dans leur conception même, il s'est agi de réagir contre la grammaire générative des années 70, dont Peters et Ritchie avaient très tôt démontré le caractère trop puissant (S. Peters et R. Ritchie 1968). L'appareil formel trop puissant de la Théorie Standard Etendue permet d'engendrer en fait n'importe quel langage récursivement énumérable et non les seules langues humaines. Il était donc difficile d'écrire des grammaires pour

ordinateur sur ce modèle et de définir des algorithmes efficaces².

D'un point de vue plus linguistique, M. Gross réagissait en 1975 contre la situation contemporaine en grammaire générative, caractérisée par "une disproportion énorme entre l'ensemble des possibilités théoriques et celui des faits linguistiques" (p. 45). Le modèle générativiste permet en effet pratiquement toujours plusieurs solutions concurrentes "entre lesquelles il semble difficile d'opérer un choix non arbitraire". M. Gross a donc réorienté les recherches linguistiques au LADL vers l' "accumulation de données empiriques aussi complètes et détaillées que possible", en partant d' "un modèle de langage aussi limité que possible et peu formalisé", suivant en cela Z. Harris 1951.

Parallèlement, ces nouveaux formalismes ont repris la double ambition de la grammaire générative à ses débuts : caractériser l'ensemble des phrases bien formées d'une langue et seulement celles-là, caractériser la classe des langues naturelles dans l'ensemble de tous les langages possibles, ambitions non réalisées par la Théorie Standard Etendue.

Ces grammaires résultent de la convergence de travaux en linguistique (analyses syntaxiques "sans transformations"), en logique (démonstration automatique de théorème) et en informatique (programmation logique, langage Prolog) qui ont abouti à la définition d'un certain nombre de langages de description linguistique. Le point commun de ces formalismes est de permettre des descriptions :

- déclaratives, et non procédurales (à la différence des réseaux de transitions). L'écriture de la grammaire peut donc être relativement indépendante de celle de l'analyseur, qui peut être de divers types : montant, descendant, mixte...
- attachées à la description directe de la surface, c'est-à-dire basées sur un seul type de représentation en général.
- basées sur des structures de traits (associant à des attributs des valeurs appartenant à un ensemble prédéfini et structuré) combinées par unification. Ces structures de traits sont un moyen d'unifier la représentation des différents niveaux linguistiques, puisque les attributs peuvent être phonologiques, syntaxiques ou sémantiques.
- monotones: les règles ne peuvent qu'ajouter, non retirer de l'information (morphologique, syntaxique, sémantique).

Si l'on se limite aux formalismes qui incorporent une théorie linguistique originale, l'on retient trois théories principales:

- la *grammaire syntagmatique généralisée* (*Generalized Phrase Structure Grammar* ou GPSG)
- la *grammaire lexicale fonctionnelle* (*Lexical Functional Grammar* ou LFG)
- la *grammaire syntagmatique guidée par les têtes* (*Head Phrase Structure Grammar* ou HPSG; issue au départ de GPSG)

On peut mentionner également la *grammaire d'unification fonctionnelle* (*Functional Unification Grammar*; M. Kay 1984) qui n'a pas été utilisée pour écrire une grammaire

² Voir cependant l'analyseur de S. Kuno 1965, et les réseaux de transition augmentés qui donnaient une approximation d'une analyse de surface, l'analyse transformationnelle restant le point délicat de l'analyse, car d'une complexité exponentielle.

de couverture significative, à notre connaissance, ainsi que la *grammaire catégorielle d'unification* (*Unification Categorical Grammar*, J. Calder et al. 1988; utilisée pour le français par G. Bès et al. 1989), et la *grammaire d'unification catégorielle* (*Categorical Unification Grammar*, L. Karttunen 1989), également en cours de développement.

Nous allons reprendre les avantages linguistiques et informatiques qu'il y a à utiliser de tels formalismes, avant de dire en quoi nous nous séparons des grammaires d'unification mentionnées ci-dessus et pourquoi nous avons choisi les grammaires d'arbres adjoints.

a) Intérêt linguistique

Les grammaires d'unification ont redéfini des notions classiques comme celles de catégorie (GPSG), de fonction grammaticale ou de structure argumentale (LFG), donné des explications nouvelles à des phénomènes linguistiques autrement décrits en grammaire générative (dépendances non bornées, coordination, relations anaphoriques...), et exploré des phénomènes peu étudiés jusque là (infinitives en hollandais, système casuel en islandais ...).

Leur intérêt majeur est d'enrichir les descriptions linguistiques que permettent les grammaires de réécriture hors contexte (*context-free*, CF). Celles-ci étaient la cible de plusieurs critiques à partir de *Structures Syntaxiques* (N. Chomsky 1957) :

- 1- capacité générative forte insuffisante: impossible de représenter les dépendances croisées, qu'on trouve par exemple dans les constructions avec *respectivement*.
- 2- mauvaise interaction du lexique et de la syntaxe, d'où une multiplication "explosive" du nombre des catégories préterminales (pour l'accord, la sous-catégorisation ...)
- 3- difficulté à représenter les constituants mobiles ou discontinus.
- 4- manque de généralisations linguistiques, pas de lien par exemple entre phrases active et passive.

Le premier point demeure, quoique assez marginal. C'est sur les autres points que porte la nouveauté.

L'interaction entre lexique et syntaxe est en revanche améliorée: à partir du moment où chaque relation de dépendance est représentée par un trait spécifique, les catégories syntaxiques proprement dites peuvent rester en petit nombre (voire être représentées elles-mêmes sous forme de traits pour permettre certaines généralisations ne recoupant pas la distinction traditionnelle en parties du discours). Contrairement aux grammaires hors contexte traditionnelles, on n'a donc pas besoin non plus de multiplier les entrées lexicales : chaque fois qu'un mot peut avoir des propriétés différentes, il faudrait le dédoubler et créer autant de sous-catégories différentes; par exemple *élève* = N_{masc} et *élève* = N_{fem}. Avec une grammaire d'unification, on peut garder une seule catégorie, et une seule entrée lexicale: il suffit de ne pas préciser de trait <genre> dans l'entrée lexicale de *élève*; c'est en contexte que le mot héritera éventuellement ce trait, puisque l'ensemble des traits associables à un item lexical est toujours ouvert.

La représentation des discontinuités peut être partiellement réglée par un système de

traits partagés par les différentes parties d'un constituant discontinu. En revanche, le traitement des constituants mobiles est plus délicat et réclame un mécanisme supplémentaire (voir la réinterprétation de la notion d'arbre chapitre 1, section 3).

S'agissant enfin du manque de généralisations linguistiques, la critique tombe en partie à partir du moment où nombre de phénomènes traités en grammaire générative par transformations (pratiquement tout ce qui était analysé comme mouvement de syntagme nominal) sont confiés à des règles lexicales, qui relient les différentes phrases entre elles via les entrées lexicales de leurs verbes principaux (c'est l'analyse "lexicale" du passif par exemple). Pour les autres (ce que recouvre le mouvement des éléments *qu-*), les relations sont exprimées non au niveau des entrées lexicales mais au niveau de règles de réécriture supérieures.

Les relations sont donc établies entre règles de réécriture (en général les règles les plus simples), et non entre arbres complets résultants et possiblement complexes comme dans l'optique transformationnelle. Sous certaines réserves (la définition des variables notamment), la capacité générative du système reste à l'intérieur des grammaires hors contexte.

Un des points communs de ces formalismes est donc la place importante donnée au codage lexical des informations linguistiques.

b) Intérêt informatique

Les différentes grammaires d'unification sont utilisées pour des applications informatiques variées, qu'il s'agisse de systèmes d'analyse, d'interrogation de bases de données en langue naturelle ou de traduction.

La plupart s'appuient sur un composant syntagmatique représenté par une grammaire de réécriture hors contexte, et les algorithmes d'analyse associés sont des "classiques" de la littérature informatique (ce sont les mêmes que ceux utilisés pour les langages de programmation, sauf que le langage naturel impose des grammaires ambiguës). Ces algorithmes sont de conception simple et efficaces et diverses techniques existent pour assurer que le temps d'analyse au pire des cas soit une fonction polynomiale de la longueur de la phrase (CKY, Earley), voire une fonction linéaire (analyseurs LR ou LL).

Il n'en est pas de même pour ce qui est d'analyser les structures de traits associées. En lui-même, le mécanisme d'unification permet l'écriture de grammaires beaucoup plus puissantes que les grammaires hors contexte.

Les grammaires Lexicales Fonctionnelles (LFG) par exemple sont dépendantes du contexte (*context-sensitive*) et leur temps d'analyse au pire des cas est une fonction exponentielle de la longueur de la phrase.

C'est pourquoi différents mécanismes ont été mis en place pour limiter l'utilisation des structures de traits : interdiction des structures récursives, usage limité des opérations booléennes sur les structures (disjonction, négation...). A ces conditions, des formalismes tels que la grammaire syntagmatique généralisée (GPSG) ont pu rester dans le domaine

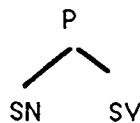
des grammaires CF.

Il n'en reste pas moins un coût de traitement non négligeable, indépendamment de la capacité générative des systèmes et de la complexité théorique au pire des cas des algorithmes d'analyse, qui tient au type même de grammaires de réécriture auxquelles ces structures sont associées.

c) Insuffisances des grammaires d'unification basées sur des règles de réécriture hors-contexte

1- un domaine de localité réduit

Le domaine de localité d'une règle est la structure à laquelle cette règle correspond, c'est-à-dire l'ensemble des mots ou des catégories qui sont accessibles en même temps directement. Par exemple, le domaine de localité d'une règle de réécriture: $P \rightarrow SN\ SV$ est le sous-arbre:



Les grammaires d'unification s'appuyant sur des règles de réécriture hors contexte, auxquelles sont associées les équations portant sur les traits, ont en général un domaine de localité réduit puisque les sous-structures manipulées sont toujours de profondeur 1. Or il n'est pas vrai que la plupart des dépendances linguistiques observées pour une langue comme le français le soient entre des éléments "frères" dans l'arbre syntagmatique. Ainsi les dépendances sujet-verbe sont des relations oncle-neveu, si l'on considère comme la plupart de ces théories qu'il y a un noeud SV entre sujet et verbe. La codépendance lexicale entre verbe principal et préposition si l'on admet un syntagme SP pour les compléments prépositionnels, comme celle entre verbe principal et syntagme interne au groupe prépositionnel complément (SN, infinitive ou *ce-que* phrase), ou celle existant entre verbe principal et déterminant du SN complément dans les phrases à verbe support (M. Gross 1981), sont quelques exemples de dépendances oncle-neveu parmi bien d'autres. Enfin, les "dépendances à distance" en cas d'extraction d'un complément verbal par exemple obligent à exprimer la relation de sous-catégorisation entre des éléments qui peuvent être infiniment éloignés (toujours en termes de profondeur).

Il ne s'agit pas là de phénomènes dont les grammaires d'unification auraient du mal à rendre compte puisqu'elles sont a priori faites pour cela. Le principe de la combinaison par unification de structures de traits associées à des règles différentes est justement de pallier l'étroitesse de ce domaine de localité. Mais le coût est certainement plus important que ne le pensaient les concepteurs de ces grammaires à leurs débuts.

Ainsi, pour vérifier des contraintes portant sur des constituants situés à différents niveaux de profondeur dans l'arbre syntagmatique, il faut "faire passer" les traits concernés entre tous les noeuds intermédiaires. Ainsi les traits de deux constituants ne pourront être comparés et mis en commun que s'ils sont copiés au niveau de leur

premier ancêtre commun. C'est-à-dire que la contrainte doit être exprimée au niveau de la règle de production de leur premier ancêtre commun. Et tous les noeuds intermédiaires devront explicitement porter le ou les traits concernés.

D'un point de vue linguistique, il va donc y avoir associés à un noeud des traits non pertinents pour ce noeud. D'un point de vue informatique, le nombre de noeuds par où il faut "faire passer" ces structures de traits est généralement élevé et peut être non borné. Les traitements sont donc coûteux en pratique.

Nous souhaitons donc pouvoir disposer d'un domaine de localité plus "naturel", c'est-à-dire plus conforme aux unités linguistiques observées. L'unification n'aurait idéalement à se faire qu'entre des structures co-présentes localement, ce qui simplifierait considérablement les traitements.

2- un traitement inadéquat des expressions figées

Cette limitation du domaine de localité, c'est-à-dire des sous-structures où peuvent s'exprimer directement des relations de dépendance entre éléments, est particulièrement gênante pour la représentation des expressions figées. Les expressions figées, étudiées de façon systématique pour le français par M. Gross 1982 et 1989b, ont en général une structure syntaxique interne et sont constituées de différents items lexicaux qui peuvent être très éloignés (en termes de profondeur syntagmatique), alors même qu'il serait souhaitable d'assigner à chacune une entrée unique dans le lexique.

Il est toujours possible de définir des traits permettant de regrouper les items concernés à un autre niveau d'analyse mais on aboutit à un système de traits très hétérogènes (des traits "lexicaux" à côté de traits syntaxiques ou lexico-syntaxiques) et l'on ne peut pas traiter des propriétés purement syntaxiques des expressions figées puisqu'elles n'ont pas d'existence à ce niveau. Les expressions figées mettent également en cause les principes de correspondance généralement retenus entre unification des traits syntaxiques et unification des traits sémantiques. Elles ont de ce fait été peu étudiées dans le cadre des grammaires d'unification (pour plus de détails voir chapitre 4).

Un de nos objectifs est d'aboutir à une représentation "naturelle" du caractère paradoxal des expressions figées à savoir leur syntaxe régulière (et possiblement complexe) et leur sémantique spécifique (car non compositionnelle).

3- un rôle important mais insuffisant du lexique

Les grammaires d'unification encodent beaucoup plus d'informations dans le lexique que ne le faisait la grammaire générative. Pourtant, on est frappé par le peu d'études lexicales menées dans le cadre de ces formalismes. La plupart en effet retravaillent les mêmes données que la grammaire générative ou des données similaires, s'appuient sur peu d'exemples et définissent des règles censées valoir sur des catégories indépendamment de leur valeur lexicale (propriétés d'extraction, coordination ...). L'on a donc l'impression que le dispositif technique est là qui donne un grand rôle au lexique (a priori tous les traits utilisés peuvent se retrouver *in fine* dans le lexique), mais sans que toutes les conséquences théoriques en soient tirées sur le plan linguistique. L'on peut ainsi se demander si ce transfert vers le lexique ne tient pas davantage de la

solution algorithmique (mise à plat des anciennes règles syntaxiques) et moins d'une réflexion proprement linguistique.

En particulier, la relation lexique-syntaxe reste plus une relation de compatibilité et de partage d'informations (par unification) qu'un véritable "principe de projection"³. Notre objectif est d'aboutir à une représentation où lexique et syntaxe formeront un seul et même composant.

Nous avons donc choisi un autre formalisme, les grammaires d'arbres adjoints (TAG), relativement ancien dans sa conception originelle (1975) mais plus récent dans sa version utilisant à son tour structures de traits et unification (1987). Il fait partie des grammaires d'unification en ce qu'il s'appuie sur des représentations linguistiques déclaratives, monotones, vise à une description directe des phénomènes de surface, et utilise (même si c'est de façon moindre) des structures de traits combinées par unification. Comme elles, il est basé sur des définitions mathématiques rigoureuses et a des propriétés informatiques intéressantes.

Mais il s'en distingue en ce que ses unités sont des arbres, non des règles de réécriture hors contexte, et qu'il les combine à l'aide d'une opération originale : l'adjonction. Il nous a semblé, sous réserve des modifications que nous y avons apportées, que ce formalisme était particulièrement bien adapté à une théorie linguistique qui donne au lexique et aux expressions figées le rôle que leur confère la théorie des lexiques-grammaires. En particulier, dans sa version "lexicalisée" que nous avons contribué à définir, ce formalisme permet une totale intégration de la syntaxe et du lexique.

Nous sommes loin d'avoir représenté tous les phénomènes syntaxiques décrits dans le cadre des lexiques-grammaires du français. Nous nous sommes principalement intéressée aux phénomènes suivants : auxiliaires, formes des complétives et des infinitives, extraction à longue distance, constructions à verbe-support, constituants discontinus et expressions figées. Certains sont traditionnellement considérés comme syntaxiques, d'autres sont à la frontière de la syntaxe et de la sémantique. Nous reprenons en effet l'idée qu'une approche syntaxique d'un certain nombre de phénomènes sémantiques (ambiguïtés, figement) est fructueuse.

Des développements vers d'autres phénomènes linguistiques sont en cours, qui poussent à enrichir le formalisme (voir chapitre 6).

Le plan de la thèse est le suivant. Dans un premier chapitre, nous exposons le cadre théorique des grammaires d'arbres adjoints et leur propriété essentielle: la "lexicalisation". Au deuxième chapitre, nous justifions les différents syntagmes que nous utilisons, et en particulier pourquoi nous n'avons pas de syntagme verbal. Au troisième chapitre, nous présentons les différentes structures élémentaires de la grammaire d'arbres adjoints du français et leur regroupement en "familles". Ensuite, nous présentons les structures utilisées pour les phrases à verbe support et les phrases figées, qui posent le problème plus général des phénomènes non compositionnels (chapitre 4),

³ Nous empruntons ce terme à la théorie chomskyenne (N. Chomsky 1981), bien qu'il n'y concerne en fait les structures syntagmatiques qu'en tant qu'elles reflètent les propriétés de sous-catégorisation des items lexicaux, et fonctionne plutôt comme un filtre sur les différents niveaux de structuration. Nous souhaitons l'entendre dans une acception beaucoup plus large (cf infra chapitre 1, section 2).

puis notre représentation des phénomènes d'extraction et des dépendances non bornées (chapitre 5). Enfin, nous présentons les recherches en cours visant à l'extension de la grammaire (chapitre 6). Au chapitre 7, nous exposons le principe de l'analyse automatique de phrases à l'aide de cette grammaire, et nous mentionnons quelques autres applications en cours de développement.

Chapitre I

Le cadre théorique

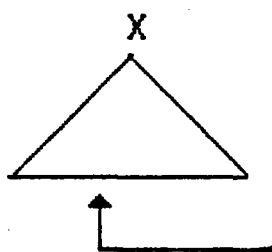
Nous allons commencer par présenter le formalisme que nous avons choisi : les grammaires d'arbres adjoints. Nous définissons les unités de base du formalisme, ainsi que les opérations qui les combinent, avant de présenter l'unification dans ce cadre particulier. Nous contrastons les propriétés formelles des grammaires d'arbres adjoints avec d'une part celles des grammaires hors contexte, d'autre part celles des grammaires contextuelles. Puis, nous présentons la contrainte de lexicalisation que nous avons ajoutée aux grammaires TAG et nous la justifions d'un point de vue linguistique. Nous montrons également en quoi la lexicalisation nécessite l'emploi de structures arborescentes et un mode de combinaison particulier de ces structures entre elles. Ce qui nous amène enfin à préciser notre usage des arbres par rapport à celui qui en est fait dans d'autres théories linguistiques.

1. Présentation des grammaires d'arbres adjoints

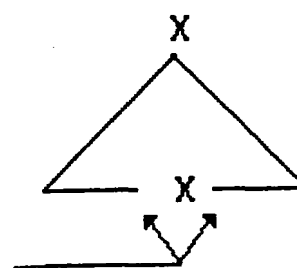
L'attrait majeur des grammaires d'arbres adjoints (TAG) vient de leurs deux caractéristiques: des unités qui sont des arbres, c'est-à-dire un domaine de localité plus étendu que celui des formalismes basés sur des règles hors contexte, et une opération originale, l'adjonction, qui permet une lexicalisation intégrale de la grammaire.

Une grammaire d'arbres adjoints (ou *Tree Adjoining Grammar* : TAG), définie par Joshi et al. 1975 et Joshi 1987, est un ensemble fini d'arbres élémentaires qui peuvent être de deux types : arbres initiaux ou arbres auxiliaires. Ces arbres dont la profondeur peut être quelconque ont à leurs feuilles des terminaux ou des noeuds à substituer. Les arbres auxiliaires ont en outre une feuille (appelée 'pied') portant un noeud non terminal et de même catégorie que leur noeud racine.

arbres initiaux α



arbres auxiliaires β



terminaux
ou noeuds à substituer

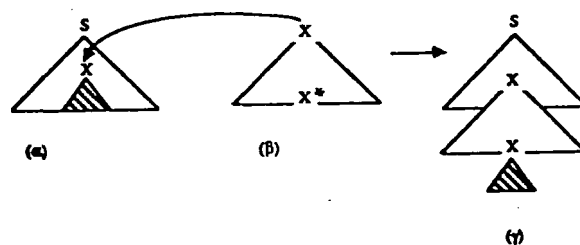
Figure 1

Le langage TAG est produit par la combinaison de tels arbres élémentaires qui donne

naissance à des arbres 'dérivés'. A chaque noeud d'un arbre élémentaire peut être associé un ensemble de traits (K. Vijay-Shanker 1987; K. Vijay-Shanker et A. Joshi 1988). L'unification exprimée sur ces traits peut contraindre les combinaisons entre arbres élémentaires (voir 1.1).

L'opération de combinaison principale est l'adjonction, qui insère un arbre auxiliaire (ou dérivé d'un arbre auxiliaire) à un noeud quelconque de même catégorie dans un arbre élémentaire ou dérivé. Nous avons ajouté une opération de substitution, qui est similaire à l'opération de concaténation utilisée en grammaire hors contexte, et donc moins puissante que l'adjonction, essentiellement pour des raisons de facilité d'expression linguistique (voir chapitre 5). Elle insère un arbre initial (ou dérivé d'un arbre initial) à la frontière d'un arbre élémentaire ou dérivé.

Adjonction:



Substitution:

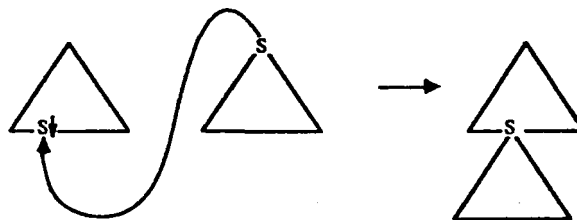


Figure 2

Un exemple d'adjonction est l'insertion d'un adverbe dans un arbre phrastique¹:

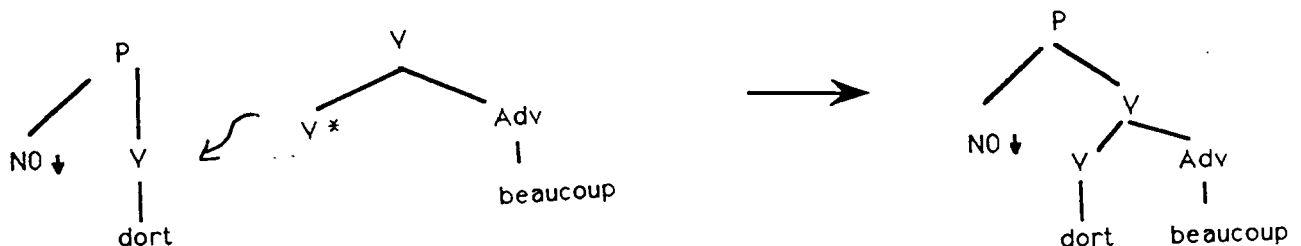


Figure 3

¹ ↓ marque les noeuds à substituer, et * les noeuds pieds.

Un exemple de substitution est l'insertion d'un groupe sujet à un noeud feuille d'un arbre phrastique:

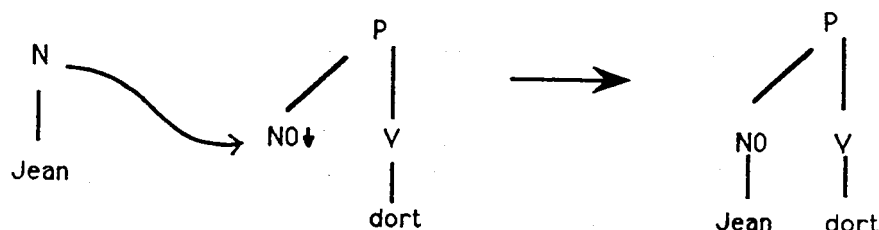


Figure 4

La substitution est toujours obligatoire, à un noeud non terminal marqué comme tel. L'adjonction est en général facultative mais peut être rendue obligatoire ou interdite par l'ajout de contraintes d'unification. Elle est toujours interdite aux noeuds à substituer et aux noeuds pieds d'un arbre auxiliaire. La substitution a lieu une fois; l'adjonction est une opération récursive et un arbre auxiliaire peut s'adjoindre sur lui-même (au noeud racine ou à tout noeud intérieur de même catégorie).

L'opération d'adjonction rend les grammaires TAG 'légèrement' plus puissantes que les grammaires hors contexte (voir infra). L'opération de substitution est similaire à celle utilisée par les grammaires hors contexte et ne change rien à la capacité générative faible de la grammaire. Elle peut être simulée par l'opération d'adjonction. Ainsi, l'adjonction du sujet *Jean* dans l'arbre phrastique se ferait ainsi:

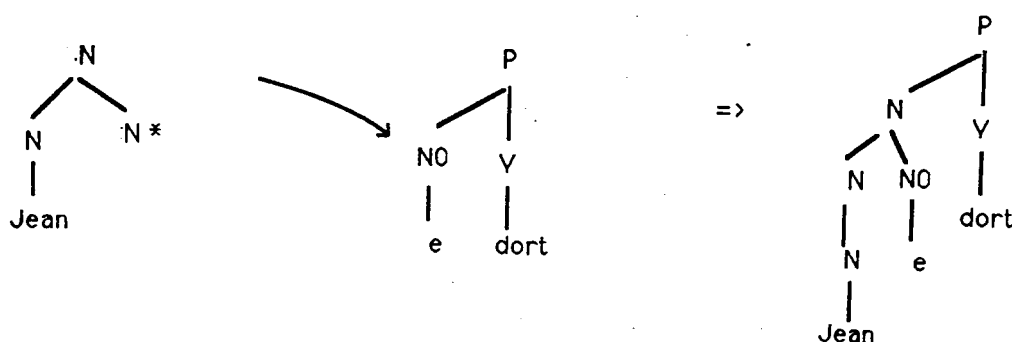


Figure 5

Ce sont des raisons linguistiques et non formelles qui nous ont poussée à introduire explicitement la substitution dans le formalisme (voir chapitres 3 et 5).

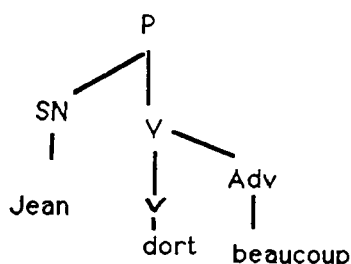
Un arbre 'dérivé' est un arbre résultant de la combinaison d'arbres élémentaires. Il sera

de type auxiliaire s'il résulte de la combinaison d'arbres auxiliaires seulement, et de type initial dans les autres cas (arbres initiaux seulement ou arbres initiaux et arbres auxiliaires). La substitution comme l'adjonction conservent donc le type de l'arbre où la substitution ou l'adjonction a lieu. Une séquence bien formée du langage TAG doit être dérivée d'un arbre initial de racine Phrase².

Dans le cas des grammaires hors contexte, l'arbre dérivé est à la fois la structure résultant de l'analyse de la phrase et l'historique de l'application des règles (la 'dérivation'). En TAG, en revanche, l'arbre dérivé est uniquement la structure résultante et ne dit rien de l'histoire de la combinaison des structures élémentaires. Celle-ci est représentée par un 'arbre de dérivation' qui est une entité spécifique.

Prenons l'exemple de la phrase *Jean dort beaucoup* :

arbre dérivé



arbre de dérivation

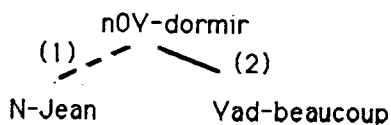


Figure 6

Un arbre de dérivation représente une combinaison d'arbres élémentaires précisant par quelles opérations (substitution ou adjonction) et à quels noeuds ceux-ci ont été combinés. Un même arbre dérivé peut correspondre à des arbres de dérivation différents, s'il peut résulter de la combinaison de différents types d'arbres élémentaires. Cette notion spécifique aux grammaires TAG est utilisée pour différencier deux types d'ambiguïtés. Aux phrases ambiguës structurellement (ou syntaxiquement) sont associés autant d'arbres dérivés (et d'arbres de dérivation) que la phrase a d'analyses possibles; aux phrases ambiguës sémantiquement seulement (avec des homonymes de même catégorie et une seule analyse syntaxique) est associé un seul arbre dérivé mais autant d'arbres de dérivation que la phrase a de lectures possibles (voir infra).

Les grammaires TAG s'inscrivent historiquement dans la lignée des grammaires en chaîne (Z. Harris 1962) avec deux modifications importantes :

- les unités de base ne sont plus des chaînes mais des arbres (A. Joshi et al. 1975).
- l'opération d'adjonction sur les arbres est plus puissante que l'opération d'adjonction

² En ne considérant pas les intitulés ou les titres (SN)

sur les chaînes (A. Joshi et al. 1972; K. Vijay-Shanker 1987)³.

Deux propriétés originales en découlent, qui font l'intérêt linguistique de ce formalisme :

- les unités manipulées ont un domaine de localité étendu, par rapport à celles des grammaires en chaîne et des grammaires hors contexte en général.
- la composante récursive de la grammaire (arbres auxiliaires) peut être disjointe de la composante qui exprime les rapports de dépendance locale (arbres initiaux), contrairement aux règles de réécriture des grammaires hors contexte qui expriment de la même façon les dépendances hiérarchiques et les phénomènes de récursion.

Si l'on ne considère que les propriétés formelles, ce formalisme appartient au groupe récemment défini (A. Joshi et al. 1989) des grammaires 'légèrement' dépendantes du contexte (*context-sensitive*) : *Head grammars* (C. Pollard 1984), *Combinatory Categorical grammars* (M. Steedman 1988), *Indexed grammars* (G. Gazdar 1988), c'est-à-dire des grammaires plus puissantes que les grammaires hors contexte mais néanmoins suffisamment contraintes pour éviter les problèmes d'indécidabilité et de complexité de traitement posés par les grammaires contextuelles en général (A. Joshi 1985).

Il s'en distingue par les deux propriétés évoquées ci-dessus et dont la conséquence importante linguistiquement est la possibilité d'une 'lexicalisation' intégrale de la grammaire (voir section 2, infra p. 19).

Nous allons maintenant expliquer le mécanisme d'unification en grammaire d'arbres adjoints, que nous comparerons à celui utilisé par les grammaires hors contexte, avant de faire un rappel plus complet des propriétés formelles des unes et des autres.

1.1. L'unification dans une grammaire d'arbres adjoints

A chaque noeud d'un arbre élémentaire est associée une structure de traits, c'est-à-dire un certain nombre de couples attribut valeur⁴. Les contraintes d'unification exprimées sur ces traits peuvent contraindre la flexion des éléments lexicaux présents dans l'arbre (elles sont alors associées à des préterminaux). Elles peuvent également contraindre la combinaison de cet arbre avec d'autres arbres élémentaires (elles sont alors associées à des noeuds feuilles à substituer ou à des noeuds intérieurs ou racines susceptibles de recevoir des adjonctions). Elles jouent donc un rôle essentiel dans la bonne formation des phrases.

3 L'utilisation de deux types d'opération: adjonction et substitution avait déjà été envisagée pour les grammaires en chaîne (voir A. Joshi 1969). L'opération de substitution est la même dans les deux cas.

4 A la limite, on peut encoder les arbres élémentaires eux-mêmes comme des structures de traits si l'on ajoute un trait fils (cf HPSG) et que l'on sépare une composante ordre des mots (voir infra p. 47).

a) L'unification dans les grammaires basées sur des règles de réécriture hors contexte

Proposée à l'origine par M. Kay 1983, elle est devenue une notion extrêmement populaire en linguistique informatique, et a permis d'enrichir considérablement le pouvoir descriptif des règles de réécriture hors contexte.

L'on définit des traits c'est-à-dire des couples attributs valeurs, les valeurs pouvant être des symboles atomiques ou des traits. De tels ensembles de traits associés à un mot ou à un syntagme sont appelés structures fonctionnelles en LFG, catégories en GPSG et HPSG. Ils peuvent se représenter comme des matrices ou comme des graphes acycliques orientés (l'arc portant l'attribut et pointant vers sa valeur)⁵.

L'unification entre deux structures de traits produit une structure résultante sauf si les deux structures portent des informations incompatibles (par exemple, si un attribut présent dans les deux structures y a une valeur différente); on dit alors que l'unification échoue. La structure résultante est la plus petite structure qui contient toute l'information contenue dans la première structure et toute l'information contenue dans la deuxième structure.

Plus précisément, on dit qu'une structure de traits A est une extension d'une structure de traits B si et seulement si :

- tous les traits à valeur atomique présents dans B sont présents dans A avec la même valeur
- pour tout trait f ayant une valeur non atomique, la valeur de f dans A est une extension de la valeur de f dans B

Soient A et B deux structures de traits. L'unification de A et de B (notée A U B) est la structure minimale qui est à la fois une extension de A et de B.

Si une telle structure n'existe pas, l'unification n'est pas définie (ou 'échoue'). Par exemple, avec deux traits f et g (dont la valeur est a ou b):

$$|f:a| \cup |g:b| = \begin{bmatrix} f:a \\ g:b \end{bmatrix}$$

$$|f:a| \cup |f:b| = 0 \text{ (échec)}$$

Figure 7

L'unification est idempotente (le résultat de l'unification d'une structure avec elle-même est cette structure), commutative et associative.

L'usage de l'unification est particulièrement utile dans le cas de structures de traits

⁵ Dans la présentation qui suit, nous adoptons un style PATR (S. Shieber 1985). Nous désignons les attributs par un nom explicite, et non par leur place dans une structure fonctionnelle à arité fixe définie a priori. Nous considérons en effet que le nombre de traits utilisés, et leur enchâssement est sans limites. Pour d'autres présentations voir notamment G. Sabah 1988 et P. Miller, T. Torris 1990.

'réentrantes', c'est-à-dire dont plusieurs attributs ont une valeur en commun. Elle permet d'exprimer une contrainte sur des attributs dont on ignore la valeur. La structure suivante, par exemple, représente un schéma de phrase intransitive avec une contrainte d'accord entre le sujet et le verbe:

$$\begin{bmatrix} \text{cat: P} \\ \text{sujet : } [\text{nombre : x}] \\ \text{prédictat : } \begin{bmatrix} \text{cat: V} \\ \text{nombre: x} \end{bmatrix} \end{bmatrix}$$

Si on la combine par unification avec un sujet singulier, elle devient :

$$\begin{bmatrix} \text{cat: P} \\ \text{sujet : } [\text{nombre : x}] \\ \text{prédictat : } \begin{bmatrix} \text{cat: V} \\ \text{nombre: x} \end{bmatrix} \end{bmatrix} \text{ U } \begin{bmatrix} \text{sujet : } [\text{cat: SN} \\ \text{nombre : sing}] \end{bmatrix} = \begin{bmatrix} \text{cat: P} \\ \text{sujet: } \begin{bmatrix} \text{cat: SN} \\ \text{nombre:sing} \end{bmatrix} \\ \text{prédictat: } \begin{bmatrix} \text{cat: V} \\ \text{nombre:sing} \end{bmatrix} \end{bmatrix}$$

Figure 8

Il peut donc y avoir partage d'informations entre certaines structures. Comme le fait remarquer S. Shieber 1986, il ne faut pas confondre des traits à valeur identique et des traits partageant la même valeur. Le résultat de l'unification de ces structures avec une troisième sera différent :

exemple de structures à traits identiques :

$$\begin{bmatrix} \text{f: [h:a]} \\ \text{g: [h:a]} \end{bmatrix} \text{ U } \text{f: [k:b]} = \begin{bmatrix} \text{f:[h:a]} \\ \text{k:b} \\ \text{g:[h:a]} \end{bmatrix}$$

exemple de structures "réentrantes" (le partage d'information étant noté par coïndicage):

$$\begin{bmatrix} \text{f: }^1\text{[h:a]} \\ \text{g: }^1 \end{bmatrix} \text{ U } \text{f:[k:b]} = \begin{bmatrix} \text{f: [h:a]} \\ \text{k:b} \\ \text{g: [h:a]} \\ \text{k:b} \end{bmatrix}$$

Figure 9

b) L'unification dans les grammaires d'arbres adjoints

Le mécanisme d'unification entre structures a été redéfini pour tenir compte de cette opération particulière qu'est l'adjonction (K. Vijay-Shanker 1987; K. Vijay-Shanker et A. Joshi 1988). A chaque noeud d'un arbre élémentaire est associée au départ une structure de traits bipartite, divisée en une partie 'top' et une partie 'bottom'. En top sont les traits indiquant les relations du noeud avec les noeuds qui le dominent, en bottom ses relations avec ceux qu'il domine. Les noeuds auxquels ne peut se produire une adjonction, c'est-à-dire les noeuds à substituer et les noeuds pieds des arbres auxiliaires, ont une structure de traits unique. A la fin d'une dérivation, parties top et bottom doivent s'unifier (selon la définition ci-dessus) à chaque noeud.

A cause de l'opération d'adjonction, l'unification a donc lieu en deux temps dans une dérivation : entre les structures de traits des arbres élémentaires combinés au fur et à mesure de leurs combinaisons, puis à chacun des noeuds de l'arbre dérivé ainsi obtenu lorsque toutes les combinaisons ont été effectuées. Lorsque deux arbres élémentaires sont combinés, leurs structures de traits sont en effet modifiées de la façon suivante :

Mise à jour des structures de traits en cas d'adjonction:

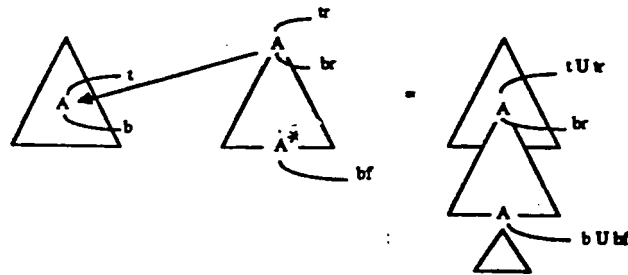
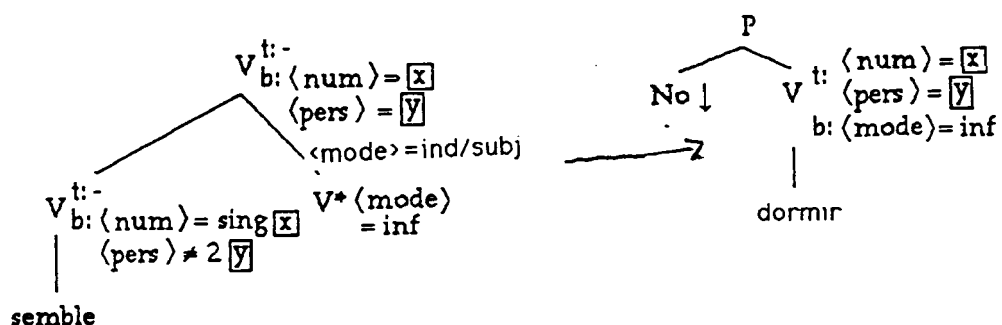


Figure 10

Prenons comme exemple la mise à jour des traits mode, nombre et personne en cas d'adjonction d'un auxiliaire :

arbres élémentaires



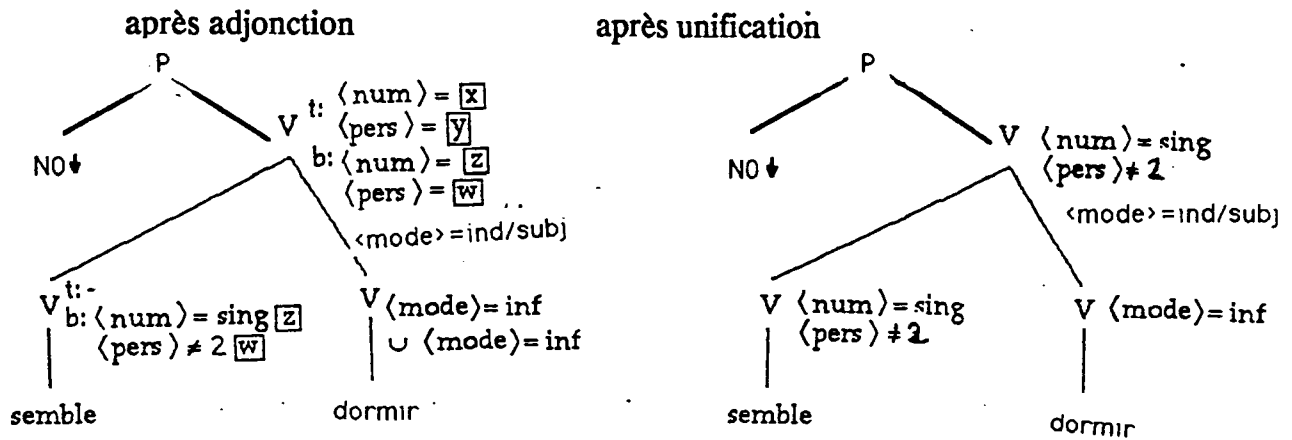


Figure 11

Mise à jour des structures de traits en cas de substitution:

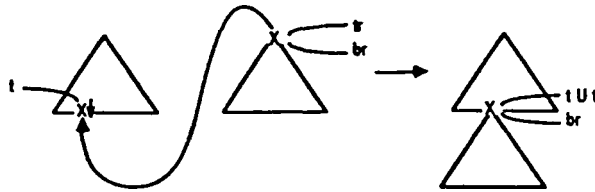
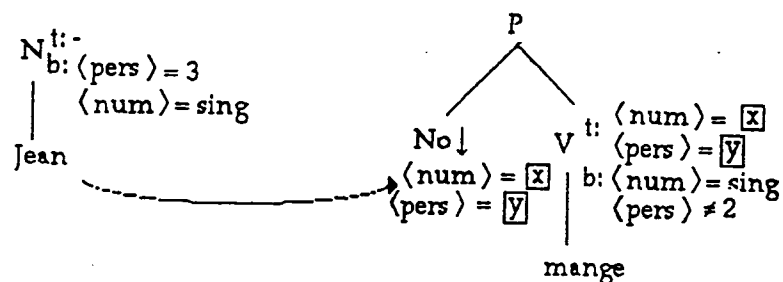


Figure 12

Prenons comme exemple la mise à jour des traits nombre et personne en cas de substitution du sujet :

arbres élémentaires



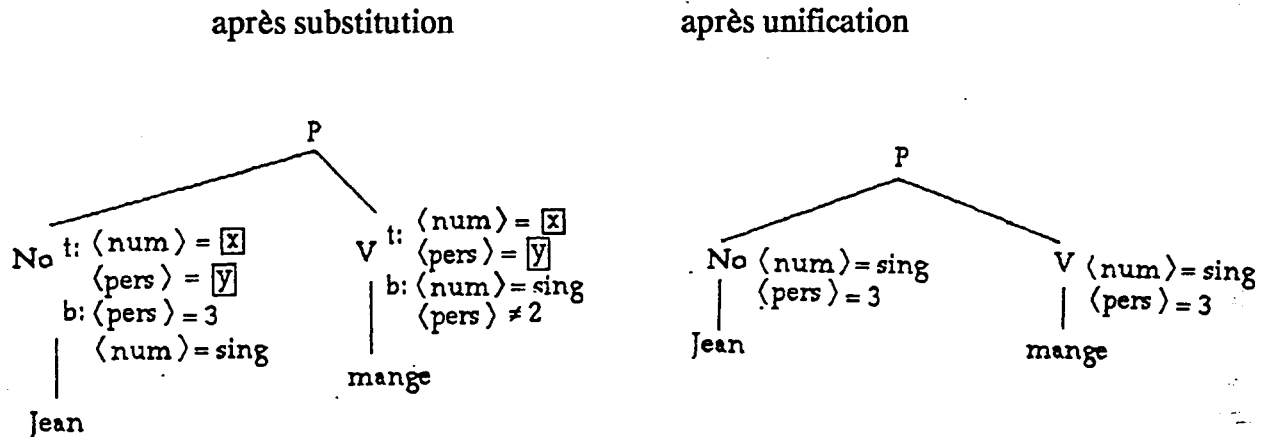


Figure 13

Pour qu'il y ait adjonction, il faut qu'il y ait unification des traits top du noeud racine de l'arbre auxiliaire avec les traits top du noeud qui reçoit l'adjonction, et des traits du noeud pied de l'arbre auxiliaire avec les traits bottom du noeud recevant l'adjonction. Pour qu'il y ait substitution, les traits top du noeud racine de l'arbre substitué doivent s'unifier avec les traits du noeud où a lieu la substitution⁶.

L'unification contraint donc les combinaisons entre arbres élémentaires et remplace les anciennes contraintes d'adjonction associables à chaque noeud (adjonction obligatoire, adjonction interdite, adjonction sélective)⁷. Les combinaisons entre arbres élémentaires (ou arbres dérivés) réglées par unification peuvent être:

- interdites: il n'y a pas d'unification possible des structures de traits des noeuds où la substitution ou l'adjonction ont lieu.
- obligatoires: les parties top et le bottom d'un noeud donné ne peuvent s'unifier: seule l'adjonction d'un arbre auxiliaire compatible permettra de résoudre ce conflit (en séparant parties top et bottom) et d'achever la dérivation.
- facultatives: les structures de traits des noeuds racine et pied d'un arbre auxiliaire peuvent s'unifier avec les parties top et bottom d'un noeud de même catégorie respectivement.

L'on voit que ces contraintes peuvent apparaître de façon dynamique à un noeud donné au fur et à mesure de la combinaison des arbres entre eux. Il convient également de noter qu'une adjonction (ou une substitution) interdite peut l'être directement (si les noeuds directement en contact en cas d'adjonction ou de substitution portent des traits

⁶ Si l'on effectue d'abord cette unification avec la partie top, et non avec la partie bottom, c'est pour obtenir exactement le même arbre dérivé que l'on opère une substitution puis une adjonction à un noeud donné, ou bien une adjonction avant de substituer l'arbre ainsi obtenu à ce même noeud.

⁷ Pour l'adjonction nulle, qui interdit toute adjonction à un noeud donné, il faudrait définir un trait tel qu'il ne puisse s'unifier avec aucune autre structure de la grammaire. Nous préférons garder la contrainte explicite (NA), ce qui revient au même.

incompatibles) ou indirectement (si cette adjonction ou cette substitution assigne à un noeud des traits top et bottom incompatibles, créant ainsi une contrainte d'adjonction obligatoire, et qu'aucun arbre auxiliaire existant dans la grammaire ne peut satisfaire cette nouvelle contrainte d'adjonction).

Si l'on interdit les structures de traits cycliques (où un attribut pourrait avoir pour valeur une structure de traits où il figurerait lui-même), les TAG avec unification ont même capacité générative que les TAG avec contraintes d'adjonction (K. Vijay-Shanker 1987).

Linguistiquement, on considère que certaines équations sur les traits sont imposées à un noeud donné par la tête lexicale de l'arbre, tandis que d'autres le sont par la structure même de l'arbre (indépendamment de sa tête lexicale).

Certains traits peuvent prendre leur valeur directement d'un élément lexical tête de l'arbre: ce sont les traits bottom, qui demeurent inchangés à ce noeud. Certains traits peuvent recevoir leur valeur par unification avec le trait d'un autre noeud: ce sont les traits top, toujours modifiés en cas de substitution ou d'adjonction.

Un même trait peut, suivant les arbres et les noeuds où il se trouve, être tantôt un trait top, tantôt un trait bottom.

Dans une grammaire TAG, l'usage de traits syntaxiques (combinés par unification) est plus limité que dans les formalismes comme GPSG ou LFG basés sur des règles hors contexte et manipulant donc des structures de base moins étendues. Ces traits, qui sont associés aux différents noeuds des arbres de la grammaire, contraignent les possibilités de combinaison de ces structures entre elles.

L'opération d'unification a été formellement redéfinie puisqu'opérant sur des arbres et non sur des règles. Tirant parti du domaine de localité étendu des grammaires TAG, nous pouvons représenter 'directement', c'est-à-dire de manière locale, des phénomènes nécessitant des équations sur des traits dans d'autres formalismes: sous-catégorisation, 'dépendances non bornées', formes figées des expressions idiomatiques. En particulier, l'unification sur des séquences de longueur indéfinie n'est pas nécessaire, étant donné le domaine de localité étendu de ces grammaires.

Dans la plupart des cas, le domaine de localité étendu des grammaires TAG va permettre d'exprimer directement des relations de dépendance entre les traits des noeuds concernés, puisqu'ici il faut et il suffit que les noeuds en question appartiennent au même arbre élémentaire, quelles que puissent être leur position et leur profondeur respectives.

Nous donnons ci-dessous, un tableau général contrastant l'usage de l'unification en TAG et dans une grammaire basée sur des règles de réécriture hors contexte de type GPSG ou LFG. Les phénomènes linguistiques mentionnés sont décrits tout au long de la thèse. Beaucoup de dépendances que les grammaires de réécriture hors contexte doivent représenter par des traits sont représentés en TAG par la topologie des arbres élémentaires eux-mêmes (par la valeur de leur(s) tête(s) lexicales ou par les catégories de leurs noeuds).

**Tableau comparatif:
grammaires hors contexte, grammaires TAG**

Phénomènes linguistiques	Usage de l'unification	
	GPSG	TAG
tête des syntagmes	oui: traits	non: directement dans l'arbre élémentaire
sous-catégorisation	oui: traits	non: structures des arbres élémentaires
complétives	oui: traits locaux	complémenteur dans l'arbre élémentaire; traits locaux pour le mode
prépositions	oui: traits locaux	non: tête dans l'arbre du prédicat
restrictions de sélection	oui: traits	oui: traits locaux
restrictions modifieur/modifié:		
catégorie	oui: traits	non: structure des arbres élémentaires
autres contraintes	oui: traits	oui: traits locaux
accord	oui: traits	oui: traits locaux
dépendances non bornées à propagation non bornée	oui: traits	oui: traits locaux
Phénomènes de contrôle:		
verbes à montée	oui: traits	non: directement dans l'arbre
verbes "equi"	oui: traits	oui: traits locaux
phrases à verbe support	-	non: structure des arbres élémentaires
expressions figées	oui: traits	non: structure des arbres élémentaires
contraintes lexicales		
sur les règles syntaxiques	-	traits sur les familles d'arbres élémentaires

Traits: locaux ou à propagation non bornée selon que l'on introduit les adverbes ou non par des règles récursives.

Traits locaux (en TAG): contraintes entre noeuds du même arbre élémentaires, ou directement en contact lors d'une substitution ou une adjonction.

1.2. Propriétés formelles

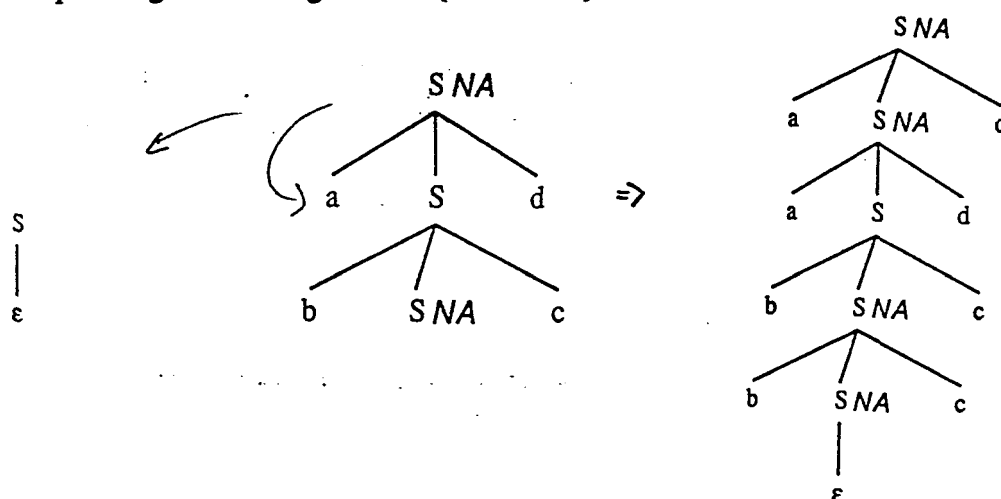
Les propriétés formelles des grammaires TAG ont été essentiellement étudiées par K. Vijay-Shanker 1987, D. Weir 1988 et Y. Schabes 1990. Nous rappelons ici les principaux résultats, dans la mesure où ils sont importants pour l'implémentation de programmes d'analyse utilisant une grammaire TAG.

a) Propriétés mathématiques

* Capacité générative faible:

Les grammaires hors contexte génèrent le langage $\{a^n b^n\}$ mais pas au delà. Le langage $\{a^n b^n c^n\}$ par exemple n'est pas un langage hors contexte (CFL). Les grammaires TAG génèrent les langages $\{a^n b^n\}$, $\{a^n b^n c^n\}$ et $\{a^n b^n c^n d^n\}$ mais pas au delà. Le langage $\{a^n b^n c^n d^n e^n\}$ par exemple n'est pas un langage TAG (TAL).

exemple de grammaire générant $\{a^n b^n c^n d^n\}$ ⁸:



Les grammaires CF génèrent le langage qui copie "en miroir" une suite quelconque de symboles (exemple les palindromes: *in girum imus nocte et consumimur igni*), et les grammaires TAG aussi.

Les grammaires CF ne génèrent pas le langage qui copie une suite quelconque de symboles alors que les grammaires TAG le génèrent.

⁸ On note 0 (ou NA) la contrainte de non adjonction à un noeud.

exemple de grammaire TAG générant le langage $\{wew/w \in \{a,b\}^*\}$

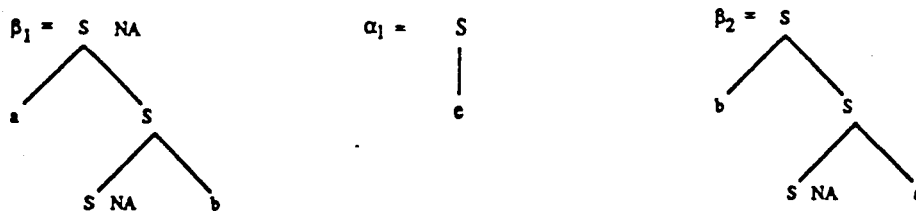


Figure 14

C'est pourquoi les grammaires TAG sont dites 'légèrement' contextuelles (*mildly context-sensitive*).

* Capacité générative forte:

Les grammaires hors contexte générant le langage $\{a^n b^n\}$ lui assignent forcément une structure emboîtée, du type de celles qu'on rencontre dans les constructions relatives: (*le fromage (que la souris (que le chat a mangée) convoitait) est intact*).

$S \rightarrow ab$

$S \rightarrow aSb$

Mais elles ne peuvent assigner au même langage une structure de dépendances croisées, de type : *le chat et la souris dorment et mangent respectivement*⁹.

$a_1 a_2 a_3 \dots b_3 b_2 b_1 \dots \in \text{CFL}$ $a_1 a_2 a_3 \dots b_1 b_2 b_3 \dots \notin \text{CFL}$

Les grammaires TAG génèrent les dépendances emboîtées mais peuvent également générer certaines dépendances croisées:

$a_1 a_2 a_3 \dots b_3 b_2 b_1 \dots \in \text{TAL}$ $a_1 a_2 a_3 \dots b_1 b_2 b_3 \dots \in \text{TAL}$

$a_1 a_2 \dots b_2 b_1 \dots c_1 c_2 \dots \in \text{TAL}$ $a_1 a_2 \dots b_1 b_2 \dots c_2 c_1 \dots \in \text{TAL}$

$a_1 a_2 \dots b_1 b_2 \dots c_1 c_2 \dots \notin \text{TAL}$

⁹ C'était d'ailleurs un des arguments de N. Chomsky 1957 pour montrer l'insuffisance des grammaires de réécriture non contextuelles.

exemple de grammaire TAG engendrant $\{a^n b^n\}$ avec des dépendances emboîtées:



Figure 15

exemple de grammaire TAG engendrant $\{a^n b^n\}$ avec des dépendances croisées:

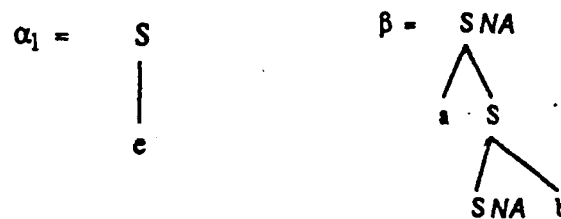


Figure 16

b) Propriétés informatiques

L'intérêt des grammaires hors contexte, qui a fait leur succès pas seulement dans le domaine du traitement du langage naturel mais aussi pour la plupart des langages de programmation, est la simplicité de l'automate qui leur correspond et l'efficacité des traitements qu'elles permettent.

* Automates

Les grammaires hors contexte correspondent à des automates à pile, les grammaires TAG à des automates à piles de piles (ou à piles emboîtées).

Un résultat récent montre que l'automate correspondant à une analyse descendante n'est pas le même que celui qui correspond à une analyse montante (Y. Schabes et K. Vijay-Shanker 1990).

* Algorithmes

Les algorithmes existant pour les grammaires hors contexte sont transposables aux grammaires 'légèrement' contextuelles, et en particulier aux TAG.

Ainsi un algorithme de type Cocke-Kasami-Younger (CKY cf D. Younger 1967) qui suppose la mise en forme normale de la grammaire (ici en arbres binaires) a été défini pour les TAG (K. Vijay-Shanker 1987).

Un algorithme de type Earley (défini pour les grammaires hors contexte par J. Earley 1970) a été défini pour les TAG par Y. Schabes, A. Joshi 1988. Comme pour les grammaires hors contexte, ce second algorithme est plus efficace en pratique (sa

complexité moyenne est inférieure à sa complexité au pire des cas pour une phrase de longueur donnée). Conformément à l'algorithme originel d'Earley, il est ascendant avec une composante prédictive et procède de gauche à droite. Au lieu de règles à points (*dotted rules*), il utilise des arbres à points qui séparent à chaque noeud l'arbre en une partie gauche (reconnue) et une partie droite (non reconnue). C'est l'algorithme que nous utilisons pour notre grammaire (cf chapitre 7).

Un analyseur déterministe ascendant de type LR (pour un sous-ensemble des TAG) a également été défini par Y. Schabes et K. Vijay-Shanker 1990.

* Complexité d'analyse

Pour les grammaires CF, le temps d'analyse au pire des cas (avec un algorithme de type Earley ou CKY) est proportionnel au produit du cube de la longueur de la séquence (en nombre de mots) et du carré de la taille de la grammaire (en nombre de règles).

Pour les TAGs, le temps de traitement au pire des cas avec les mêmes algorithmes est proportionnel au produit de la puissance 6 de la longueur de la séquence (en nombre de mots, notée n) et du carré de la taille de la grammaire (en nombre d'arbres élémentaires, notée G).

La place occupée en mémoire au pire des cas est proportionnelle au produit du carré de la longueur de la séquence par la taille de la grammaire pour les grammaires hors contexte, et au produit de la taille de la grammaire par la puissance 4 de la longueur de la séquence dans le cas des grammaires TAG.

Avec un algorithme de type Earley, la complexité en temps au pire des cas, pour des grammaires hors contexte non ambiguës, est proportionnelle à n^2 . Pour des grammaires TAG non ambiguës, elle est proportionnelle à $G^2 n^4$.

Il est intéressant de noter que la complexité de traitement est dans le meilleur des cas une fonction linéaire de la longueur de la phrase à analyser, pour les TAG comme pour les grammaires hors contexte (avec un algorithme de type Earley, cf Y. Schabes 1990).

**Tableau comparatif:
grammaires hors contexte, grammaires TAG, grammaires contextuelles**

	CFG	TAG	CSG
langages	$\{a^n b^n\}$ * ww	$\{a^n b^n c^n d^n\}$ ww	$\{a^n b^n c^n d^n e^n f^n \dots\}$ www...
dépendances	emboîtées	emboîtées et croisées	emboîtées et croisées
automate	à piles	à piles de piles	à mémoire linéairement bornée
complexité avec un algorithme Earley (au pire des cas)			
temps	$G^2 n^3$	$G^2 n^6$	exponentielle
espace	$G n^2$	$G n^4$	exponentielle

Les grammaires TAG sont donc un peu plus coûteuses pour l'analyse automatique que les grammaires hors contexte: l'automate est plus complexe, et le temps d'analyse d'une phrase au pire des cas est proportionnel à la puissance 6 (et non au cube) de sa longueur. Il faut cependant considérer deux points :

- Les grammaires basées sur des règles de réécriture hors contexte font un plus grand usage des contraintes d'unification que les grammaires TAG. Elles doivent en particulier vérifier la compatibilité entre les valeurs de traits qui peuvent être à des distances non bornées, ce qui peut rendre leur traitement plus long en pratique .
- Le traitement des langues naturelles (par opposition aux langages de programmation) concerne en pratique des phrases assez courtes et des grammaires de grande taille. Le coût supplémentaire des grammaires TAG concernant la seule longueur de la phrase (n^6 et non n^3), et non la taille de la grammaire (toujours G^2), la différence en temps de traitement ne devrait pas être très sensible en pratique¹⁰.

¹⁰ Nous n'avons pas pu effectuer de comparaisons empiriques, faute d'une grammaire hors contexte ayant exactement la même couverture que la nôtre.

2. Pour une "lexicalisation" intégrale de la grammaire

Une contrainte supplémentaire a donc été ajoutée au formalisme originel: la lexicalisation. Tout arbre élémentaire doit avoir à sa frontière au moins un terminal servant de tête lexicale. La catégorie de cet item lexical dans un lexique TAG est cet arbre élémentaire. Le lexique TAG est donc une grammaire.

La justification d'une telle "lexicalisation" est d'abord linguistique, mais elle a aussi un intérêt informatique. Nous montrerons enfin en quoi il s'agit d'une propriété originale de notre formalisme arborescent: en effet les grammaires de réécriture hors contexte (avec ou sans unification) ne sont pas "lexicalisées" (et en général ne pourraient pas l'être sauf à être traduites sous forme de TAG).

2.1 Justification linguistique

Nous ne considérons pas les structures de constituants comme un composant séparé du lexique, contrairement à la plupart des grammaires¹¹. L'intégration du lexique et de la syntaxe se justifie de plusieurs points de vue, et a été abondamment discutée dans les publications du LADL. Nous allons simplement opérer quelques rappels.

a) Les propriétés traditionnellement associées aux mots sont généralement dépendantes du contexte où ces mots apparaissent.

Par exemple, quand on dit que *taureau* a pour pluriel *taureaux*, il s'agit de relier des phrases telles que:

J'ai vu un taureau
J'ai vu des taureaux

Cette relation n'est pas toujours observée, puisque à la phrase:

Jean a pris le taureau par les cornes

ne correspond pas de la même façon la phrase:

Jean a pris les taureaux par les cornes

Comme conclut M. Gross 1989, la relation entre singulier et pluriel n'est donc pas une relation entre mots isolés.

Il en est de même d'autres relations morphologiques. Il a été souvent noté (A. Meunier 1977, J. Giry-Schneider 1978...) que la relation de nominalisation devait s'exprimer entre phrases simples et non entre un nom et un verbe. Ainsi, le nom *attaque* peut être relié au verbe *attaquer* dans le contexte suivant:

¹¹ GPSG définit par exemple des règles de réécriture non lexicales (notamment pour la Phrase) et LFG a un niveau syntaxique indépendant du lexique (appelé structure-c). Des traits appropriés (et les principes de propagation gouvernant ces traits) sont ensuite chargés de relier les deux composants.

Jean a lancé une attaque contre le fort
 = *Jean a attaqué le fort*

Mais non dans cet autre contexte:

Max a eu une attaque
 ≠ *Max a attaqué X*
 ≠ *X a attaqué Max*

Il en est de même des relations définies par la sémantique lexicale, telles que synonymie, antonymie, hyponymie ou hyperonymie. Ainsi, *voler* est synonyme de *dérober* dans le contexte:

Jean a volé 3 F à Marie
 = *Jean a dérobé 3F à Marie*

Mais non dans le contexte:

L'oiseau a volé vers son nid
 = * *L'oiseau a dérobé vers son nid*

Campagne peut s'opposer à *ville* dans le contexte:

Jean aime la campagne
 ≠ *Jean aime la ville*

Mais non dans ceux-ci:

Jean fait une campagne publicitaire.
Jean aime les campagnes électorales.

Il semble donc que peu puisse être dit sur des mots isolés de leur contexte syntaxique et lexical. Ceux-ci ne sont donc pas les unités pertinentes en morphologie, en syntaxe ou en sémantique.

b) Réciproquement, les propriétés traditionnellement associées aux syntagmes, de même que les relations entre phrases, sont généralement dépendantes du type d'items lexicaux en jeu.

Cela est manifeste pour les conditions de bonne formation sur les séquences de catégories ou pour les règles de réécriture.

De telles règles, sans information lexicale, tout au moins pour le français, d'une part contraignent très peu la bonne formation des phrases; d'autre part engendrent beaucoup d'ambiguïtés erronées. En effet, si l'on espère pouvoir exclure à ce niveau les séquences

de catégories non dérivables à partir d'un noeud Phrase, très peu d'entre elles vont être exclues. Une séquence *Det Det Det* va l'être, mais même les séquences suivantes, a priori suspectes, devraient être acceptées en raison des phrases ci-dessous¹² :

- 1: ? P = * = > Prep Prep Det N V
- 2: ? P = * = > Det Prep N V
- 3: ? P = * = > Det V Det

- 1: *Près de trois personnes écoutent*
- 2: *Une d'elles rit*
- 3: *Plusieurs trouvent quatre*

D'autre part, ces règles syntagmatiques non lexicalisées vont engendrer beaucoup d'ambiguïtés erronées. Considérons les dérivations partielles suivantes, qui sont toutes possibles¹³ :

- 1: P = * = > SN V SN
- 2: P = * = > SN V SN Adj
- 3: SN = * = > Det N Adj
- 4: SN = * = > Det N

Non seulement des phrases agrammaticales comme:

* *Jean (dort + meurt + ressemble) son chien malade*

vont être acceptées au même titre que:

Jean (trouve + caresse + promène + tue) son chien malade,

Mais elles vont toutes être analysées comme systématiquement ambiguës (dérivation 1 ou dérivation 2), puisqu'à ce niveau aucune information n'est disponible sur la sous-catégorisation des verbes. Ajouter, comme le font des formalismes comme GPSG ou HPSG un trait < sous-cat > aux règles syntagmatiques générales réécrivant V et SV est une solution de type purement technique, qui peut se révéler coûteuse et peu justifiée. Ici, le nombre de verbes dont la sous-catégorisation correspond aux dérivations 1 et 2 (*rendre, trouver, voir...*) est bien inférieur à celui des verbes dont la sous-catégorisation correspond à 1 seulement (*tuer, caresser, promener...*). A la limite, il faudra une combinaison de traits pour chaque verbe, si l'on envisage de surcroît la commutation de la séquence *SN SA* avec une complétive ou une infinitive: *rendre* n'autorise ni l'une ni l'autre, *trouver* seulement la complétive, *voir* autorise les deux:

12 Sauf si l'on introduit des catégories vides, permettant de régulariser les séquences acceptables, par exemple SN = Det N avec Det = *Plusieurs* et N = e. Il reste que les catégories vides elles-mêmes ne peuvent être introduites indépendamment de la valeur lexicale des catégories en jeu. La séquence *Det e* ne sera pas acceptable si Det = *le* ou *ce*.

13 Conformément à l'usage, nous notons = * = > une dérivation obtenue par l'application d'un nombre quelconque de règles de réécriture (notées ->). Ainsi, nous ignorons ici l'existence éventuelles de constituants intermédiaires.

- * *Je rends (que le chien est malade + (Luc + E) battre du tambour).*
- Je trouve (que le chien est malade + *(E + Luc) battre du tambour).*
- Je vois (que le chien est malade + (?E + Luc) battre du tambour).*

Il en serait de même pour des règles générales permettant de dériver un SN à partir d'une séquence *Prep Prep Det N* ou *Det*, nécessitant ensuite une machinerie de traits complexe pour spécifier quels déterminants peuvent se comporter en SN, ou prendre des modificateurs prépositionnels.

Si des structures syntagmatiques (ou des combinaisons de structures) sont ainsi limitées à un nombre extrêmement réduit d'items lexicaux, les introduire indépendamment du lexique dans la grammaire pour les contraindre ensuite par des traits (qui ne feront que vérifier la présence de ces éléments lexicaux) est un procédé coûteux et qui ne fait gagner aucune généralisation.

Un résultat concordant pour l'anglais est fourni par K. Church 1988. Celui-ci constate qu'il est impossible de désambiguer les mots d'un texte (selon leur catégorie) sur la base de séquences a priori de catégories. S'il l'on prend la phrase simple : *I see a bird*, chaque mot est catégoriellement ambigu (Pronom/Nom, Nom/ Verbe, ou Nom/Article). Ce qui fait 16 suites de catégories possibles pour cette phrase. Et toutes ces séquences seront "bien formées" en anglais¹⁴ : [N N N N], [N V N N], [N N N V] ...

La bonne formation ne peut donc se définir sans tenir compte des éléments lexicaux en jeu.

Il en est de même si l'on considère les relations syntaxiques entre phrases (transformations). Il n'est pas vrai par exemple, comme l'a montré M. Gross 1969, qu'à toute construction transitive corresponde une phrase passive, ou une pronominalisation par *le*:

Jean aime Marie
Maire est aimée par Jean
Jean l'aime

Ce livre vaut les 50 F que j'ai payés.
 * *Les 50 F que j'ai payés sont valus par ce livre*
Ce livre les vaut

Ce livre coûte 50 F
 * *50 F sont coûtés par ce livre*
 * *Ce livre les coûte.*

14 K. Church, qui se situe dans une optique d'étiquetage catégoriel probabiliste, doit donc avoir recours à des statistiques lexicales (quelle est la catégorie la plus probable de tel mot ?) et non syntagmatiques (quelle est la séquence de catégories la plus probable ?) afin d'analyser correctement *I see a bird*.

Les travaux du LADL sur les phrases simples du français ont bien montré l'importance des facteurs lexicaux en jeu, aboutissant à de petites classes lexicales conditionnant les relations telles que passif, pronominalisation, neutralité ou symétrie.

Les travaux plus récents sur les phrases à verbe support, et ceux en cours sur les phrases figées ne font que confirmer ce résultat, qu'ils étendent notamment aux questions et aux relatives¹⁵ :

Jean caresse cet espoir
Cet espoir est caressé par Jean
Quel espoir Jean caresse-t-il ?

Jean a brisé la glace
La glace a été brisée par Jean
Quelle glace Jean a-t-il brisée ?

Le projet a fait des vagues
** Des vagues ont été faites par le projet*
** Quelles vagues le projet a-t-il faites ?*

Peu semble donc pouvoir être dit des structures syntaxiques considérées indépendamment des éléments lexicaux en jeu (en particulier du verbe).

2.2. De nouvelles unités lexico-syntaxiques

Les unités linguistiques pertinentes doivent donc intégrer éléments lexicaux et structures syntaxiques. Nous considérons de telles catégories étendues et structurées comme les unités de base de la grammaire.

Nous n'avons pas par exemple de règles indépendantes réécrivant la Phrase, le syntagme verbal ou le groupe nominal puisque à toute règle est associé au moins un élément lexical. Un verbe comme *caresser* va introduire la règle: P -> SN V SN, et un verbe comme *trouver* les règles: P -> SN V SN, P -> SN V SN SA et: P -> SN V P. Ces règles doivent être en fait des structures de profondeur au moins égale à 2, c'est-à-dire des arbres, pour introduire en même temps le constituant Phrase, la catégorie morphologique V et la valeur lexicale du verbe¹⁶ :

¹⁵ On note # l'impossibilité de la lecture idiomatique de la phrase.

¹⁶ La profondeur serait de 3 s'il s'agissait de donner l'équivalent des grammaires habituelles qui placent un noeud SV intermédiaire entre V et P. Nous avons abandonné ce noeud non motivé linguistiquement (voir chapitre 2).

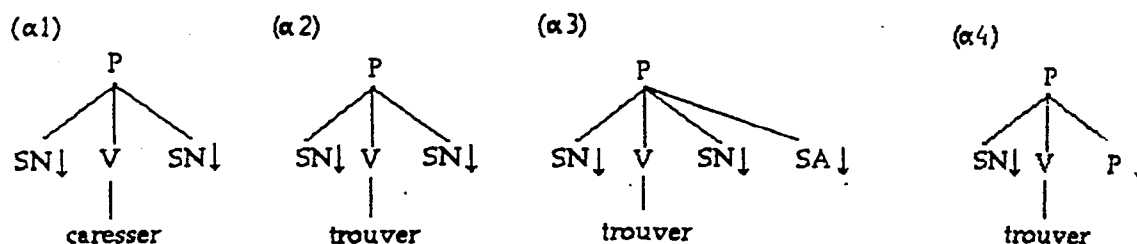


Figure 17

Cette lexicalisation des structures de base de la grammaire conduit donc à regrouper en un arbre plusieurs sous-structures correspondant à des règles différentes dans un formalisme non contextuel. Par exemple, aux arbres α_1 et α_3 ci-dessus correspondraient les règles:

α_1 :
 $P \rightarrow SN \ V \ SN$
 $V \rightarrow caresser$

α_3 :
 $P \rightarrow SN \ V \ SN \ SA$
 $V \rightarrow trouver$

De ce fait, les traits chargés de gérer les relations entre ces sous-structures dans les autres formalismes, tels que <sous-cat>, <tête> et <barre> en GPSG, deviennent caducs. Cette "lexicalisation" de la grammaire permet donc d'éliminer les traits qui dans les autres formalismes ne font que gérer les risques de discordance entre information syntaxique et information lexicale, et de construire une grammaire véritablement lexico-syntaxique.

Il en est de même pour les noms, les adjectifs et les adverbes dans la mesure où ils peuvent être également "têtes" de phrases. En considérant qu'il y a sous-catégorisation des sujets comme des compléments, c'est en effet la valeur lexicale de l'adjectif qui détermine la forme du sujet dans les phrases copulatives, celle du nom prédicatif qui la détermine dans les phrases à verbe support et celle de l'adverbe dans les phrases à verbe support d'adverbes (voir chapitre 3). Enfin, dans les expressions figées, c'est à des niveaux de profondeur variés qu'il faut pouvoir associer les différents éléments figés de l'expression (voir chapitre 4):

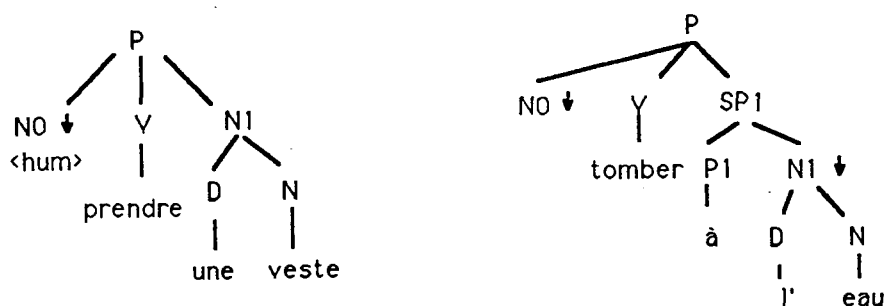


Figure 18

Les règles de réécriture classiques sont équivalentes à des arbres de profondeur 1. Pour "lexicaliser" la grammaire, il faut donc définir comme unités de base des arbres de profondeur quelconque (finie mais sans limite supérieure), qui sont équivalents à plusieurs règles de réécriture prises ensemble. Tant qu'on ne fait que cela, on reste dans un cadre non contextuel, puisque ces unités arborescentes pourraient être combinées par substitution. Nous allons maintenant montrer en quoi cette opération de substitution est insuffisante.

2.3. La lexicalisation comme caractéristique originale des grammaires TAG

La "lexicalisation" complète de la grammaire, outre la notion d'arbre élémentaire, nécessite également un mode original de combinaison des structures lexicales entre elles, qui correspond à notre opération d'adjonction (sauf à perdre la distinction entre arguments et ajouts selon laquelle les premiers font partie de la structure lexicale du prédicat mais pas les seconds).

Définissons de façon plus précise ce que nous entendons par "lexicalisation". Une grammaire G est lexicalisée si elle consiste en:

- un ensemble fini de règles ou structures élémentaires associées à chaque item lexical (qui en est la tête).
- une ou plusieurs opérations pour combiner ces règles ou structures entre elles.

Si l'on prend l'ensemble des grammaires hors contexte il n'est pas vrai qu'à chacune correspond une grammaire lexicalisée totalement équivalente (c'est-à-dire engendrant le même langage et construisant les mêmes structures) si l'on s'en tient à la seule opération de combinaison des grammaires hors contexte à savoir la substitution. La démonstration générale a été faite par Y. Schabes, A. Abeillé et A. Joshi 1988¹⁷.

Par exemple, la grammaire $G1$ suivante (qui est ambiguë) n'est pas lexicalisée suivant notre définition:

$S \rightarrow SS$
 $S \rightarrow a$

Et elle ne peut l'être si la seule opération de combinaison définie est la substitution. Considérons par exemple la grammaire $G2$ comportant les structures suivantes:

¹⁷ Si l'on se contente d'une grammaire lexicalisée faiblement équivalente à la grammaire de départ (engendrant le même langage mais non les mêmes structures), il y a de nombreux procédés permettant de rester dans le cadre non contextuel : mise en forme normale de Greibach, traduction en grammaire catégorielle applicative...

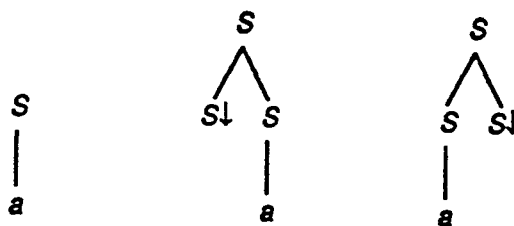


Figure 19

La grammaire G2 est lexicalisée car toutes ses structures ont bien chacune une tête lexicale et elle engendre le même langage que G1. Mais elle n'engendre pas toutes les structures engendrées par G1. Ainsi G2 engendre l'arbre (1) mais pas l'arbre (2), alors que tous deux sont engendrés par G1.

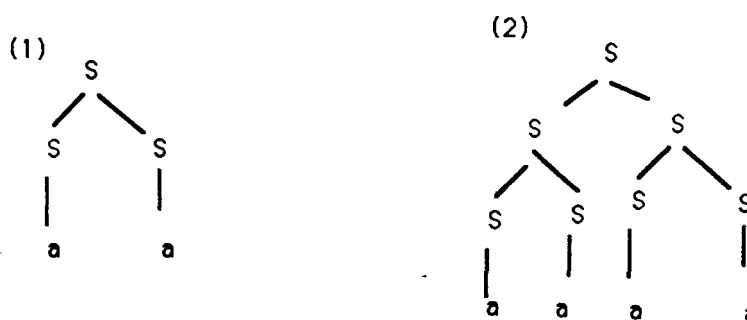


Figure 20

Il en serait de même si l'on prenait une grammaire lexicalisée avec des arbres plus profonds au départ, car leur profondeur sera toujours finie. Il y aura toujours un niveau où la différence apparaîtra puisque dans les arbres engendrés par G1 tous les terminaux a peuvent être à une profondeur infinie alors que dans une grammaire lexicalisée, au moins un a (celui du premier arbre utilisé) est par définition à une profondeur finie.

En revanche, si l'on introduit l'opération d'ajonction, il devient possible de construire une grammaire lexicalisée exactement équivalente à G1. Il suffit de deux arbres de base, dont l'un est de type auxiliaire¹⁸ :

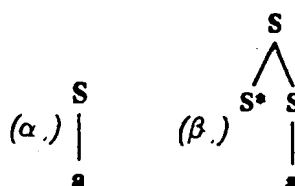


Figure 21

Par adjonctions successives de l'arbre (β) sur l'arbre (α) ou sur lui-même, l'on engendre

¹⁸ ↓ marque les noeuds à substituer, et * les noeuds pieds.

bien les mêmes structures que G1, par exemple:

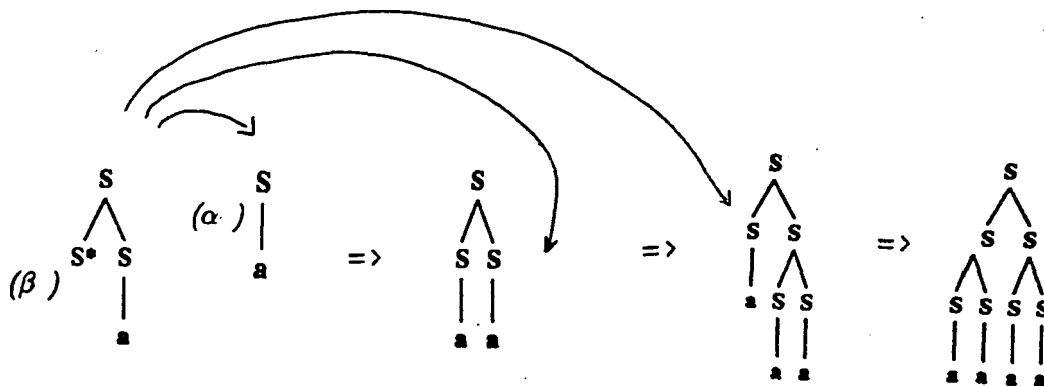


Figure 22

La conclusion générale est donc que c'est l'opération d'adjonction (spécifique aux grammaires TAG et responsable de leur caractère contextuel) qui permet la lexicalisation des grammaires¹⁹.

Si l'on s'en tient à des grammaires plus proches du langage naturel et nettement moins récursives, la conclusion est un peu différente: certaines grammaires CFs peuvent être lexicalisées sans être transformées en TAG, mais obligent alors à définir une tête lexicale qui n'est pas forcément motivée linguistiquement (Y. Schabes, A. Abeillé, A. Joshi 1988).

Considérons la grammaire G3:

P -> SN V
 V -> dort
 SN -> Jean
 V -> V Adv
 Adv -> beaucoup

Elle engendre des phrases du type: *Jean dort beaucoup, Jean dort beaucoup beaucoup ...* Si l'on prend les structures lexicalisées suivantes, la nouvelle grammaire résultante n'engendrera pas le langage attendu. La phrase *Jean dort beaucoup* ne sera pas engendrée:

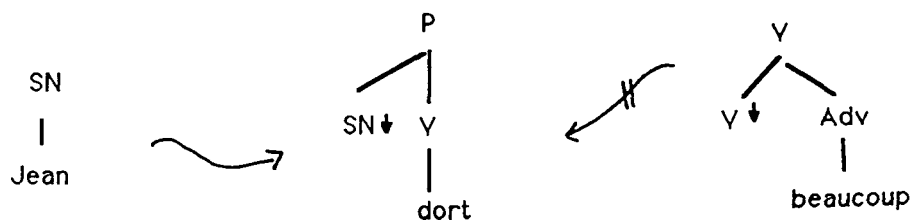


Figure 23

¹⁹ Plus exactement la lexicalisation de toutes les grammaires CF sauf de celles comportant des chaînes récursives du type: $S \rightarrow S$ ou $S \rightarrow A$ et $A \rightarrow S$, c'est-à-dire des grammaires infiniment ambiguës.

Si l'on avait pris une structure de base plus complexe pour le noeud phrase, avec déjà un adverbe, ç'aurait été la phrase à deux adverbes qui n'aurait pas été engendrée, et ainsi de suite.

La seule solution, avec la substitution pour seule opération, est de choisir les structures suivantes, qui sont peu motivées linguistiquement puisque l'adverbe doit être obligatoirement tête de la phrase:

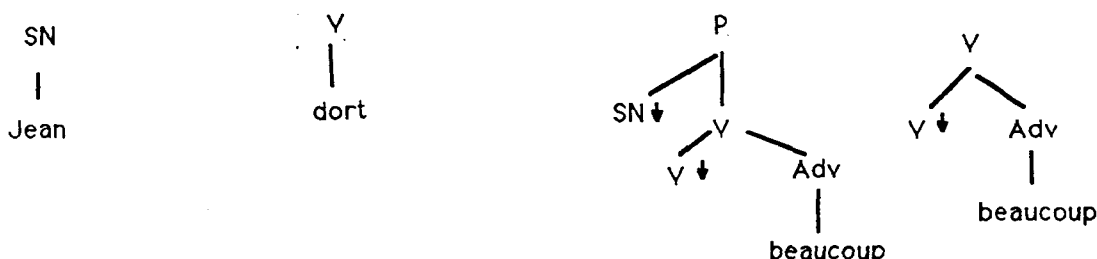


Figure 24

Le langage généré est bien celui de G3 et les structures assignées sont les mêmes. Mais l'adverbe n'est pas la tête linguistique désirée de la structure de la phrase simple.

Notons que ce n'est pas en soi la présence d'une règle récursive dans la grammaire de départ qui fait ici problème, c'est le fait que la récursion porte sur un noeud situé entre l'élément qu'on souhaite définir comme tête et la racine de la structure de base. Une règle de production des complétives de type : $P \rightarrow SN V P$ ne poserait pas problème, pas plus d'ailleurs qu'une règle à deux récursions du type: $P \rightarrow P V P$ (pour les verbes à complétives sujet et objet).

En revanche, avec l'opération d'adjonction, le problème disparaît. L'on peut lexicaliser les grammaires tout en choisissant aux structures de base des "têtes" motivées linguistiquement. En effet, avec des règles récursives d'introduction de modifieurs telles que : $V \rightarrow V Adv$ (ou $N \rightarrow N Adj$), comme en comportent la plupart des grammaires hors contexte, la tête lexicale (ici le V ou le N le plus bas) risque d'être éloignée à une profondeur potentiellement infinie du noeud racine de la structure (ici P ou SN). Il faut donc représenter ces ajouts par des arbres auxiliaires, qui en s'adjoignant créent (par définition) un niveau de profondeur supplémentaire. Ainsi la grammaire G3 ci-dessus peut être lexicalisée de la façon suivante:

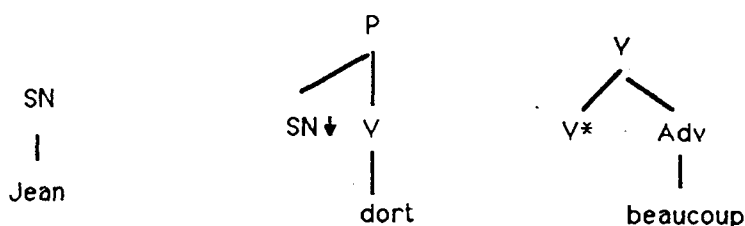


Figure 25

Une objection peut se présenter à l'esprit: il n'est pas sûr que le mode récursif d'introduction des modifieurs envisagé ci-dessus soit justifié linguistiquement. S'il l'on décidait de d'exclure de nos grammaires des règles du type $X \rightarrow X Y$ pour l'introduction des modifieurs, l'adjonction ne serait plus l'opération adéquate pour les lexicaliser.

Deux possibilités se présentent. Si l'on s'en tient à la seule opération de substitution, il faut représenter les modifieurs directement dans la structure de l'élément à modifier, à l'aide de l'opérateur de Kleene qui rend compte à la fois de leur optionalité et de leur itérabilité potentiellement infinie. Il s'agirait donc toujours d'une grammaire d'arbres (pour les raisons mentionnées plus haut) mais la seule opération de combinaison admise serait la substitution non l'adjonction, la grammaire restant donc hors contexte (l'opérateur de Kleene, noté $*$, est défini pour les langages réguliers et ne change pas la capacité générative des grammaires hors contextes). Tous les modifieurs seraient donc au même niveau:

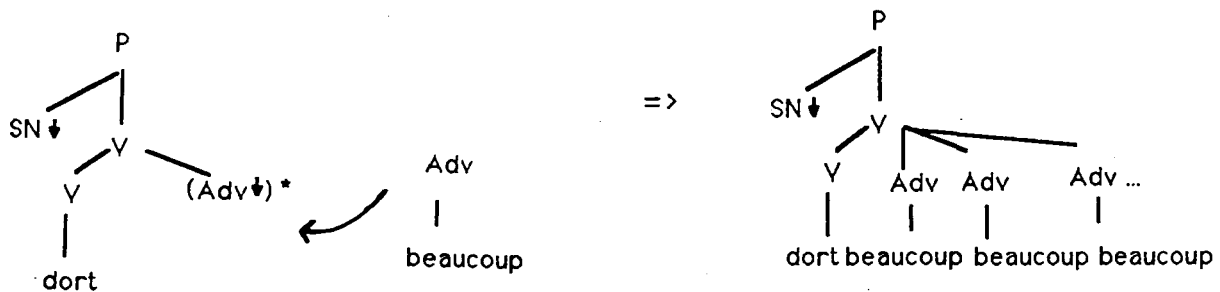


Figure 26

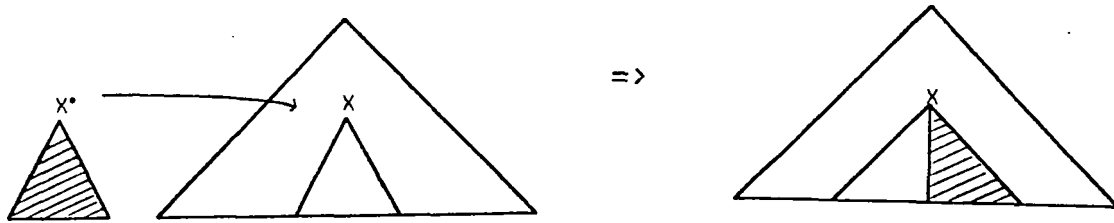
En soi, la structure résultante où arguments et modifieurs sont mis au même niveau n'est pas forcément contestable. Ce qui est contestable, c'est que leur mode d'introduction soit le même c'est-à-dire qu'ils soient les uns comme les autres parties prenantes de la représentation lexicale du prédicat. Rien ne va formellement distinguer les modifieurs des arguments optionnels (et itérables par coordination). Il est toujours possible de définir les structures prédicats-arguments à un autre niveau (par exemple sémantique) mais elles n'auront plus de lien avec les phénomènes de sous-catégorisation que devraient représenter les structures syntaxiques de base.

L'inconvénient majeur de cette représentation est donc qu'il n'y a plus de correspondance naturelle entre structure syntaxique et structure prédicat-argument. L'on perd la notion de "phrase simple" au sens de Z. Harris 1962 et M. Gross 1975, ou, pour le dire autrement, le domaine de prédication ne correspond plus à un domaine de localité syntaxique.

Une autre possibilité consiste à introduire les modifieurs dans la structure du prédicat mais sans pour autant créer un niveau d'enchâssement supplémentaire. C'est ce que fait par exemple l'opération de furcation (notée $\circ$) proposée récemment dans le cadre des

Segment grammars (De Smedt et Kempen 1990), par ailleurs assez proches des TAG²⁰. Les propriétés de cette opération ne sont pas encore totalement définies, mais on peut la présenter ainsi²¹ :

furcation à droite:



furcation à gauche:

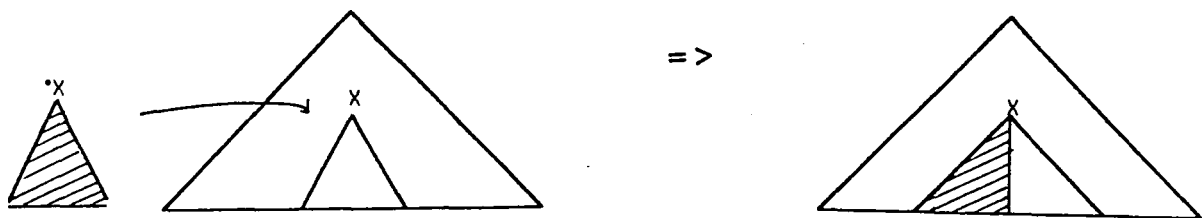


Figure 27

Les structures permettant de dériver *Jean dort (beaucoup)** deviennent donc:

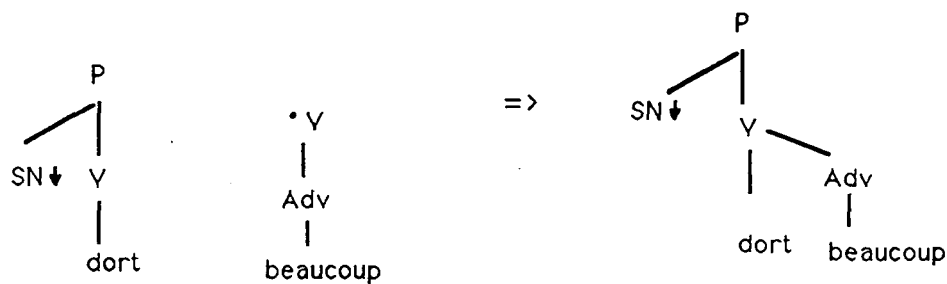


Figure 28

20 Cette opération est équivalente (sur des arbres) à l'opération d'"adjonction" utilisée par les grammaires en chaîne (sur des chaînes).

21 La furcation est mal définie si l'on veut ajouter du matériel de part et d'autre d'un noeud, par exemple *ne...pas* de part et d'autre du verbe conjugué. Ce qui se fait en revanche naturellement avec l'adjonction (Cf chapitre 3).

Il nous semble cependant que cette mise à plat forcée n'est pas forcément très heureuse. Un argument traditionnellement avancé pour l'enchâssement des adjectifs et des adverbes concerne les phénomènes de portée. Il s'agit par exemple de faire la différence entre *Jean est venu trois fois un samedi* (où l'interprétation préférentielle est qu'il s'agit de trois samedis différents) et *Jean est venu un samedi trois fois* (où l'interprétation préférentielle est qu'il s'agit du même samedi). Ces phénomènes sont en fait limités à un petit nombre d'adjectifs et d'adverbes, et la différence dans l'ordre linéaire rend suffisamment compte des différences.

Nous souhaitons avancer deux autres types d'arguments: la modification elle-même par son caractère non monotone conduit souvent à distinguer deux catégories différentes: la catégorie modifiée et la catégorie résultante. D'autre part, il est intéressant de considérer comme des ajouts des structures qui peuvent difficilement être mises au même niveau que ce qu'elles modifient à savoir les phrases à complétives et les déterminants complexes.

Si l'on considère les auxiliaires (auxquels nous associons les modaux et verbes opérateurs au sens de M. Gross 1975), il est clair qu'ils sont porteurs de traits différents de ceux des verbes qu'ils modifient: les verbes modifiés sont au participe passé ou à l'infinitif, les auxiliaires à tous les modes. Leur accord peut se faire également différemment: *avoir* s'accorde avec le sujet, le participe qui le suit est soit invariable soit s'accorde avec l'objet. Il n'est pas possible de rendre compte de ces différences de propriétés si le verbe et le groupe qu'il forme avec son ou ses auxiliaires partagent le même noeud, c'est-à-dire si les auxiliaires sont introduits par furcation:

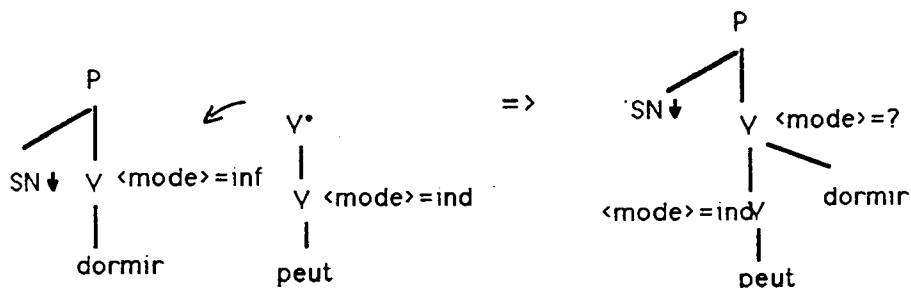
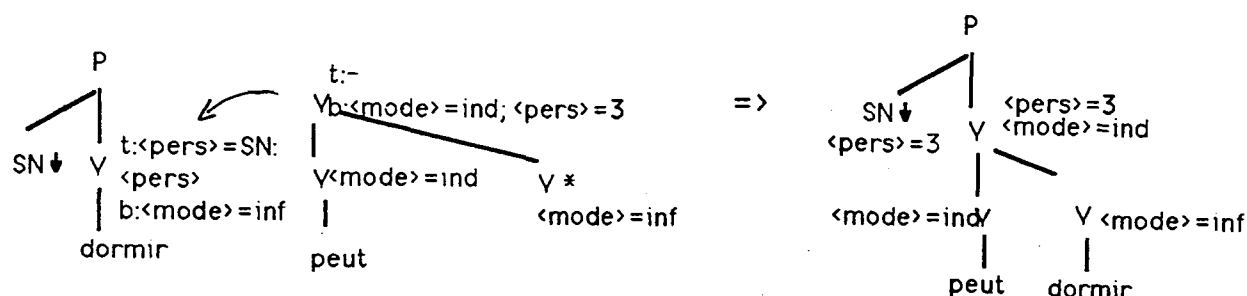


Figure 29

Notons qu'il ne s'agit pas non plus de considérer un mode de combinaison des structures de traits concernées non monotone, les propriétés du verbe étant simplement remplacées par celles de l'auxiliaire. En effet, on a besoin à la fois des propriétés de l'auxiliaire (qui s'accorde avec le sujet et conditionne le mode de la phrase) et de celles du verbe (qui peut s'accorder avec l'objet). Ainsi, le placement de *pas* qui s'effectue après un verbe fini n'est bien décrit que si l'on conserve l'information que le verbe est au participe passé (tandis que l'auxiliaire est à l'indicatif) dans: *Jean a mangé*. Sinon, on aurait aussi bien: * *Jean n'a mangé pas* que: * *Jean n'a pas mangé*. De même, les propriétés du groupe auxiliaire-verbe ne sont pas forcément les mêmes que celles de l'auxiliaire. Ainsi, le trait <passif> est un trait du groupe et non de l'auxiliaire *être* ou

du participe pris isolément (Cf J-C. Milner 1986). C'est pourquoi l'opération d'adjonction paraît bien adaptée à l'introduction des auxiliaires puisqu'elle permet la distinction de trois noeuds :



Il en est de même des déterminants, à partir du moment où l'on considère qu'ils ne sont pas des arguments déjà prévus dans la structure du nom (c'est-à-dire en dehors des combinaisons figées Det-N; voir chapitre 3). Il est clair que les propriétés du nom, du déterminant et du groupe Det-N peuvent être à chaque fois différentes. C'est d'ailleurs le principe de la distinction classique de deux niveaux N et SN, ayant chacun leurs modificateurs spécifiques, ce qui explique par exemple qu'un adjectif (modifieur de N) puisse précéder un nom commun mais non un déterminant ou un nom propre²² :

un grand fils
 * *grand un fils*
 ? * *grand Jean*

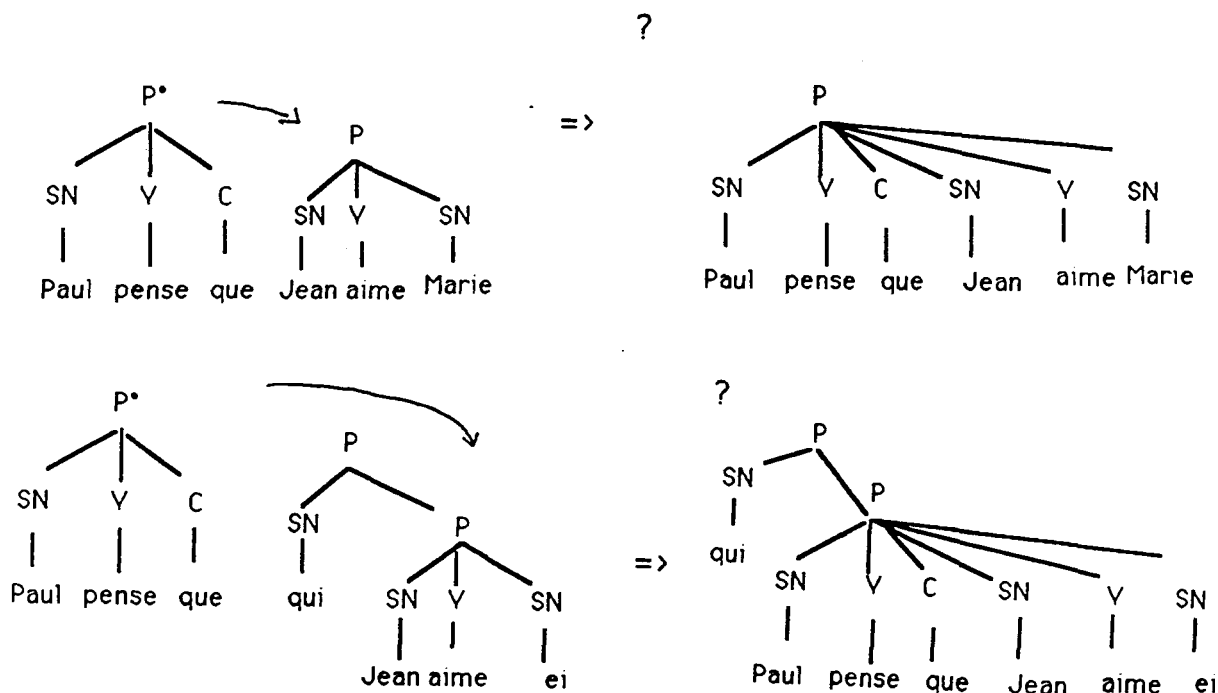
Si l'on ajoute le cas des déterminants complexes et prédéterminants (comme: *la plupart des N, un (grand) nombre de N ..*), un autre inconvénient de leur mise à plat avec le nom apparaît. Il devient alors difficile de faire apparaître une structure de groupe prépositionnel ou de SN complexe pourtant suggérée par la séquence des catégories et l'insertion régulière de modificateurs internes: *la (grande) majorité de ces gens*. L'adjonction permet bien l'empilement successif de ce type d'ajouts, tout en construisant à chaque fois des groupes nominaux et prépositionnels réguliers: *la majorité d'un grand nombre de gens ...* (voir chapitre 3).

Un autre type de construction pour lequel l'adjonction est nécessaire est représenté par les verbes à arguments phrastiques. Il est en effet intéressant de considérer les "principales" comme des ajouts à leur complétive et comme susceptibles de pouvoir s'insérer à l'intérieur de celles-ci. Prenons l'exemple des dépendances non bornées. L'élément antéposé peut être un argument, libre ou figé, infiniment séparé de son prédicat qui apparaît en fin de phrase. Si l'on souhaite conserver l'hypothèse majeure qui sous-tend la lexicalisation, à savoir que les relations prédicat-argument sont définies par (et définissent) les structures de base de la grammaire, cet élément antéposé doit faire partie du même arbre élémentaire que son prédicat éloigné. Selon ce traitement

²² Pour une discussion plus détaillée de cette distinction voir chapitre 2. Ce n'est pas la différence de nom (de catégorie) qui importe ici, c'est la distinction de deux noeuds porteurs de traits distincts.

"local" des dépendances non bornées, il faudrait donc (avec la furcation) introduire les "principales" de la façon suivante:

Paul pense que Jean aime Marie



Qui Paul pense que Jean aime ?

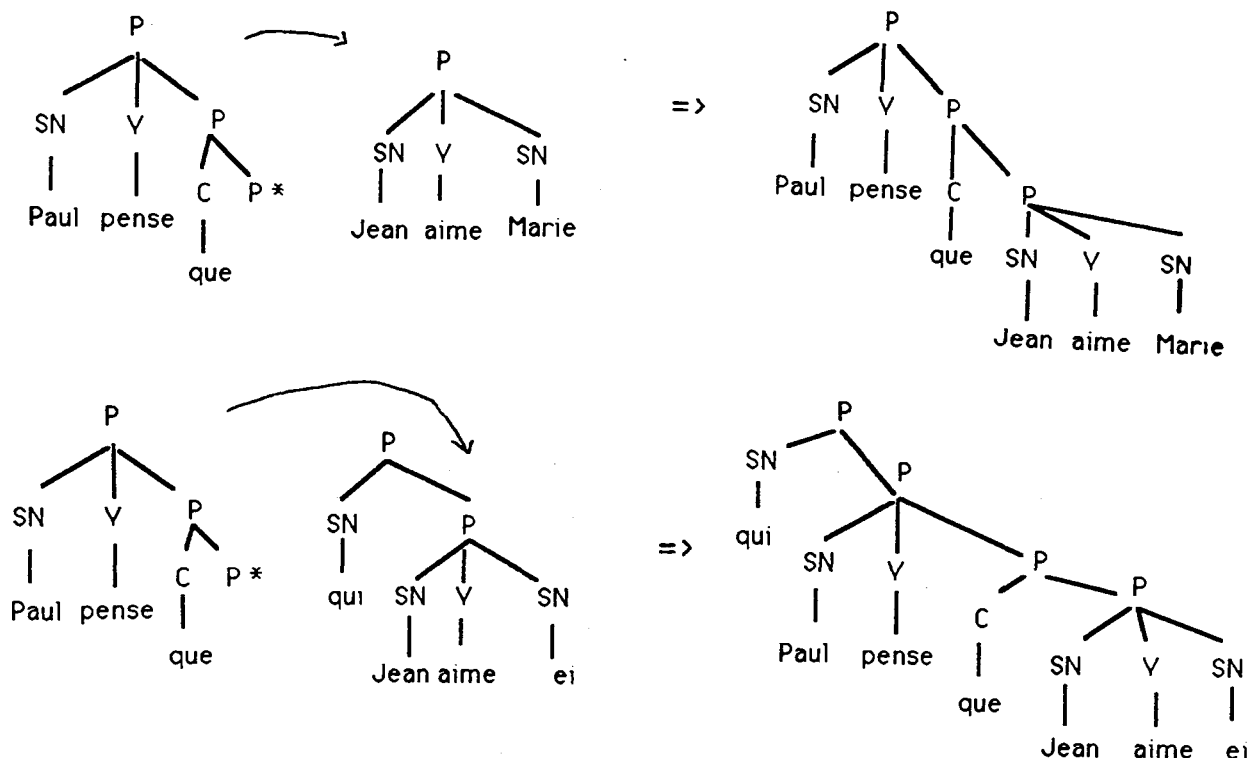
Figure 30

Mais c'est alors la relation prédicat-argument entre "principales" et complétives qui se perd.

En revanche, si l'on considère la principale comme adjointe entre l'élément antéposé et son prédicat, c'est-à-dire à l'intérieur de l'arbre élémentaire correspondant, le problème disparaît²³. Ainsi pour nous, la "principale" est un cas mixte: modifieur en tant qu'elle peut commuter avec un groupe adverbial et peut se mettre en incise; prédicat en tant que la complétive commute souvent avec un groupe nominal abstrait. Elle est donc naturellement représentée comme un arbre auxiliaire (voir chapitre 5):

²³ Une solution alternative (mais à l'intérieur d'un cadre strictement hors contexte) serait de considérer l'élément antéposé comme une structure séparée, à coïncider avec la structure prédictive dont il dépend (comportant éventuellement une trace). Les dépendances non bornées seraient donc constituées par empilements successifs de principales et en dernier de cet élément *qu*-. C'est globalement le traitement donné en GPSG, nous expliquons au chapitre 4 pourquoi il est malheureux, particulièrement lorsque l'élément antéposé est un morceau d'expression figée.

Paul pense que Jean aime Marie



Qui Paul pense que Jean aime ?

Figure 31

Nous n'avons pas trouvé d'arguments similaires pour l'emboîtement des adjectifs ou des adverbes. Il conviendrait peut-être de les traiter différemment. Il faudrait ainsi distinguer selon les modifieurs, et l'opération de furcation pourrait aussi être utilisée de façon alternative pour les adjectifs et les adverbes, l'opération d'adjonction concernant plutôt les déterminants, les auxiliaires et les phrases à complétive.

Nous avons préféré nous en tenir à notre formalisme originel et ne pas ajouter la furcation comme opération supplémentaire, d'autant plus que ses propriétés formelles sont encore mal étudiées (particulièrement la prise en compte de l'ordre respectif modifieur/modifié et la définition de l'unification).

C'est pourquoi nous avons choisi la lexicalisation à l'aide de l'adjonction. Le prix à payer est que tous les modifieurs deviennent des ajouts structuraux et sont forcément à des niveaux différents (les syntagmes intermédiaires ainsi créés ne se justifiant pas toujours directement).

Cette "lexicalisation" des structures de base de la grammaire est donc rendue possible par les deux propriétés des grammaires TAG : domaine de localité étendu et autonomie de la composante récursive, comme l'ont montré Y. Schabes, A. Abeillé, A. Joshi 1988.

L'opération d'adjonction permettrait également de 'lexicaliser' aisément la plupart des grammaires hors contexte (sauf celles comportant des chaînes récursives du type: $S \rightarrow S$) tout en permettant une certaine liberté dans le choix des têtes lexicales.

Cette lexicalisation permet de simplifier considérablement le système des traits associés aux catégories dans le lexique et dans les règles syntaxiques. Elle permet aussi de définir une nouvelle stratégie de traitement automatique (voir chapitre 7).

3. Sur la notion d' arbre

Il doit être clair à présent que notre usage de représentations arborescentes a peu à voir avec celui qui en a été fait à partir de N. Chomsky 1957 en grammaire générative. Il faut également noter que nous faisons nôtres la plupart des critiques qui ont été adressées à ce mode de représentation, et que c'est cette démarche qui nous a amenée, aussi paradoxal que cela puisse paraître, à choisir le formalisme TAG. Nous allons donc rappeler brièvement ces discussions afin de préciser notre propre position.

En grammaire générative transformationnelle, les arbres de la structure profonde correspondent à l'application (potentiellement récursive) de règles de réécriture considérées comme largement indépendantes du lexique. Ils définissent l'ensemble des fonctions grammaticales. L'ensemble des phrases de la langue (les arbres de surface) est obtenu par application de transformations sur ces structures profondes. Ces transformations permettent donc de représenter des relations entre arbres tenues pour au moins aussi importantes que les arbres eux-mêmes.

Cet usage des arbres a été abondamment critiqué et des représentations alternatives ont été proposées en linguistique théorique comme en linguistique informatique.

3.1. Critiques théoriques

La critique la plus radicale a sans doute été portée par M. Gross 1975 dans le cadre d'une critique plus générale des méthodes et objectifs de la grammaire générative. Pour M. Gross, il est fait des arbres des usages hétérogènes, voire contradictoires, qui reflètent la surabondance des modes de représentation propres à la grammaire générative et qui sont le symptôme du caractère pauvre, voire redondant, de ce type de représentation. Ils ont pour tâche de représenter l'organisation de la phrase en syntagmes, syntagmes pour lesquels existent déjà des définitions distributionnelles et transformationnelles (qui ne se recoupent d'ailleurs pas), et différentes propriétés de ces syntagmes (propriétés là encore aussi bien transformationnelles que combinatoires et qui ne sont que superficiellement étudiées). Or ces propriétés peuvent être représentées, de façon assez arbitraire,

aussi bien par le recours à différents symboles auxiliaires que par des points d'attachement différents dans un arbre, et ceci dans l'arbre profond ou dans l'arbre de surface. Un ordre différent d'application de certaines transformations peut également être envisagé. Or la solution "arborescente" (différents points d'attachement dans un arbre) est sans doute la moins bonne, car la plus rigide. Elle suppose par exemple qu'à une position (par exemple le SN à droite du verbe directement dominé par SV) correspond une et une seule fonction syntaxique (ici objet direct) et que celle-ci recouvre un ensemble constant et homogène de propriétés. Or il n'en est rien (Cf M. Gross 1969) et cette définition géométrique des fonctions grammaticales a été abondamment critiquée (Cf le développement des grammaires relationnelles et de LFG qui considèrent celles-ci comme des primitives non déductibles de configurations arborescentes).

J-C. Milner 1984, faisant état du moindre rôle joué par les représentations arborescentes dans la théorie chomskyenne plus récente, s'est également livré à une critique assez générale. Il souligne l'hétérogénéité des informations censément apportées par les arbres. Ainsi les arbres expriment le fait que les structures linguistiques sont hiérarchisées en un certain nombre de niveaux (et l'on peut discuter sur la pertinence de ces distinctions, Cf notre chapitre 2) et identifient les relations entre différents niveaux alors qu'il est douteux que la relation de *Jean* à N soit la même que celle de N à GN, et encore la même que celle de GN au groupe Phrase. De plus, les représentations en arbres avec leurs enchâssements possiblement infinis créent des problèmes d'ambiguïtés pour l'application des règles (Cf le principe A/A, voir chapitre 5). Les domaines syntaxiquement pertinents ne recoupent en fait pas ces arbres, et la théorie a besoin de notions toutes différentes (c-commande, gouvernement, projection) pour neutraliser les différences de position non pertinentes et établir des différences entre des positions que les arbres confondent. J-C. Milner suggère en note une représentation plus combinatoire, basée sur des séquences de catégories sous-catégorisées lexicalement puis combinées entre elles selon des listes préétablies, tout en s'efforçant par ailleurs de garder une conception géométrique de la syntaxe (J-C. Milner 1989).

Pour résumer les différentes critiques apportées aux arbres tels qu'ils sont utilisés en grammaire générative, nous avons retenu les points suivants:

3.1.1 L'ordre des mots.

Les constituants mobiles ou discontinus s'attachent malaisément à l'intérieur d'une représentation arborescente (telle qu'en génèrent des règles de réécriture hors contexte). C'est pourquoi un formalisme comme la *Grammaire syntagmatique généralisée* (GPSG, G. Gazdar et al. 1985) sépare les relations de dominance et celles de précédence que les arbres regroupent. Les arbres n'expriment donc plus alors que de pures relations de dominance, les relations de précédence étant laissées à un autre composant (format ID/LP qui sépare les règles de dominance:

(ID) des règles d'ordre linéaire (LP)).

La même distinction a été proposée par A. Joshi 1987. Comme pour le format ID/LP de GPSG, deux problèmes se posent: formellement, le nouveau système est beaucoup plus puissant que l'ancien, on sort des grammaires 'légèrement' contextuelles aux propriétés informatiques bien définies. Linguistiquement, il apporte une trop grande flexibilité pour une langue à ordre contraint comme le français. Nous ne l'avons donc pas adopté et avons gardé l'encodage des relations de précedence dans nos arbres élémentaires. Il faudra néanmoins trouver un équivalent plus contraint pour les adverbes mobiles (voir chapitre 6)

Nous développons au chapitre 4 le cas de certains constituants généralement considérés comme discontinus (ex: *N0 prendre N1 en compte*) pour lesquels nos représentations arborescentes semblent bien adaptées.

Il reste les parenthèses et incises, pour lesquelles J. McCawley 1982 proposait des arbres exotiques du type:

I think, my friend, that they are wrong

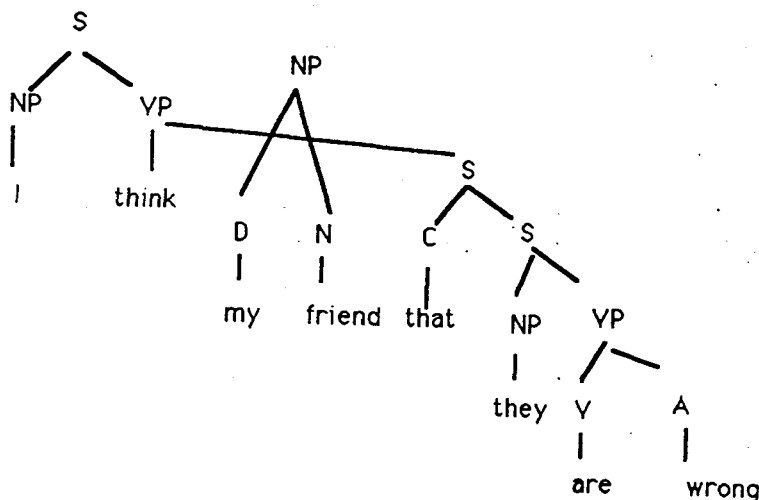


Figure 32

Nous n'avons traité que de certaines incises, que nous relions à des phrases à complétives (voir chapitre 6).

3.1.2 Double analyse et réanalyse

Les arbres sont également gênants si l'on considère que certaines phrases peuvent avoir des découpages concurrents (sans être pour autant ambiguës), et certains éléments occuper la même fonction grammaticale à l'intérieur d'une même phrase.

Un exemple est celui des phrases à verbe support où l'on constate une triple

extraction en *c'est que* (M. Gross 1976):

Luc a commis une agression contre Max

C'est contre Max que Luc a commis une agression.

C'est une agression contre Max que Luc a commise

C'est une agression que Luc a commise contre Max

Si la dernière phrase peut éventuellement s'expliquer par extraposition du complément prépositionnel *contre Max*, les deux phrases précédentes posent un problème d'attachement apparemment insoluble: *contre Max* est-il un complément de *commettre* (déplaçable de façon autonome) ou un complément de *agression* (avec lequel il se déplace) ?

Le problème peut être résolu si l'on considère la propriété "complément de" non seulement en terme de point d'attachement mais aussi comme signifiant que l'élément appartient à la même structure de base que ce dont il est complément. Ainsi, dans notre représentation, il devient possible de dire que *contre Max* est à la fois complément de *commettre* et de *agression*, sans pour autant supposer deux arbres ou une réanalyse (voir chapitre 4).

3.1.3. Contraintes lexicales sur les règles syntaxiques

Les structures d'arbres, en tant qu'elles correspondent à un composant syntagmatique séparé du lexique sont extrêmement pauvres. C'est par exemple le cas avec les grammaires lexicales fonctionnelles (J. Bresnan 1982) qui représentent la plupart des informations importantes à un deuxième niveau: la structure fonctionnelle qui est elle directement reliée au lexique (et se présente comme une structure de traits)²⁴.

Nous avons vu que les mêmes motivations nous ont poussée à une lexicalisation radicale de la grammaire, laquelle nous a conduite à définir des unités lexicales structurées, c'est-à-dire des arbres "élémentaires". Nous sommes tout à fait d'accord pour considérer que les arbres syntagmatiques classiques sont extrêmement pauvres, voire trompeurs.

Notre "réhabilitation" des arbres, fondée sur une lexicalisation intégrale de la grammaire, s'inscrit donc dans la droite ligne de la revalorisation du lexique et du mouvement d'appauvrissement du composant syntagmatique des 10 dernières années.

²⁴ L'on peut considérer que nous rassemblons dans nos unités TAG les deux structures (structure de constituants et structure fonctionnelle) distinguées par les grammaires LFG. Toutefois, nous différons en ce que nous n'utilisons pas de fonctions grammaticales explicites (qui sont les attributs de leur structure fonctionnelle). Etant donné l'indépendance de la plupart des propriétés syntaxiques, et le fait qu'elles peuvent être partagées par plusieurs éléments dans une même phrase (ex double passif), nous préférons abandonner toute référence explicite aux fonctions grammaticales et encoder directement au niveau des unités lexicales concernées (nos arbres élémentaires) les règles syntaxiques concernées.

3.2. Les arbres des grammaires TAG

Il y a en fait trois usages fort distincts du mot "arbre" dans la définition classique des TAGs : les 'arbres élémentaires' sont les unités de base de la grammaire, les 'arbres dérivés' résultent de la combinaison de ces arbres élémentaires entre eux et les 'arbres de dérivation' donnent l'histoire de ces combinaisons.

Nous allons préciser ici ces notions de base. Les arbres élémentaires définis pour la grammaire du français seront exposés aux chapitres 3 et 4.

3.2.1 Les arbres élémentaires

Répetons-le, c'est une notion à première vue paradoxale puisque l'on a coutume d'entendre par 'arbre' le résultat d'une analyse et non son point de départ.

Elle est moins surprenante si l'on songe à toute la tradition qui a plaidé pour des données lexicales à structure interne, et particulièrement aux catégories complexes associées aux items lexicaux en HPSG ou en grammaire catégorielle.

Nous n'avons donc pas de règles syntagmatiques autonomes (règles de réécriture) et nous considérons la grammaire davantage comme combinaison finie d'éléments lexicaux que comme processus récursif (sur notre limitation de la récursion, voir chapitre 7).

Nos arbres élémentaires sont les catégories des éléments lexicaux. Notre grammaire se présente donc comme un grand dictionnaire (voir annexe 3).

Ces arbres élémentaires sont des unités syntaxiques et sémantiques (cadres de sous-catégorisation complets, relations filler-gap et prédicat-arguments). Certaines ambiguïtés peuvent être levées à ce niveau. Il y aura par exemple, un arbre élémentaire différent pour chaque emploi d'un verbe comme *voler* ou *affecter* (Cf chapitre 3):

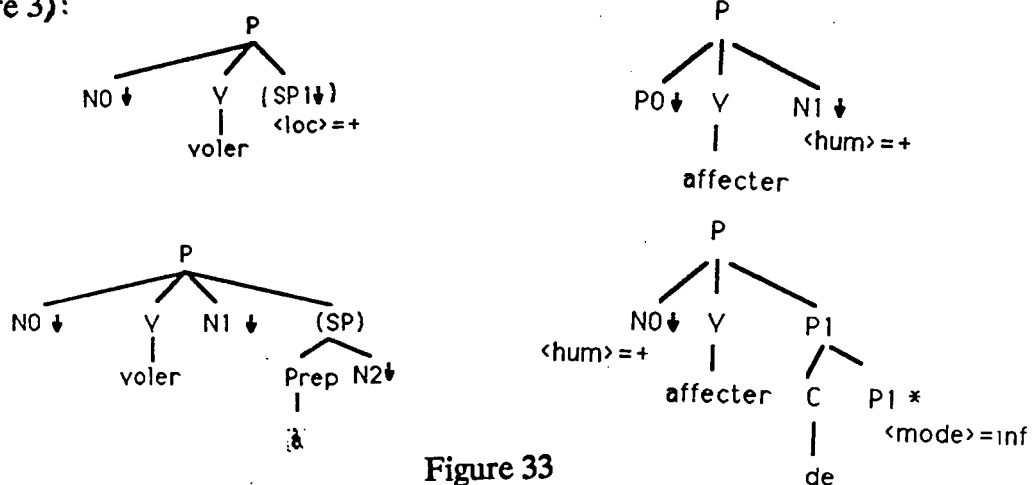


Figure 33

C'est également à ce niveau que s'appliquent les règles syntaxiques. La relation actif-passif est définie comme une relation entre arbres élémentaires, entre 1 et 2 par exemple (voir chapitre 3):

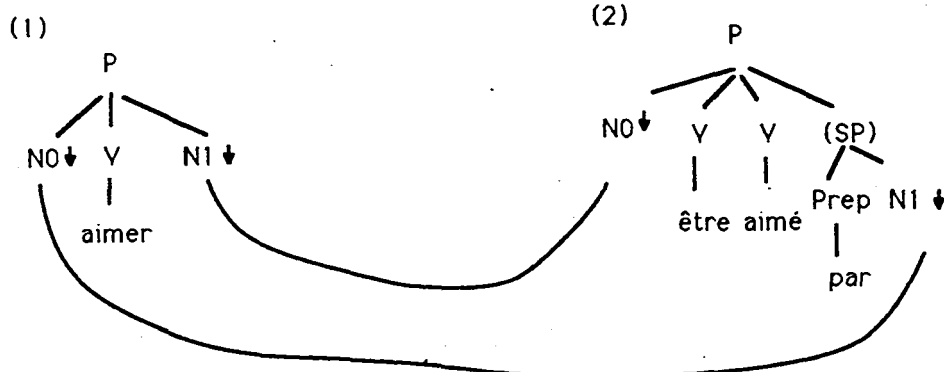


Figure 34

Il est donc aisé d'exprimer les contraintes lexicales éventuelles contraignant tel ou tel schéma syntaxique (ex: verbes sans passif).

Les règles morphologiques opèrent aussi à ce niveau. Ainsi, la relation *attaquer* et *attaque* est une relation entre arbres élémentaires (ce qui permet de distinguer *attaque* dans: *Y a lancé une attaque contre X* et dans *Y a eu une attaque*).

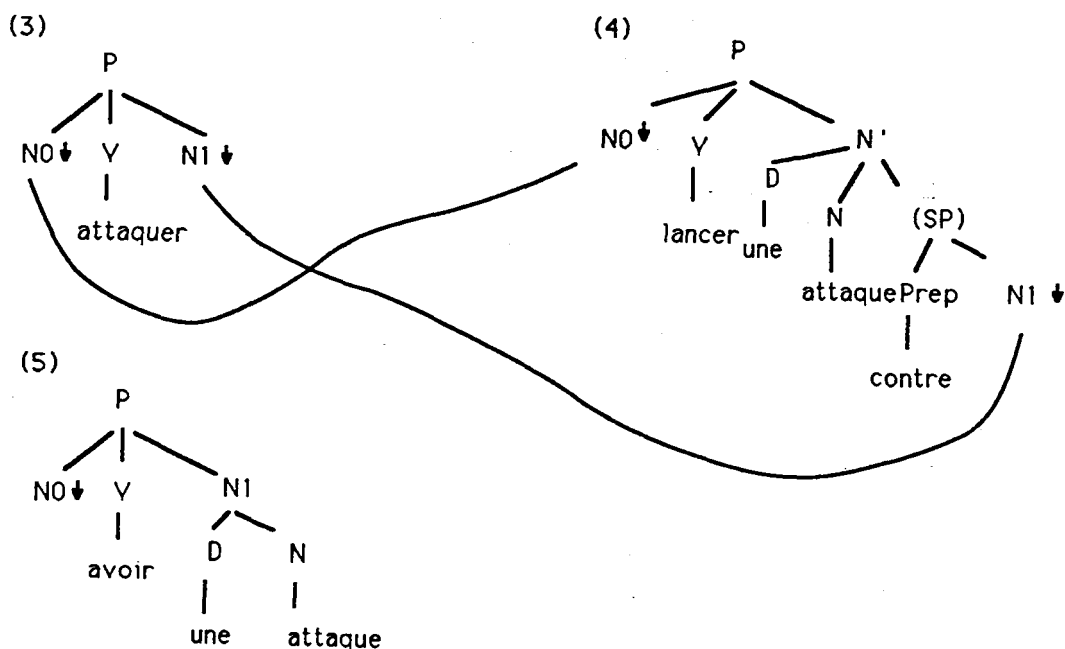


Figure 35

Il y a un parallèle à effectuer avec d'autres théories qui mettent en avant la notion de "catégorie", considérée comme ayant une structure interne possiblement complexe pour proposer une représentation unifiée, sous forme de structures de traits, des "préterminaux" lexicaux et des syntagmes. Les grammaires catégorielles vont jusqu'à ignorer une telle distinction et notent de la même façon un verbe intransitif et le groupe verbal formé par un verbe transitif et son objet direct: NP\S. Les grammaires syntagmatiques (GPSG et HPSG), elles, conservent la distinction entre préterminals et syntagmes mais en donne une représentation unifiée sous forme de structures de traits.

En GPSG les catégories syntaxiques sont exprimées (suivant en cela Z. Harris 1951 pour la définition d'un trait 'barre' et N. Chomsky 1971 pour la factorisation des parties du discours traditionnelles) en termes de structures de traits, spécifiées

en partie lexicalement, en partie lors de la combinaison en catégories plus complexes (par unification). L'essentiel des contraintes linguistiques s'exprime au niveau des principes portant sur ces structures de traits (compatibilité des traits à l'intérieur d'une structure, partage de traits entre structures) et non au niveau des règles de réécriture.

La *grammaire syntagmatique guidée par les têtes* (HPSG; C. Pollard et I. Sag 1987) poursuit le mouvement en considérant toutes les représentations de la grammaire comme des structures de traits (les règles de réécriture sont remplacées par des principes de bonne formation des structures de traits). Ces structures de traits peuvent être enchâssées les unes dans les autres et les relations de dominance être elles-mêmes exprimées comme un trait particulier. La composante syntagmatique est réduite à un trait "filie" (*daughter*) chargé de distinguer entre mots et syntagmes.

Il convient de noter qu'il y a formellement équivalence entre de telles structures de traits et une représentation en arbres. Ces représentations sont en effet des graphes acycliques orientés où chaque noeud a un père unique et le graphe une racine unique²⁵.

L'on pourrait encoder les arbres élémentaires des TAGs sous forme de structures de traits (plus complexes que les actuelles structures associées par HPSG aux items lexicaux si l'on veut que l'ensemble de la grammaire soit lexicalisé). La différence essentielle est dans l'opération qui combine ces structures entre elles, puisque que nous avons besoin d'un équivalent de l'adjonction. Celle-ci a été récemment redéfinie en terme de contraintes sur les structures de traits (K. Vijay-Shanker et A. Joshi 1990).

On peut aussi représenter les arbres élémentaires comme des structures de traits combinées par unification, une structure (possiblement complexe) étant associée à chaque arbre au lieu de l'être à chaque règle de réécriture comme dans les grammaires d'unification plus classiques (K. Vijay-Shanker 1987). L'opération de substitution s'exprime naturellement. En revanche, l'opération d'adjonction nécessite un aménagement particulier. Une possibilité est de définir les arbres initiaux comme des structures de traits "classiques" et les arbres auxiliaires comme des fonctions sur les structures de traits du sous-arbre du noeud où a lieu l'adjonction.

3.2.2. Les arbres dérivés

Les arbres dérivés correspondent à ce que les autres théories entendent par arbres²⁶. Ce n'est pas une notion originale aux TAG, et c'est une notion qui joue pour nous un rôle assez secondaire.

Les seules informations qui y apportées sont la flexion morphologique des items

²⁵ Il y aurait de même équivalence avec une représentation par automates finis.

²⁶ Avec cette différence que les arbres dérivés en TAG ont des chemins non contextuels (emboîtés) alors que les arbres des règles de réécriture hors contexte des autres théories ont des chemins réguliers. La capacité formelle est différente mais la notion linguistique est la même.

lexicaux en jeu et l'ordre des mots²⁷. Pour les retrouver il faut se reporter à l'arbre dérivé.

Ils encodent aussi certaines ambiguïtés syntaxiques: *Paul pense que Jean est venu aujourd'hui* (*aujourd'hui* attaché au noeud P de l'arbre *pense* ou à celui de *venu*):

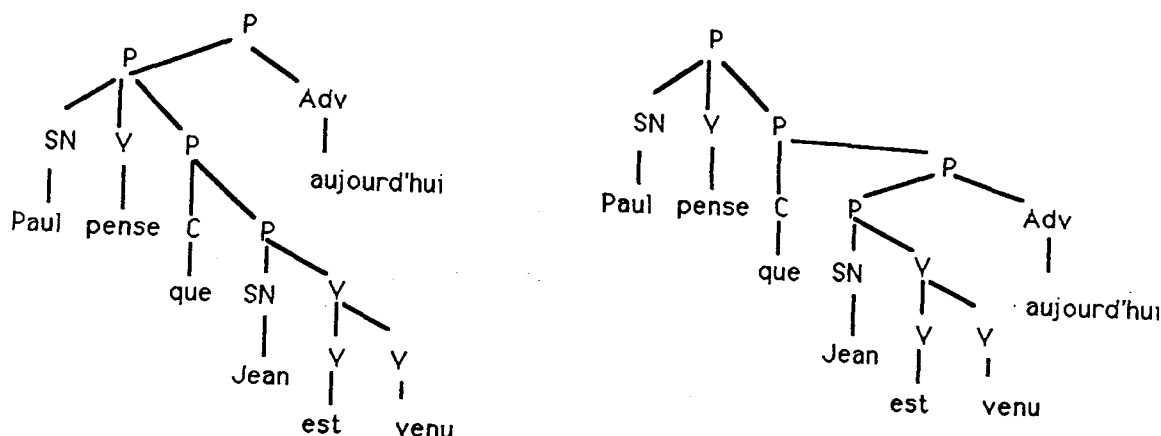


Figure 36

Mais une telle différence apparaît également dans les arbres de dérivation correspondants:

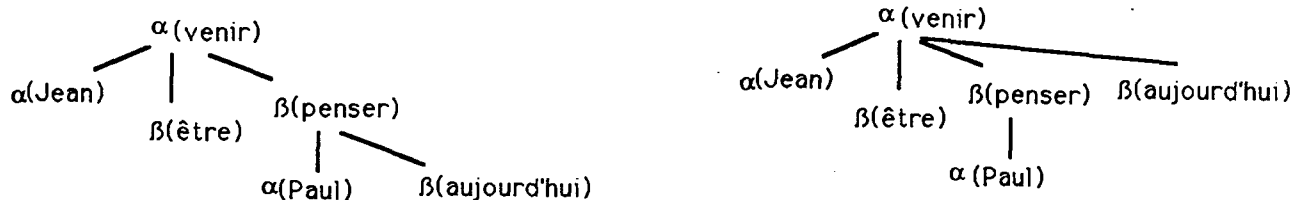


Figure 37

Ce ne sont donc pas les informations les plus importantes pour l'analyse syntaxique qui sont apportées par ce que nous appelons "arbres dérivés" et qui correspond à

²⁷ Une variante des TAG, proposée par Joshi 1987b, ôte les relations de précédence aux arbres élémentaires pour les exprimer par une composante séparée. Du coup, l'information apportée par les arbres dérivés est encore appauvrie.

ce que les autres théories appellent "arbres"²⁸.

3.2.3. Les arbres de dérivation

La différence entre arbre de dérivé et arbre de dérivation est essentielle en TAG. c'est une distinction propre aux grammaires qui ont des unités de base structurées (arbres élémentaires) et qui n'existe pas dans le cas des formalismes basés sur des règles de réécriture de chaînes(context-free).

L'arbre de dérivation et l'arbre dérivé sont deux "vues" d'une analyse (c'est à dire d'un couple séquence de mots/structure). A partir de l'arbre de dérivation on peut retrouver l'arbre dérivé mais l'inverse n'est pas vrai. Il y a donc plus d'informations dans l'arbre de dérivation.

Ils sont différents des "arbres de transformation" utilisés en grammaire transformationnelle (ou dans la grammaire en chaîne de N. Sager 1981) pour représenter l'histoire transformationnelle d'une phrase (quelles transformations ont été appliquées et dans quel ordre).

En TAG, l'histoire de la dérivation d'une phrase est également importante, mais c'est une histoire toute différente. Nous ne distinguons pas plusieurs étapes d'analyse et il n'y a ni règles ni principes généraux pour passer d'un arbre à un autre. (Il y a seulement les deux opérations de combinaisons des arbres entre eux que sont l'adjonction et la substitution) Cette histoire est exprimée par un arbre: l'arbre de dérivation mais c'est une histoire lexicale. Elle indique quelles structures lexicales ont été combinées entre elles pour former un arbre dérivé (et un seul) et non quels arbres dérivés (structures phrastiques complètes) ont été transformés en quels autres.

Les arbres de dérivation ont leurs branches étiquetées avec les adresses des noeuds où a lieu la combinaison entre arbres élémentaires.

L'arbre de dérivation a deux vertus: d'une part il "condense" l'arbre dérivé correspondant, d'autre part il le rend non ambigu.

Il le condense au sens où ce qui y apparaît comme un noeud est toujours un sous-arbre dans l'arbre dérivé (c'est-à-dire correspond à au moins deux noeuds: le mot et sa catégorie préterminale). Ses noeuds (qui sont très exactement les arbres élémentaires du lexique) sont donc les unités correspondant à la structure argumentale de la phrase et il fait apparaître plus explicitement les relations prédicats/arguments qui sont "noyées" dans l'architecture syntagmatique de l'arbre

²⁸On peut rapprocher ces arbres "pauvres" de ceux que fournissent les analyses des grammaires catégorielles d'unification. Ils sont également tenus pour inessentiels et c'est la structure de traits résultante qui est plus importante. La différence est que les arbres catégoriels sont forcément binaires, puisque ce formalisme repose sur une dichotomie foncteur/argument, et qu'ils sont en général plusieurs pour une même phrase (il y aura deux arbres possibles pour une phrase à verbe transitif selon que l'on combine le verbe d'abord avec son sujet ou avec son objet, qui seront réputés équivalents; cf L. Karttunen 1989 et C. Gardent, G. Bès 1989). Ces grammaires multiplient donc en fait les niveaux syntagmatiques (*spurious ambiguities*).

dérivé²⁹. Ainsi les arguments sont toujours dominés directement par leur prédicat dans l'arbre de dérivation, alors qu'ils peuvent en être infiniment éloignés (en termes d'ordre des mots et de niveau de profondeur) dans l'arbre dérivé.

Il lève l'ambiguïté de l'arbre dérivé en ce sens qu'à un même arbre dérivé peuvent correspondre plusieurs arbres de dérivation différents. C'est-à-dire qu'un même arbre dérivé peut être obtenu à partir de plusieurs ensembles distincts d'arbres élémentaires. Cela est utilisé pour rendre compte d'ambiguïtés sémantiques entre phrases ayant par ailleurs la même structure syntaxique, par exemple l'ambiguïté entre interprétation littérale et figée d'une phrase telle que : *Jean prend une veste*. Pour chacune des deux interprétations, l'arbre dérivé sera le même, mais la différence se fera par les arbres de dérivation. Dans le cas littéral, il y aura eu combinaison de quatre arbres élémentaires: celui de *Jean*, celui de *prendre*, celui de *un* et celui de *veste*; dans le cas figé, seuls deux arbres élémentaires sont en jeu: celui de *Jean* et celui de *prendre une veste*³⁰:

Jean prend une veste

Arbre dérivé

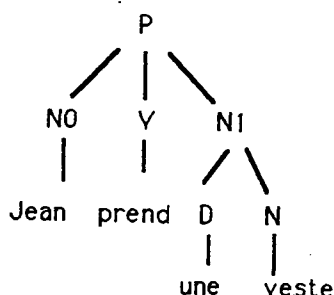


Figure 38

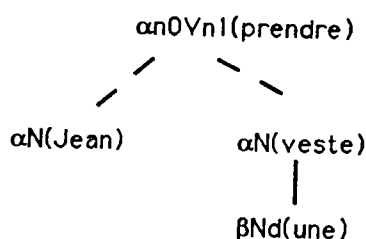
²⁹ Il y a une différence de position dans l'arbre dérivé entre arguments et ajouts (liée à leur différents modes d'insertion): les arguments peuvent être au même niveau que le prédicat dont ils dépendent, les ajouts sont toujours à un autre niveau ("au dessus" de ce qu'ils modifient).

Mais cela ne suffit pas pour avoir une définition purement configurationnelle (dans l'arbre dérivé) des arguments, puisque:

- dans tous les cas où l'on a affaire à un prédicat complexe ou à une construction à extraction certains arguments sont à un niveau différent de tout ou partie de leur prédicat.
- on ne peut savoir ce qui est le prédicat en regardant l'arbre dérivé, puisque c'est l'arbre élémentaire (dont le prédicat est par définition la tête lexicale) qui définit celui-ci.

³⁰ Les expressions figées sont en effet représentées par des arbres élémentaires étendus, dont la tête lexicale comporte tous les éléments figés de l'expression, et qui sont en général semblables à ce qu'on obtiendrait par la combinaison régulière de structures élémentaires déjà existantes. L'on rend compte ainsi à la fois (et de façon indissoluble) de la syntaxe régulière des expressions figées et de leur sémantique non compositionnelle (voir chapitre 4).

Arbre de dérivation (littéral)



arbre de dérivation (figé)

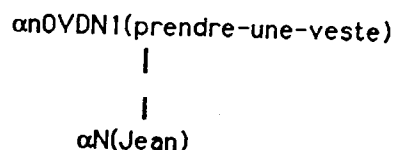


Figure 39

L'ambiguïté sémantique entre interprétation littérale et interprétation figée est représentée de façon structurale au niveau des arbres de dérivation, non au niveau de l'arbre dérivé, c'est-à-dire en fait en termes des unités lexicales (arbres élémentaires) combinées.

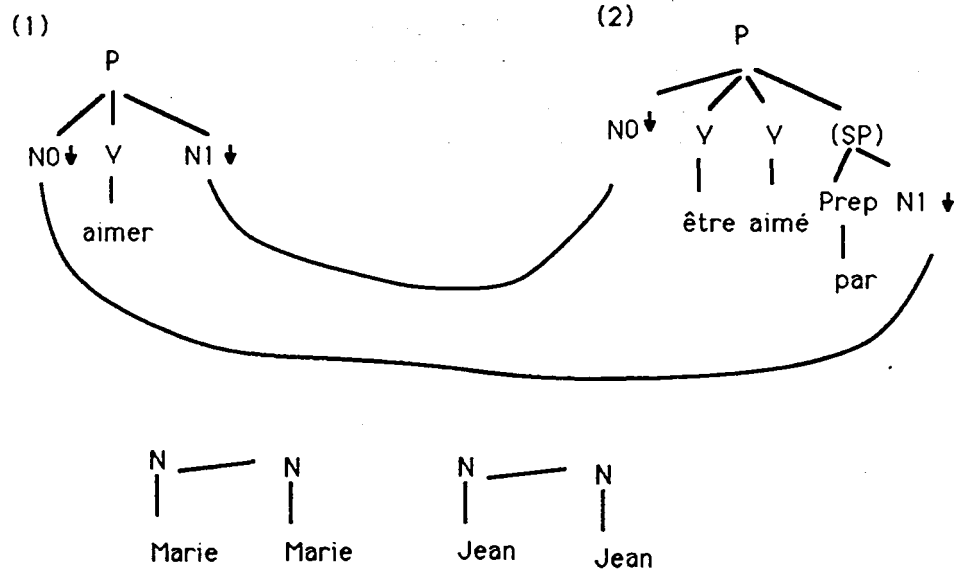
La relation de l'arbre dérivé à l'arbre de dérivation n'est donc pas du tout la même que celle de l'arbre de surface à l'arbre profond en grammaire générative classique. D'une part il s'agit de la même unité, considérée sous deux angles différents, d'autre part le lien entre les uns et les autres n'est pas assuré par des transformations mais par consultation des entrées lexicales.

Considérons maintenant le cas inverse de phrases à structures syntaxiques différentes mais à interprétation sémantique identique. A un arbre de dérivation ne peuvent, à proprement parler, correspondre plusieurs arbres dérivés, puisque l'arbre de dérivation "contient" l'arbre dérivé de manière univoque. Mais il est possible de définir des arbres de dérivation équivalents: lorsque des liens sont définis entre deux arbres élémentaires (c'est-à-dire entre leurs noeuds qui peuvent être à des adresses différentes), des arbres de dérivation différents sont considérés comme équivalents si les mêmes combinaisons (ou des combinaisons équivalentes) ont lieu aux noeuds ainsi liés. Ainsi l'équivalence entre une phrase active et la phrase passive correspondante s'exprime d'abord au niveau des arbres élémentaires (il s'agit de deux structures possibles pour un même verbe) et doit être conservée au fur et à mesure de la dérivation des phrases (qui peuvent être complexes):

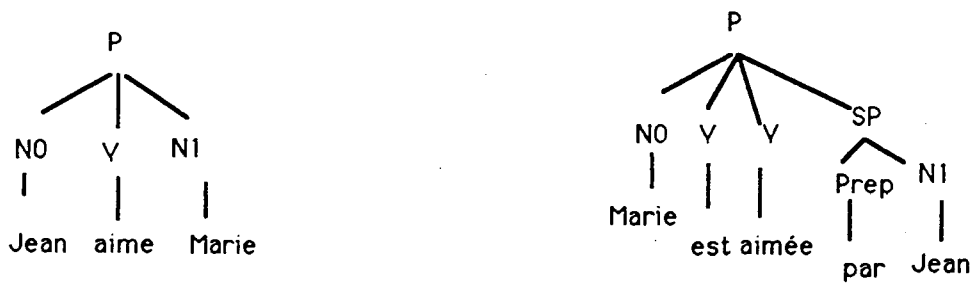
Jean aime Marie

Marie est aimée par Jean

arbres élémentaires



arbres dérivés



arbres de dérivation



Figure 40

Les dérivations sont considérées comme équivalentes car "synchrones" bien que les arbres de dérivation soient de structure différente.

Les arbres élémentaires de l'actif et du passif sont regroupés pour un même verbe à l'intérieur d'une "famille" d'arbres (voir chapitre 3). Les liens entre arbres élémentaires sont développés aux chapitres 3 et 7.

Chapitre II.

Justification des syntagmes employés

Une structure élémentaire est une structure lexicale (c'est-à-dire comportant au moins une feuille occupée par un élément lexical) qui correspond soit à une phrase simple, soit à ce qui, dans le cadre des lexiques-grammaires, s'analyse comme une réduction de phrase simple¹. Elle peut être de catégorie et de structure interne variées. Plusieurs questions générales se posent pour notre grammaire: quelles sont les catégories distinguées ? lesquelles sont préterminales, lesquelles peuvent être la racine de structures élémentaires ? quelles sont les têtes lexicales possibles pour une structure élémentaire donnée ? Nous allons les prendre successivement en commençant par la notion de tête, d'abord de façon générale, puis en considérant chaque syntagme l'un après l'autre.

Notre notion de tête est un peu différente que celle des autres théories linguistiques. Elle est plus simple que celle qu'on trouve dans la plupart des grammaires d'unification. Ainsi, la notion de tête en GPSG ou HPSG recouvre les têtes lexicales (c'est-à-dire des catégories préterminales) et les têtes syntagmatiques qui servent à propager d'un niveau à un autre un certain nombre de traits listés a priori comme traits de tête (par exemple <personne>, <mode> ou <nombre>). Ainsi, en GPSG ou HPSG le verbe est tête du syntagme verbal et le syntagme verbal tête de la phrase. Nos têtes sont toujours des têtes lexicales et ne portent que les traits qui y sont localement pertinents (essentiellement des traits morphologiques). Les traits concernant les autres catégories de la structure sont assignés directement au bon noeud sans devoir être copiés sur le noeud tête.

LFG accorde un statut particulier aux phrases à verbe finis qu'elle considère comme une catégorie exocentrique c'est-à-dire sans tête. Nous n'avons pas de telles catégories.

Enfin, la récente théorie générative (N. Chomsky 1981 et 1986) ne considère que des têtes lexicales (préterminaux) qui déterminent la catégorie de la structure dont elles sont têtes (théorie X-bar) et généralise la notion de tête à toutes les catégories, au prix d'une multiplication des catégories vides. Nous n'avons pas de têtes vides et nous ne suivons pas strictement un modèle de type X-bar.

Pour nous, les têtes sont les éléments lexicaux qui déterminent la structure argumentale de l'expression, c'est-à-dire le sujet et les compléments éventuels mais aussi la catégorie que la structure peut modifier s'il s'agit d'une structure d'ajout. Ainsi, le verbe est la tête de la structure de la phrase active (pour une phrase simple), mais c'est l'auxiliaire du passif et le verbe au participe passé qui sont conjointement têtes de la phrase passive (toujours pour la phrase simple). Pour les relatives, qui sont représentées comme des structures d'ajouts (arbres auxiliaires), c'est le pronom relatif (qui détermine l'antécédent de la relative) et le verbe (qui détermine la structure interne de la relative) qui sont conjointement têtes de la structure (Cf chapitre 3 section 3).

Les têtes ne correspondent pas forcément à un préterminal de la même catégorie que

¹ Les adverbes, groupes nominaux ... y sont en effet analysés comme réduction des phrases à verbe support correspondantes, Cf chapitre 3

celle du noeud racine de l'arbre qu'elles sélectionnent. En particulier les adverbiaux sélectionnent des arbres dont le noeud racine est de la catégorie de ce qu'ils modifient. Ainsi un adjectif épithète est tête d'une structure de racine N.

Un arbre élémentaire peut avoir plusieurs têtes, de catégories différentes et situées à des niveaux de profondeur variés. L'ensemble des têtes doit former le prédicat correspondant à la structure. On parle alors de têtes multiples (ou de co-têtes). C'est le cas notamment pour les expressions figées (Cf chapitre 4) mais aussi des verbes à complément prépositionnels (Cf chapitre 3). Ainsi le prédicat de la phrase simple *Jean ressemble à Marie* est *ressembler-à*; celui de *Jean se souvient de Marie*, *se-souvenir-de*. Dans le cas de structure à tête multiples, un ou plusieurs éléments têtes peuvent être optionnels (par exemple *de* dans *se-souvenir-de*).

Pour ce qui est de la structure interne des arbres élémentaires, nous distinguons le minimum de niveaux hiérarchiques. En fait, ce sont les expressions figées qui amènent à définir des structures avec un grand nombre de strates. Ce sont ces différents niveaux que nous allons maintenant définir et justifier.

On peut distinguer plusieurs types de critères pour définir des catégories non terminales:

a) "distributionnel"

Quand il y a commutation entre des suites différentes de catégories (pour une tête lexicale donnée), on a intérêt à les regrouper sous une même variable. Cela évite d'avoir à répéter à chaque fois toutes les combinaisons possibles. L'on doit ainsi avoir un symbole SN (ou l'équivalent) pour définir une suite du type:

$$SN = (Det) (Adj)^* N (Adj)^* (SP)^* (Rel)^* \dots$$

Il est d'ailleurs impossible de décrire toutes ces combinaisons sans appel soit à une définition récursive, soit à des variables internes (c'est-à-dire à d'autres catégories). C'est le cas ici puisque le complément prépositionnel, comme la relative, peut également comporter un SN (en excluant ici les relatives en *qui*):

$$\begin{aligned} SP &= Prep (Det) (Adj)^* N (Adj)^* (SP)^* (Rel)^* \dots \\ &= Prep SN \end{aligned}$$

$$\begin{aligned} Rel &= Pronom-rel (Det) (Adj)^* N (Adj)^* (Rel)^* V \dots \\ .. &= Pronom-rel SN V \dots \end{aligned}$$

La traduction dans notre cadre est de considérer comme une même catégorie un groupe et ses modifieurs éventuels, c'est-à-dire tout ce qui peut recevoir une adjonction.

b) "transformationnel"

Plusieurs règles sont possibles: est un constituant ce qui est remplaçable par un pronom clitique, par un pronom interrogatif, ce qui peut être extrait par *c'est .. qu-*, ce qui est

topicalisable... Ce type de critère permet de désambiguer un certain nombre de séquences considérées comme ambiguës avec le premier critère, par exemple: Det N Prep Det N = [Det N]SN [Prep N]SP ou [Det N [Prep N]SP]SN.

Un problème général se pose néanmoins. La classification complète des phrases simples du français menée au LADL a en effet montré que ces différentes règles ne fournissaient pas en général de résultats convergents (il peut y avoir convergence seulement à l'intérieur de petites classes de verbes). Ainsi selon la transformation considérée, on risque d'aboutir à des types de constituants différents. Nous avons donc choisi de retenir en cas de conflit un seul critère: l'extraction en *c'est... que*, qui est apparu comme le plus opératoire.

Ce principe apparaît dans notre cadre pour la définition des "familles" d'arbres (voir chapitre 3), c'est-à-dire les relations autorisées entre noeuds d'arbres élémentaires paraphrastiques (ou proches sémantiquement).

c) "par analogie"

C'est le critère proposé par M. Gross 1989 pour régulariser la description des expressions figées du français. Dans la plupart des cas, ces expressions relèvent du même découpage catégoriel et syntagmatique que les expressions "libres", comme en témoigne leur flexibilité syntaxique (passif, extraposition, nominalisation...). Par exemple, *N₀ mettre la main à la pâte* a bien le même découpage en deux compléments du verbe dans l'interprétation idiomatique et dans l'interprétation littérale, comme en témoigne la possibilité de (1) et l'interdiction de (2), pour les deux interprétations:

- (1) *La main a été mise à la pâte*
- (2) * *La main à la pâte a été mise*

Dans les quelques cas où aucune transformation n'est possible, c'est un raisonnement par analogie avec une expression libre homonyme qui permet le découpage. Ainsi, pour *N₀ donner sa langue au chat*, ni le passif ni l'extraction ne permettent de trancher puisque toutes les possibilités sont interdites (avec l'interprétation idiomatique):

- # *C'est sa langue que Jean a donnée au chat*
- * *C'est sa langue au chat que Jean a donnée*

C'est donc par analogie avec le découpage de la phrase dans son interprétation littérale (*au chat* complément de *donner* et non de *langue*), que la phrase figée reçoit le découpage: *N₀ donner (sa langue)SN (au chat)SP*.

Nous utilisons également ce critère.

Une notion particulière à notre formalisme est celle de "syntagme étendu". Elle découle en partie du principe qui consiste à faire apparaître au moins une tête lexicale dans une structure donnée. Les syntagmes aux noeuds racines des arbres élémentaires sont forcément étendus (par exemple les arbres phrastiques vus au chapitre précédent). Les syntagmes immédiatement dominés par le noeud racine d'un arbre auxiliaire (et

différents du noeud pied) le sont également. Ainsi des arbres auxiliaires comme les suivants pour un syntagme prépositionnel modifieur de phrase ou de groupe nominal sont mal formés car ils ne comportent pas de tête lexicale:

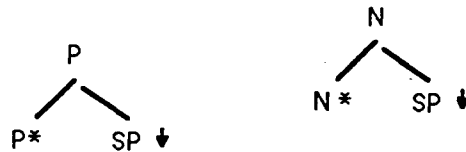


Figure 1

Il faut obligatoirement les "étendre" pour faire apparaître la valeur de la préposition:

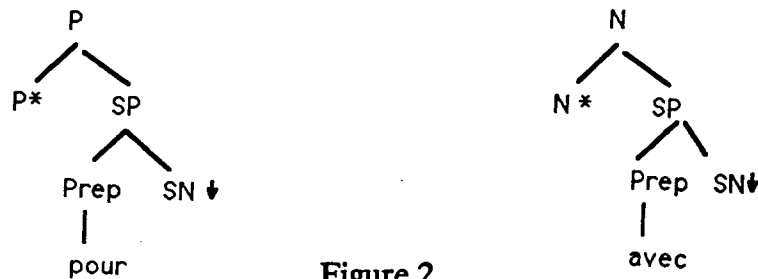


Figure 2

L'on peut ainsi faire apparaître aisément des dépendances entre la valeur de la préposition (ou la structure interne du groupe prépositionnel) et les catégories que le groupe prépositionnel peut modifier (Cf chapitres 3 et 7).

Nous utilisons également des syntagmes étendus dans d'autres cas que ceux strictement exigés par la condition de lexicalisation. Les syntagmes ne sont en général pas étendus dans l'arbre où ils sont arguments (par exemple les sujets sont des noeuds feuilles de catégorie groupe nominal ou phrase). Nous les "étendons" néanmoins soit quand ils ne se comportent pas comme des îlots (pour la sous-catégorisation ou l'extraction; Cf chapitre 5), soit quand il s'agit de compléments figés (que nous souhaitons faire apparaître comme co-têtes lexicales de l'expression ; Cf chapitre 4).

Nous allons maintenant examiner successivement les syntagmes suivants: nous définissons brièvement ce que nous entendons par phrase, nous expliquons ensuite pourquoi nous avons rejeté la notion de syntagme verbal, puis nous examinons la notion de syntagme nominal. Ensuite, nous définissons les différents types de syntagmes prépositionnels distingués dans notre grammaire. Nous considérons enfin les autres catégories syntagmatiques: syntagme adjectival et syntagme adverbial.

Ce chapitre présente les justifications linguistiques théoriques qui sous-tendent les structures arborescentes de notre grammaire TAG. Ces arbres élémentaires sont présentés aux chapitres 3 (phrases simples) et 4 (phrases à verbe support et phrases figées).

1. La phrase

Dans les grammaires hors contexte, le symbole Phrase est distingué comme étant le point de départ d'une dérivation bien formée.

En TAG, il faut distinguer le noeud étiqueté Phrase (qui peut être un noeud racine, un noeud intérieur ou une feuille), et les arbres élémentaires de type initial et de noeud racine Phrase qui sont les unités de base de la grammaire. Ce sont ces arbres initiaux qui sont le point de départ de toute dérivation bien formée. Le noeud racine ne porte pas forcément les mêmes informations (les mêmes traits) que le reste de l'arbre; par exemple, la valeur lexicale de la tête (le verbe dans les cas standards) qui joue un rôle essentiel dans la détermination des propriétés syntaxiques de la phrase simple n'est pas notée au niveau du noeud racine. De plus, certains arbres ont pour racine un noeud Phrase sans pour autant correspondre à une phrase simple, par exemple les adverbes modificateurs de phrase.

Les arbres élémentaires initiaux (et certains arbres auxiliaires) de racine Phrase correspondent à la notion de phrase simple (Z. Harris 1962, M. Gross 1975). Celle-ci se définit comme un prédicat (en général un verbe) et tous ses arguments. Nous représentons de la même façon les expressions figées et les phrases à verbe support.

Les arguments sont dans l'arbre élémentaire au même niveau que le prédicat, sauf pour les compléments prépositionnels ou certains compléments phrastiques (où l'on introduit un niveau SP et P'). Nous n'avons pas de noeud Syntagme verbal, qui placerait le sujet à un niveau distinct (voir section suivante). Pour nous le "syntagme verbal", au sens de domaine d'expansion du verbe, c'est la phrase simple.

Toutes les catégories peuvent être têtes d'arbres phrastiques élémentaires. La différence est que les verbes peuvent être à eux seuls têtes de P, tandis que les autres catégories ne le sont en général que dans la mesure où elles sont accompagnées d'un verbe, et il s'agit alors d'une phrase "à plusieurs têtes" c'est-à-dire figée ou à verbe support (voir chapitre 4)².

Dans quelques cas, des catégories non verbales seront considérées comme têtes de phrases à elles seules. C'est le cas notamment des phrases "nominales", souvent assez figées. Des noms comme *non* ou *bonjour* ont une structure phrastique à côté de leur structure nominale³:

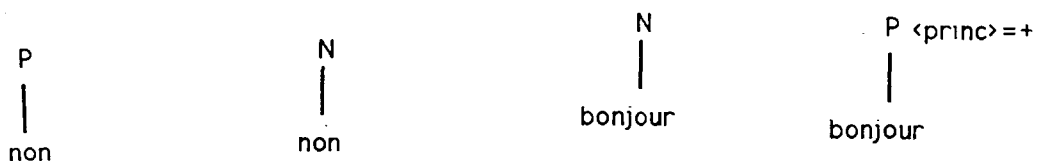


Figure 3

2 L'on pourrait rendre compte plus explicitement de cette présence d'un verbe en choisissant d'appeler la phrase syntagme verbal (c'est-à-dire SV et non P). Nous ne l'avons pas fait de crainte de créer des confusions, puisque nous rejetons parallèlement la notion "classique" de syntagme verbal.

3 Avec cette différence que la phrase en *bonjour* ne peut être enchâssée:

Jean répondit (que + :) non

*Jean répondit (*que + :) Bonjour*

Cela est représenté par un trait différent sur le noeud P.

En effet, nous n'avons pas de mécanisme transformationnel qui permettrait de les engendrer à partir d'une phrase verbale par ellipse du verbe. Mais le lien éventuel avec des phrases verbales synonymes est à prévoir (pour les liens de synonymie voir chapitre 7):

Joyeux anniversaire ! = je (te + vous) souhaite un joyeux anniversaire

Non ! = je (te + vous) dis (que) non

Dehors ! = je (te + vous) dis d'aller dehors

Nous n'avons donc pas de noeud SV, mais nous n'avons pas une structure totalement "plate" de la phrase pour autant, puisque les niveaux catégoriel (catégorie V) et lexical (valeur du verbe) sont toujours présents.

Nous regroupons également, sous un même syntagme phrastique, complémentateur et phrase enchâssée : *Je demande (qui tu as vu)* ou *Je crains (que Paul soit venu)*. En effet, les tests distributionnels ou transformationnels montrent bien qu'il s'agit de syntagmes (contrairement à la mise à plat effectuée par exemple par M. Salkoff 1973 dans sa grammaire en chaîne).

Je demande qui est venu

Je demande cela.

? C'est qui est venu que je te demande.

Je crains que Paul soit venu

Je crains cela.

C'est que Paul soit venu que je crains.

Nous considérons la plupart des compléments à l'infinitif comme des phrases (voir section suivante), et traitons de la même façon: *Je te demande (de venir)* ou *Il cherche (à venir)*.

Nous ne créons pas de catégorie spécifique du type S' ou C', comme en grammaire générative, la même catégorie nous suffit et des traits marquent la différence entre (*Jean part*), (*que Jean part*) et (*qui Jean aime*). En effet, les phrases à complémenteurs *que*, *si*, *à* ou *de* ne sont pas introduites de la même façon que celles introduites par un élément *qu-*. Les éléments de type *qu-* en position complémentateur appartiennent à l'arbre élémentaire du verbe dont il sont un argument, ce qui est naturel puisqu'un arbre élémentaire est par définition une structure prédicat-argument complète. Ils reçoivent la même catégorie que dans une autre position (N, SP). Les complémenteurs *que*, *à* ou *de*, qui sont considérés comme "vides" sémantiquement, appartiennent à l'arbre du verbe principal, ce qui ne les empêche pas de former un syntagme avec la complétive, celle-ci étant représentée par un "syntagme étendu" (pour plus de détails, voir chapitres 3 et 5):

Jean pense que Marie promettra à Paul que Bob partira

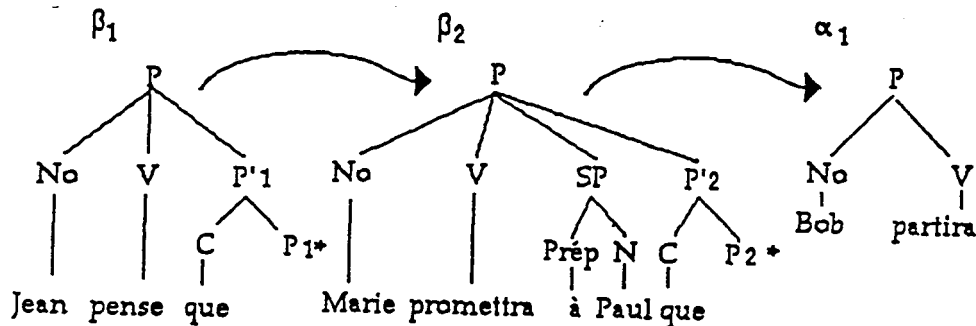


Figure 4

2. Pourquoi nous n'avons pas de syntagme verbal

Nous n'avons pas de "syntagme verbal" en tant qu'il serait différent de la phrase proprement dite. La notion de syntagme verbal, introduit dans *Structures Syntaxiques* est une notion confuse, source de nombreuses discussions et nous allons discuter en détail pourquoi nous la rejetons.

La notion de syntagme verbal, inspirée de la notion de prédicat, reprise par la grammaire générative à une longue tradition logico-grammaticale⁴, est utilisée par la plupart des théories linguistiques contemporaines⁵, en tant qu'universel langagier ou de façon plus limitée pour des langues à ordre SVO comme le français ou l'anglais.

C'est pourtant une notion fragile, dont les justifications explicites sont d'ailleurs peu nombreuses dans la littérature, et qui a été critiquée à plusieurs reprises, du moins pour le français. Nous allons reprendre brièvement ces discussions théoriques, en développant des arguments esquissés dans M. Gross 1975 et en ajoutant quelques autres, avant de présenter la solution que nous avons adoptée pour notre grammaire TAG, à savoir une structure de phrase sans SV intermédiaire.

La discussion porte sur deux points distincts: la définition de la phrase et le statut des infinitives (et des gérondifs pour l'anglais). L'idée que toute phrase a un sous-constituant

4 Comme le note R. Donzé 1971, "il est difficile de dire exactement quand les termes de sujet et de prédicat ont passé de la logique à la grammaire". Port-Royal, par exemple, propose une tripartition logique de toute proposition en sujet, attribut (prédicat) et affirmation (copule), tout en reconnaissant que la réalisation linguistique de ces termes peut prendre des formes variées: amalgame attribut-affirmation (*Petrus VIVIT* = *Petrus est vivens*), amalgame sujet-affirmation (*SUM homo* = *ego sum homo*), amalgame sujet-attribut-affirmation (*VIVO* = *ego sum vivens*), amalgames plus complexes (*PUDET me* = *pudor est tenens me*). C'est le premier type d'amalgame qui a, semble-t-il, conduit la plupart des grammairiens à partir de Dumarsais à adopter une analyse bipartite des phrase non copulatives.

5 Aux notables exceptions près des lexiques grammaires de M. Gross, des grammaires relationnelles, des grammaires de valence et de certaines versions des grammaires catégorielles.

SV est la plus généralement partagée en linguistique théorique comme en linguistique informatique. Le second point, qui se limite pour le français aux infinitives, est l'objet, notamment depuis J. Bresnan 1978, d'une abondante controverse qui n'est pas close et qui oppose grammairiens chomskyens (pour qui les infinitives sont des phrases) et grammairiens "unificationnistes" pour qui ce sont des groupes verbaux (LFG, GPSG, HPSG). Nous allons examiner ces deux aspects successivement.

2.1. SV comme sous-constituant de la phrase

Cette conception assez répandue a été pourtant critiquée à plusieurs reprises.

L. Tesnière, dès les années 50, critiquait ainsi, au nom de la syntaxe, la distinction traditionnelle entre sujet et prédicat en tant qu'elle donne un statut privilégié au sujet, alors que selon lui le sujet est "un complément comme les autres"⁶. Plus récemment, M. Gross 1975 a remis en cause la notion de syntagme prédicatif comme celle de syntagme verbal, mais dans le cadre d'une mise en cause plus générale de la grammaire générative. S. Keyser et P. Postal, dans leur manuel de grammaire générative de 1978, utilisent une structure phrastique "plate" sans syntagme verbal, mais sans s'en justifier davantage. J.C. Milner 1984 suggère en note la fragilité théorique de cette notion, mais s'attache davantage à critiquer la notion d'arbre syntaxique en général. Nous avons déjà développé notre position au sujet des représentations arborescentes (Cf chapitre 1). Nous voulons simplement montrer ici que, du moins pour une langue comme le français, ce constituant n'a pas de raison d'être.

2.1.1 définition distributionnelle

Jean (dort + n'a pas dormi + regarde Marie passionnément + lui donne une pomme...)

La commutation de la suite de catégories [(Adverbes)* (Clitiques) (Aux) Verbe (Compléments) (Adverbes)*] avec un verbe simple a pu conduire à assigner à celle-ci un statut de constituant, de type verbal. C'est le principe des analyses distributionnelles classiques, qui tendent à regrouper les éléments substituables sous un même type de syntagme.

Si l'on se reporte au texte de R. Wells 1947, qui fonde le principe de l'analyse en constituants immédiats, il justifie par la possibilité de substituer *The king of England* par *John* et *opened Parliament* par *worked* une coupure binaire de la phrase *The king of England / opened Parliament*. Mais, comme le note Wells, l'on pourrait envisager de couper la même phrase après *The king* (qui commute également avec *John*) ou avant *Parliament*. Il faut donc ajouter d'autres critères, qui ne sont pas purement distributionnels. R. Wells prend pour principe de choisir toujours, en cas de doute, le découpage binaire qui aboutit aux séquences les plus courtes: ainsi *The king of England*

⁶ La distinction sujet-prédicat est pour L. Tesnière un obstacle à toute analyse syntaxique par permutation entre sujet et compléments, c'est-à-dire par permutation d'actants. Il proposait notamment d'analyser ainsi la relation passif-actif (L. Tesnière 1959).

/opened the Parliament session peut être réduit en *John / worked*, alors que *The king of England opened / the Parliament session* serait réduit seulement à *John saw / Mary*. C'est un principe arbitraire et comme tel difficilement discutable.

Ce qui nous intéresse ici, c'est que cette définition distributionnelle éclate dès que l'on veut la relier à des propriétés syntaxiques ou interprétatives régulières.

R. Wells lui-même souhaitait corréler ses divisions à un principe syntaxique: chaque constituant distingué doit être autonome et indépendant, c'est-à-dire pouvoir former "une phrase en soi" ("a sentence by itself"). Ce qui est le cas, selon lui, pour *The king of England*, qui répond à la question 1: *Who opened Parliament ?* et pour *Opened Parliament* qui répond à la question 2: *What did the king of England do ?*, mais non pour une suite telle que *of England opened Parliament* qui ne répond à aucune question.

Mais il faut noter que la question 2 est en fait réservée à certaines phrases seulement (celles qui ont une interprétation "agent-action"). Il faut également noter qu'elle ne s'applique pas au français. De plus, même pour l'anglais, la similarité apparente entre la séquence qui suit le sujet et celle qui sert de réponse à la question 2 est illusoire. Si la phrase de départ est au présent, la réponse obtenue à la deuxième question ne sera pas forcément une séquence équivalente à celle de départ:

The king of England opens Parliament.

What does the king of England do ?

- * *Opens Parliament.*
- *Open Parliament.*

Et si la phrase de départ comporte un auxiliaire, celui-ci ne sera pas repris dans la réponse:

The king of England could open Parliament.

*What (could + *did) the king of England do ?*

- * *could open Parliament.*
- *Open Parliament.*

On voit là en germe toutes les discussions ultérieures sur l'appartenance ou non des auxiliaires et des marques flexionnelles au groupe verbal (la théorie chomskienne faisant de ces derniers des constituants n'appartenant pas à SV, ce qui n'est pas le cas pour des théories comme LFG ou GPSG). Ces difficultés vont aller croissant avec la théorie transformationnelle.

2.1.2 définition transformationnelle

La définition distributionnelle ci-dessus est intenable dans un cadre transformationnel, où la notion de syntagme doit s'appuyer sur des règles de déplacement ou d'effacement qui lui fassent référence. Plusieurs définitions alternatives ont été proposées, mais sans qu'aucune conclusion claire ne s'en dégage (tout au moins pour le français).

M. Gross notait en 1975 : "Il n'existe aucune transformation qui mettrait en jeu un

déplacement de SV. Dans les cas limités (conjonction de coordination) où l'on pourrait avoir l'idée de déplacer un SV, il existe toujours une solution concurrente et au moins aussi motivée qui ne comporte que des déplacements de SN"(p. 33)⁷. C'est ce point que nous allons développer maintenant, notamment à la lumière de discussions plus récentes. L'on peut en effet montrer que les phénomènes invoqués aboutissent toujours à des "constituants" différents.

La constitution interne du syntagme verbal a considérablement varié au cours de l'histoire de la grammaire générative. Il est remarquable que ces discussions n'aient abouti à aucun test fiable pour déterminer si un groupe donné appartient ou non à SV. Nous n'avons retenu que quelques exemples, parmi les plus significatifs.

1) Dans *Structures Syntaxiques* (1957), Chomsky s'appuie sur les phénomènes de coordination pour justifier ce qu'il appelle syntagme verbal (SV), identifié à toute la phrase moins le groupe sujet⁸. Il faut d'abord noter que c'est le seul cas où un syntagme n'est défini par aucune transformation à part la coordination. Il est aisé de montrer la faiblesse de cette justification mais ses conséquences pour les applications informatiques méritent qu'on s'y arrête. Nous allons examiner successivement les coordinations avec et sans ellipse.

a) coordination sans ellipse

Dans une théorie de la coordination limitée à la coordination de constituants, le noeud SV a engendré des prédictions manifestement erronées. Ainsi, N. Chomsky 1957 s'appuie sur cette notion pour interdire la phrase (1) (ou lui affecter des traits phonologiques spéciaux en tant que "semi-grammaticale") tout en autorisant pleinement la phrase (2):

- (1) *John enjoyed and my friend liked the play*
(Jean a apprécié et mon ami a aimé la pièce)
- (2) *John enjoyed the book and liked the play*
(Jean a apprécié le livre et aimé la pièce)

Une telle distinction n'est pas juste, et il existe d'autres phrases sur le même modèle que (1), par exemple: *Jean importe et Luc exporte des bananes*, qui sont presque aussi naturelles que (2). Il faut donc admettre d'une part que la coordination peut dupliquer des groupes au delà des frontières de constituants (ou dupliquer des constituants en introduisant des "gaps" dans certains), d'autre part que des différences d'acceptabilité ou

⁷ On trouve une idée similaire dans J.-C. Milner 1984 : "Il est douteux que la notion de groupe verbal ou de V" ait un véritable statut: alors que la notion de groupe nominal est constamment mentionnée dans les propositions de la théorie [chomskyenne], qu'il s'agisse de transformations ou de règles d'interprétation ou de conditions sur les règles), alors qu'elle y est effectivement traitée comme un groupe, c'est-à-dire comme une unité, la notion de groupe verbal apparaît n'être jamais indispensable" (p. 9).

⁸ N. Chomsky reprend donc sous le terme SV ce que les grammaires traditionnelles appelaient prédicat: tout ce qui n'est pas le sujet dans la phrase, et non ce qu'elles appelaient "groupe verbal" c'est à dire le verbe et son entourage immédiat (auxiliaires, éventuellement certains adverbes).

d'intonation sont difficiles à mettre en évidence pour distinguer coordination respectant et coordination violant les frontières de constituants. Ceci est clair si l'on considère d'autres cas de coordination de plus d'un constituant et parfaitement naturels tels que: *Jean a offert des pommes à Marc et des oranges à Catherine.*

b) coordination avec ellipse

Il est aujourd'hui généralement admis qu'une théorie de la coordination ne peut se passer d'un système d'ellipse (ou d'un mécanisme équivalent). Un autre argument en faveur de SV est alors de considérer que l'ellipse ne peut s'appliquer qu'à des constituants. Le groupe SV serait donc nécessaire pour définir l'ellipse constatée dans la phrase suivante (à partir du moment où les adverbes de temps sont en dehors de SV):

Jean va à la pêche le jeudi et Paul le lundi.

Il n'en est rien, car le matériel élide par coordination ne peut servir à définir les constituants de façon stable. Ainsi on peut élider le verbe avec ou sans certains de ses compléments, avec ou sans adverbes, avec ou sans certains auxiliaires. A. Zribi-Hertz 1986 donne des exemples, transposant au français des discussions plus anciennes sur l'anglais:

Marlon [a parlé]; à Raquel [de notre projet]; et Burt e; à Jane.

Marlon [a parlé à Raquel]; de notre projet, et Burt e; de nos ennuis.

Marlon [a mangé]; rapidement [son éclair]; et Raquel e; lentement.

Marlon [a mangé rapidement]; son éclair, et Raquel e; son baba.

Marlon [veut essayer de commencer à écrire]; une pièce, et Raquel e; un roman.

L'ellipse (*gapping*) peut donc s'appliquer à des séquences, continues ou discontinues, qui ne forment pas forcément un syntagme (selon les définitions habituelles)⁹. Il n'y a donc pas là d'arguments pour un SV en structure syntaxique de surface.

c) application à l'analyse automatique

La différence semble donc être quantitative. Si l'on a un groupe SV, l'on pourra considérer un peu plus de cas de coordination comme simple duplication de constituants, tandis que sans groupe SV, un peu plus de phrases coordonnées seront considérées comme des coordinations à *gapping*.

Les constructions à ellipse sont plus difficiles à traiter d'un point de vue formel: si l'on peut élider n'importe quelle suite de constituants, on aboutit à une grammaire susceptible d'engendrer tous les langages récursivement énumérables; ce qui n'est pas très plausible s'agissant des langues naturelles et conduit à des traitements inefficaces et des problèmes potentiellement indécidables.

Cependant, si l'on désire construire une grammaire de bonne couverture linguistique, il faudra bien intégrer une procédure d'ellipse, en la maintenant dans des limites qui

⁹ A. Zribi-Hertz propose de considérer les éléments "gappés" comme formant un constituant en forme logique.

restent d'ailleurs à définir. Le "gain" apporté par SV, et dont se contentent la plupart des grammaires informatisées actuelles¹⁰, est donc illusoire. On court même le risque, à partir du moment où l'on a à la fois un constituant SV et un mécanisme d'ellipse, d'aboutir à une multiplication mal contrôlée des analyses, ce qui est une menace constante pour les systèmes d'analyse automatique.

Nous n'avons pas fait une étude complète, mais si l'on prend deux grammaires, comportant toutes deux un mécanisme de duplication de constituants et autorisant l'ellipse, l'une sans SV et l'autre avec SV, il semble que celle avec SV est linguistiquement moins adéquate. En effet, l'alternative entre analyse par duplication de constituants et analyse avec ellipse semble correspondre à des cas d'ambiguïtés linguistiques dans la grammaire sans SV, alors que la grammaire avec SV aboutit à des ambiguïtés peu justifiées linguistiquement.

Ainsi une phrase à sujets coordonnés telle que:

L'oxygène et l'hydrogène explosent.

est analysée par les deux grammaires comme ambiguë: coordination (par duplication) de SN ou coordination de phrases simples (avec ellipse du verbe). Cette ambiguïté semble linguistiquement justifiée puisque (sauf à faire intervenir des connaissances relevant de la chimie plus que de la linguistique) on ne sait pas si c'est le mélange des deux qui explose (coordination de SN) ou s'il s'agit de deux explosions indépendantes (coordination de phrases). En revanche, la phrase suivante:

(3) *Marie mange une pomme et boit du café*

est analysée comme ambiguë seulement par la grammaire avec SV (duplication de SV ou coordination avec ellipse du sujet). Une telle ambiguïté ne semble pas justifiée. De même, une phrase telle que :

(4) *Marie voudrait du whisky et de l'eau* (O. Ducrot 1968)

devient triplement ambiguë: coordination de SN objets, de SV avec effacement du verbe, ou de phrases avec effacement du verbe et du SN sujet.

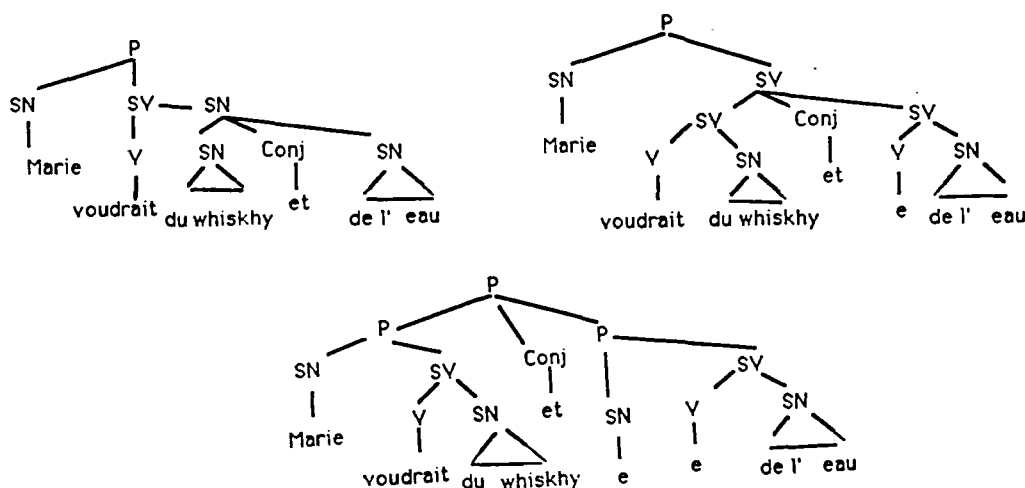


Figure 5

10 Aux notables exceptions près de la grammaire en chaîne de M. Salkoff et de certaines grammaires catégorielles.

Alors que la grammaire sans SV ne donne que deux analyses: duplication des SN objets ou coordination de phrases (avec ellipse). Il semble bien que deux interprétations seulement, selon que Marie veut un verre de ce mélange ou deux boissons distinctes, soient possibles¹¹.

De plus la grammaire avec SV conduit à des dissymétries entre des phrases que la théorie linguistique devrait pourtant avoir pour tâche de regrouper. En effet, dans la grammaire avec SV, l'on crée une dissymétrie entre les cas de coordination avec sujets différents, qui ne pourront être analysés que comme coordinations de phrases, et les autres qui seront analysés comme ambigus entre coordination de phrases et coordination de SV. Prenons par exemple une paire de phrases telle que :

- (5) *Marie adore mais Jean déteste cette fille*
- (6) *Cette fille est adorée par Marie mais détestée par Jean*

La première ne sera pas analysée comme ambiguë par une grammaire avec SV, puisqu'il ne peut s'agir que d'une coordination de phrases. En revanche son passif, qui n'est pourtant pas plus ambigu, recevra une double analyse, puisqu'il s'agit soit d'une coordination de phrases, soit d'une coordination de SV.

Le lecteur peut aisément vérifier qu'avec les mêmes règles de coordination, mais une grammaire sans SV, les phrases (3), (4) et (6) cessent d'être faussement ambiguës.

Dans notre grammaire TAG actuelle, nous n'avons implémenté que la coordination par duplication de constituants. (voir chapitre 6 pour plus de détails et des suggestions pour la coordination à ellipse).

2) Dans *Aspects de la théorie syntaxique* (1965), N. Chomsky s'appuie sur un autre critère, les phénomènes de sous-catégorisation, qui n'est d'ailleurs curieusement pas de nature transformationnelle, et aboutit à une nouvelle définition du groupe verbal: la phrase moins le sujet, moins les auxiliaires, moins certains adverbes.

N. Chomsky distingue les phénomènes de sous-catégorisation stricte des simples restrictions de sélection et pose le postulat que les premiers sont locaux, c'est-à-dire que les verbes sont sous-catégorisés seulement par les constituants de même niveau. Ce postulat a été peu critiqué (les phrases à verbe support sont néanmoins des contre-exemples; Cf chapitre 4); en revanche l'idée que la sous-catégorisation des verbes se ferait par les seuls compléments l'a été davantage.

Les discussions ont surtout porté sur les sujets. Il est vrai qu'il y a des verbes qui ne peuvent avoir de compléments mais pas de verbes qui ne peuvent avoir de sujets. Mais la sous-catégorisation ne portant pas seulement sur la présence ou l'absence d'un constituant mais aussi sur sa catégorie, la distinction nécessaire entre verbes autorisant un sujet phrastique (*ennuyer, réjouir...*) et verbes à sujet uniquement nominal (*manger,*

11 La même chose pour une phrase telle que: *Le curé a marié Marie et Pierre*, qui est doublement ambiguë (selon qu'il les marie ou non l'un avec l'autre), mais non triplement ambiguë.

fumer.) montre bien l'existence d'un parallélisme entre sélection des sujets et sélection des compléments. Cela est encore plus net dans les cas où cette sélection doit se faire simultanément, comme par exemple avec les verbes symétriques qui sélectionnent des couples sujet-compléments:

*Jean alterne avec (Marie + * la pluie)*
Le soleil alterne avec (Marie + la pluie)*

Il a aussi été noté que les sujets phrastiques commutaient systématiquement avec des syntagmes nominaux abstraits (ou des nominalisations) alors que cela est moins systématique pour les compléments phrastiques. De là l'idée que la sélection des sujets serait un phénomène plutôt sémantique et non une stricte sous-catégorisation syntaxique. Mais il n'est pas vrai que, réciproquement, un groupe nominal correspondant sémantiquement à une phrase, du type *le fait que P*, puisse toujours commuter avec un sujet phrastique, comme l'ont noté par exemple J. Bresnan 1982 et G. Gazdar et al 1985 pour l'anglais:

The fact that we invested on time made us rich
** That we invested on time made us rich (J. Bresnan 1982)*
 (Le fait + *E) que nous ayons investi à temps nous a enrichi

En français également, deux verbes acceptant des nominalisations pour sujets peuvent différer en ce que l'un acceptera un sujet phrastique et pas l'autre, et il s'agit donc bien d'une contrainte syntaxique :

Le départ de Paul importe à Marie
Que Paul soit parti importe à Marie.

La difficulté de ce travail rebute Marie
 (Le fait + ?*E) que ce travail soit difficile rebute Marie.

La sous-catégorisation du sujet est donc une propriété lexicale du verbe (ou de l'adjectif dans les phrases copulatives ou du groupe verbe-nom dans les phrases à verbe support; voir chapitre 4)¹².

La sous-catégorisation des auxiliaires a été moins étudiée. N. Chomsky 1965 mentionne lui-même en note le cas du présent progressif interdit en anglais avec certains verbes (*own, know.*). L'on peut trouver en français des interdictions similaires portant sur l'auxiliaire de temps (en plus des contraintes évidentes portant sur sa valeur lexicale (*avoir* ou *être*) qui ressortissent, elles, à des phénomènes de sélection. M. Gross 1989 donne un tel exemple à propos du verbe *remonter* dans la construction: *L'accident remonte à 10 ans*. Ce verbe n'admet aucun auxiliaire de temps:

¹² Pour des distinctions entre propriétés syntaxiques des sujets et des compléments phrastiques, voir le chapitre 5 sur les phénomènes d'extraction. Celles-ci sont sans lien direct avec la sous-catégorisation.

- * *L'accident (a + est) remonté à 10 ans*
- * *L'accident (aurait + serait) remonté à 10 ans...*

Il semble que l'on puisse trouver d'autres exemples qui tous montrent l'existence d'un lien entre cadre de sous-catégorisation du verbe et possibilité d'avoir un auxiliaire de temps (le verbe *dater* dans *Ce vin date de 1956*, le verbe *aller* dans *Cette robe va à Marie...*)

Les phénomènes de sous-catégorisation ne poussent donc pas à regrouper verbe et compléments tout en éloignant sujets et auxiliaires.

N. Chomsky 1965 distingue également deux types d'adverbes: ceux dont la combinaison est obligatoire ou interdite avec certains verbes (e.g. adverbes de manière) et ceux qui sont censés se combiner librement avec n'importe quel verbe (e.g. adverbes de temps ou de lieu). Il rattache les premiers seulement au syntagme verbal (les seconds étant mis au même niveau que les auxiliaires formant un syntagme prédicatif qui domine le syntagme verbal).

Cependant, en l'absence d'étude complète sur les co-sélections verbes-adverbes, le critère est inopérant; et tous les adverbes semblent pouvoir être interdits par certains verbes (Cf G. Lakoff et J. Ross 1966 pour l'anglais, M. Gross 1975 et 1990 pour le français).

3) J. Ross 1967 et J. Emonds 1970 ont proposé des critères véritablement transformationnels. Ils s'appuient d'une part sur la formation des "tag-questions" qui suggère une règle d'effacement de tout ce qui n'est ni le sujet ni l'auxiliaire:

The king opened Parliament
Did he open Parliament ?
 - Yes, he did (E + open Parliament).

Ce phénomène semble propre à l'anglais et nous ne le discuterons pas ici. Notons toutefois qu'avec ce critère, tous les adverbes devraient appartenir au syntagme verbal, ce qui n'est donc pas la même définition que N. Chomsky 1965.

Ils s'appuient d'autre part sur des phénomènes d'ellipse et de détachement, pour lesquels il proposent les règles "VP-deletion" et "VP-fronting", comme par exemple:

James Dean was crying in court, and James Mc Cord was too.
 James Dean pleurait au tribunal et James Mc Cord aussi.

The CIA guards our freedom and the FBI does too. (Ross 1967)
 La CIA protège notre liberté et le FBI aussi

They all said that John would pass an exam one of these days, and pass an exam he did ! (J. Emonds 1970)

Ils disaient tous que John allait réussir un examen un de ces jours, et il l'a fait !

Observables avec la plupart des auxiliaires anglais, ces phénomènes sont plus limités en français (essentiellement avec *pouvoir*, *devoir*, *avoir failli*, *être en train de*)¹³. Des exemples ont été donnés par J. Emonds 1978 et A. Zribi-Hertz 1986:

- ? *Vous auriez pu prendre des vacances, et nous aurions pu aussi.* (Emonds)
 ? *Jean a dit que je pourrais manger ces chocolats, mais je ne dois pas.* (idem)
Pierre n'a pas acheté un hélicoptère mais il a (bien) failli. (Zribi-Hertz)
Pierre n'a pas fini de taper sa thèse, mais il est en train. (idem)

De telles constructions sont très contraintes en français. On peut montrer qu'elles n'aboutissent pas à une définition plus satisfaisante d'un constituant SV.

a) Ellipses de SV

On observe celles-ci dans plusieurs contextes: coordination, conjonction de subordination, comparatives, relatives.

- Il a mangé tous les gâteaux mais il n'aurait pas dû.*
Il a mangé tous les gâteaux dès qu' il a pu.
Il a mangé plus de gâteaux qu'il n'aurait dû.
Il a mangé tous les gâteaux qu'il a pu.

Ces constructions ont été abondamment discutées pour l'anglais (surtout les cas de coordination et de comparaison). Elles conduisent à des difficultés pour le statut des adverbes, puisque les deux ellipses sont possibles dans le même contexte:

- Marie voudrait aller à Princeton cette année mais elle ne devrait pas, vu qu'elle pourra aussi l'an prochain.* (adapté au français d'après I. Sag 1976)

La première ellipse, après *devoir*, est (*aller à Princeton cette année*), la deuxième, après *pouvoir*, est (*aller à Princeton*). Il faut donc admettre, avec I. Sag 1976, que l'ellipse peut ne porter que sur "une partie de SV", ce qui montre bien qu'elle n'est pas un test pour la détermination des syntagmes¹⁴. En fait, pour les contextes de coordination (ou de comparaison), le même raisonnement que ci-dessus (2.1.2 1-b) s'applique : ce sont des cas de "gapping" et on ne peut les utiliser pour la détermination des constituants.

Pour les ellipses dans les relatives, qui semblent limitées exclusivement à *pouvoir* et *devoir* en français, il est également difficile d'y voir de bons indicateurs syntagmatiques. L'ellipse doit porter sur le verbe et le SN complément moins la relative, c'est-à-dire sur une "partie de SV", sinon il y aurait risque de régression à l'infini. De plus, il n'est pas

¹³ Le problème de l'appartenance ou non à SV des autres auxiliaires reste donc entier.

¹⁴ Supposer un SV récursif (du type SV - SV Adv), et donc deux SV emboîtés dans ((*aller à Princeton*) *cette année*) ne résout pas le problème puisqu'il faudrait expliquer pourquoi l'ellipse ne porte pas toujours sur le SV supérieur. La conclusion de I. Sag 1976 était que ces ellipses portent en fait sur des "constituants en forme logique", et peuvent donc être assez indépendantes des structures syntagmatiques.

clair qu'il y ait véritablement ellipse du verbe et de ses compléments éventuels, comme le montre le contraste suivant:

Jean a donné des gâteaux aux filles à qui il a pu en donner
 ?* *Jean a donné des gâteaux aux filles à qui il a pu.*

La difficulté à avoir une préposition devant la relative réduite s'expliquerait mieux si l'on voyait là plutôt un cas de construction transitive de *pouvoir* (du type: *N₀ pouvoir quelque chose*)¹⁵. Ce qui est confirmé par la possibilité d'une relative marginale en *que* avec *pouvoir* seul, qui serait impossible quand on a aussi le verbe et son premier complément :

? *Jean a donné des gâteaux aux filles qu'il a pu*
 * *Jean a donné des gâteaux aux filles qu'il a pu donner des gâteaux*

b) topicalisations de SV

A. Zribi-Hertz 1986 donne les exemples suivants de topicalisation¹⁶:

- (1) *Accorder le piano, quand même, il aurait pu*
- (2) *Accorder le piano, vraiment, il aurait dû*
- (3) *Etre malade, alors là, il a bien failli*
- (4) *Rédiger sa thèse, justement, il est en train*

De telles phrases, qui relèvent en général de l'oral ou d'un style familier, sont difficiles à juger. On peut simplement remarquer que la topicalisation est en général un phénomène très contraint, bien plus, par exemple, que la dislocation à gauche ou la construction en *c'est que*. Ainsi, peu de verbes l'autorisent pour leurs compléments, et à condition que ceux-ci soient d'interprétation définie ou générique:

Les fraises, il aime beaucoup.
 ?* *Les fraises de ma ferme, il mange volontiers.*
 * *Ces cigarettes, il fume rarement.*

On aurait donc ici une situation particulière puisque la topicalisation serait possible alors que la dislocation à gauche et l'extraction en *c'est que* sont interdites:

- (1a) * *C'est accorder le piano qu'il aurait pu*
- (1b) ?* *Accorder le piano, il l'aurait pu*
- (2a) * *C'est accorder le piano qu'il aurait dû*
- (2b) ?? *Accorder le piano, il l'aurait dû*
- (3a) * *C'est (d') être malade qu'il a bien failli*
- (3b) * *Etre malade, il l'a bien failli*

¹⁵ Je suis redevable à Danièle Godard de cette suggestion.

¹⁶ A. Zribi-Hertz donne aussi des exemples avec des verbes à complétives (*vouloir, éviter*), mais il n'y a pas de raison de faire des infinitives après ces verbes des SV (voir infra).

- (4a) * *C'est (de) rédiger sa thèse, justement, qu'il est en train*
 (4b) * *Rédiger sa thèse, il en est en train*

De plus, la préposition n'est pas prise dans la topicalisation, alors qu'elle l'est en général:

De ces fraises, tout le monde raffole
 * *De rédiger sa thèse, il est en train*

Il est donc difficile d'introduire dans la grammaire du français un constituant SV sur la seule base des phrases (1) à (4) ci-dessus.

4) R. Jackendoff 1977 a poursuivi cet "éclatement" de la notion en proposant, dans le cadre de la théorie X-barre, un "groupe verbal" à trois niveaux: V''' (c'est-à-dire la phrase), V'' (le syntagme prédicatif de Chomsky 1965) et V' (le syntagme verbal de 1965), mais basée sur de nouveaux critères: distinction des adverbes entre V''' et V'' selon leurs propriétés d'extraction et de négation; distinction entre V'' et V' (c'est-à-dire entre adverbes et compléments) selon la reprise par *do so*. Ces deux tests sont transposables au français et nous allons les reprendre puisque nous contestons l'existence d'un niveau V' comme celle d'un niveau V''.

a) Pronominalisation en *le faire* (distinction V' - V'')

C'est l'équivalent de la reprise en *do so* proposée pour l'anglais pour isoler un groupe V' = V (sans auxiliaire ni marque flexionnelle) + compléments sous-catégorisés. Elle a été étudiée par J-P. Boons, A. Guillet, C. Leclère 1976.

Notons d'abord, comme le notait déjà D. Michelson 1969 à propos de *do so* en anglais, que ce critère n'est pas définitoire de V' puisque les adverbes aussi peuvent être repris par *le faire*. En fait, il faudrait poser un V' récursif puisque un même adverbe peut être repris ou non par *le faire* :

Jean écoute la radio le matin et Marc le fait aussi.
Jean écoute la radio le matin et Marc le fait l'après-midi.

Marc a dardé des regards perçants sur Sophie sans qu'elle s'en rende compte et Yves l'a fait aussi.

Marc a dardé des regards perçants sur Sophie sans qu'elle s'en rende compte mais quand Yves l'a fait elle s'en est aperçu.

La définition de V' devient donc plutôt floue.

On peut également penser que la reprise en *le faire* permettra de distinguer entre adverbiaux et compléments du verbe. Cela semble être vrai dans certains cas et le complément en *sur* de *N₀ compter sur N₁* est obligatoirement inclus dans *le faire* tandis que l'adverbial correspondant avec *N₀ écrire (à N₁)* ne l'est pas forcément:

? *Jean compte sur Marie et Paul le fait aussi.*
 * *Jean compte sur Marie et Paul le fait sur Jeanne.*
 ? *Jean écrit sur la table et Paul le fait sur le bureau.*

Mais ce test ne donne pas toujours les résultats escomptés. J-P. Boons, A. Guillet, C. Leclère 1976 ont tenté d'utiliser la reprise en *le faire*, pour plusieurs centaines de verbes français, comme test du "degré de cohésion" d'un complément avec le verbe, c'est-à-dire pour distinguer compléments "essentiels" du verbe et compléments "circonstanciels". Leur conclusion est sans appel : "Le critère en *le faire* est d'autant plus opératoire qu'on peut s'en passer, et d'autant plus flou qu'il serait utile."

Comme le notent ces auteurs, la reprise en *le faire* se limite à certains verbes (agentifs):

* *Marc craint Sophie et Paul le fait aussi.*

Le complément de *craindre* est pourtant obligatoire et clairement sous-catégorisé. De plus, pour certains verbes, *le faire* peut ne reprendre que le verbe et pas ses compléments:

Jean a opté pour des chaussettes rouges et Marie l'a fait pour des chaussettes vertes

Or il s'agit bien d'un groupe prépositionnel obligatoire: * *Jean a opté*

Alors que la reprise en *le faire* s'applique à des verbes pour lesquels le groupe prépositionnel est optionnel:

Paul a menti à Marie et Paul l'a fait aussi
 ? * *Jean a menti à Marie et Paul l'a fait à Jeanne.*

Il est difficile de dire qu'*opter pour des chaussettes vertes* n'est pas un V' alors que *mentir à Marie* en serait un.

Enfin, pour certains verbes à deux compléments, *le faire* peut reprendre le verbe et seulement un de ses compléments:

Jean a donné un gâteau à Marie et Paul l'a fait aussi
Jean a donné un gâteau à Marie comme Paul l'a fait à Anne

Il est difficile de dire qu'on a affaire à deux V' superposés.

La construction en *le faire* n'est donc pas un critère pour la mise en évidence d'un constituant V'. S'il s'agit d'établir une différence entre compléments et circonstanciels, d'autres critères existent, qui sont plus précis: forme de la question, possibilité d'antéposition, caractère obligatoire ou non. S'il s'agit de représenter les contraintes propres à la construction en *le faire* (ou à *do so*), ce constituant, que l'on doit définir comme ayant une composante récursive, n'aide en rien.

b) Propriétés des adverbes (distinction V'' - V''')

La distinction est ici entre (au moins) deux types d'adverbes, c'est-à-dire entre V''' = SN V'' Adv1 et V'' = Aux V' Adv2. Il s'agit de relier différentes propriétés syntaxiques et sémantiques des adverbes (extraction ou non de l'adverbe par *c'est ... que*, possibilité ou non pour l'adverbe d'être focus d'une question...) à une différence d'appartenance structurale. L'on distingue ainsi classiquement (Cf R. Martin 1974) les adverbes qui portent sur toute la phrase et ceux qui modifient plutôt le procès exprimé par le verbe :

Jean partira probablement

Probablement, Jean partira

* *C'est probablement que Jean partira*

* *Ce n'est pas probablement, c'est certainement que Jean partira*

Jean partira rapidement

* *Rapidement, Jean partira*

C'est rapidement que Jean partira

Ce n'est pas rapidement, c'est lentement que Jean partira

On est donc en présence d'un contraste assez net, qui permet de donner un fondement formel à l'intuition sémantique selon laquelle ces deux adverbes ont une "portée" différente, l'un qualifiant le procès, l'autre l'affirmation de la phrase elle-même.

Mais il n'y a aucune raison de préférer une bipartition adverbe de phrase/adverbe de syntagme verbal à une bipartition de type adverbe de phrase/ adverbe de verbe. Il est difficile de trouver des tests appuyant une différence entre adverbes de V et adverbes de SV, ou des cas d'ambiguïté correspondant à une telle alternative.

Syntaxiquement, on a donc besoin de deux types d'attachement, non de trois. S. Schlyter 1974 propose trois niveaux d'attachement : P, Spred et SV, distinguant adverbes de phrase et adverbes de manière, et parmi ces derniers ceux qui "impliquent un rapport avec le sujet" ou non, c'est-à-dire le type *soigneusement* (*Pierre a soigneusement chargé le camion*) et le type *lourdement* (*Pierre a lourdement chargé le camion*). Il est clair que trois niveaux seront insuffisants puisque les adverbes de phrase eux aussi peuvent "impliquer un rapport au sujet" ou non (voir infra).

Une bipartition P/V est de plus toujours préférable. En effet, si l'on veut prendre en compte la portée éventuelle de l'adverbe sur le sujet, mise en évidence par une paraphrase adjectivale, une sous-distinction est nécessaire aussi bien pour les adverbes modificateurs de phrase que pour les autres. Pour les adverbes de phrase, elle permet de distinguer *probablement* d'*intelligemment* :

Max a probablement invité Luc à dîner

* *Max a été probable.*

Intelligemment, Max a invité Luc à dîner.

Max a été intelligent d'inviter Luc à dîner (ou: *Max a eu l'intelligence d'inviter Luc à dîner*)

Pour les autres, elle permet de distinguer *mortellement* de *clairement*:

Max a frappé Marie mortellement

* *Max a été mortel.*

Max a exposé clairement le processus de fermentation du vin

Max a été clair (E + dans son exposé du processus de fermentation du vin).

Qu'une telle sous-distinction vaille dans les deux cas est rendu fort obscur si l'on croit que le second type d'adverbe est dans le domaine du SV, la raison d'être de SV étant de barrer l'accès au sujet. Il vaut donc mieux préférer un attachement à V ou à P, selon le résultat du test d'extraction en *c'est .. que*, et une structure de phrase sans SV, qui permet de représenter les sous-phénomènes de portée sémantique dans toute leur variété, tous les adverbes ayant accès à tous les arguments du verbe¹⁷.

L'adverbe peut en effet porter de façon privilégiée sur le sujet, comme *rageusement* ou *énergiquement*, qui demandent un agent humain dans la phrase, ou sur un des compléments du verbe, comme *complètement* ou *intégralement*, qui demandent un verbe transitif. Comme le note M. Gross 1990, ces phénomènes de portée peuvent être redoublés par des contraintes d'accord (entre le déterminant possessif à l'intérieur du groupe adverbial et l'argument de la phrase sur lequel il porte de façon privilégiée):

Max m'a rageusement frappé

*Max m'a frappé dans (sa + *ma + *leur ...) rage*

Jean a légué intégralement ses biens au Metropolitan

*Jean a légué ses biens dans (leur + *son + *mon ...) intégralité au Metropolitan.*

De telles variations montrent bien que la représentation des phénomènes de portée va au delà de celle de leur simple attachement structural¹⁸.

Une autre difficulté créée par un noeud SV pour l'analyse des adverbes est la mobilité des adverbes de phrase. Si l'on adopte une "convention de transportabilité" (pour reprendre l'expression de S. J. Keyser 1968) qui permet le libre placement de ces adverbes dans les limites des frontières de constituants dominés par le noeud Phrase, on exclut bien (1) et (2) tout en autorisant (3) et (4):

- (1) * *Le probablement garçon mangera une banane*
- (2) * *Le garçon mangera une probablement banane*
- (3) *Le garçon mangera une banane probablement*
- (4) *Le garçon probablement mangera une banane*

17 C. Molinier 1984 distingue ainsi, parmi les adverbes de manière en *-ment*, les adverbes de phrase, les adverbes sujets-phrases, les adverbes de verbe et les adverbes sujet-verbe. M. Gross 1990 parle plus généralement des phénomènes de double portée adverbiale.

18 M. Gross 1990 propose de baser leur description détaillée sur la mise en relation des phrases à adverbe avec une phrase à verbe support.

Mais, avec une structure de phrase à SV, on exclut aussi la phrase (5) qui est pourtant correcte:

(5) *Le garçon mangera probablement une banane*

Sauf à faire de SV un syntagme exception pour la convention de transportabilité, ce qui n'est guère satisfaisant.

5) Autres définitions : Les travaux plus récents en grammaire chomskyenne n'ont pas, à notre connaissance, développé de critères nouveaux pour la mise en évidence d'un constituant SV. En revanche, ils ont considérablement contribué à "remotiver" cette notion qui semble jouer un rôle crucial dans la définition de la plupart des principes sur lesquels la théorie s'appuie: théorie du gouvernement, principes de coindexation (ou "liage")... et qui a donc des conséquences sur la plupart des phénomènes syntaxiques, bien qu'elle ne soit pas pour autant plus claire¹⁹.

La notion de syntagme verbal y est souvent utilisée (quoique indirectement) pour rendre compte de différents cas d'asymétries, par exemple entre sujets et compléments, à l'aide d'une contrainte de type géométrique (c-commande, gouvernement...).

Cette situation témoigne simplement du fait que la théorie en question définit ses notions de base en termes largement syntagmatiques²⁰, et des théories différentes comme les grammaires d'unification ont en général développé sur plusieurs points des solutions alternatives qui ne font pas directement référence à la structure syntagmatique et donc à SV²¹.

Empiriquement, l'on constate que ces différents phénomènes d'asymétrie (tels qu'extraction ou relations anaphoriques) ne convergent pas vers une bipartition simple du type appartenance ou non appartenance à SV. Ainsi, s'il s'agit de contraintes d'extraction (îlots *wh*-), la différence est à faire entre sujets et adverbiaux (voire compléments nominaux) d'une part, et compléments (de verbe) d'autre part. S'il s'agit de contraintes sur la forme du complémenteur, c'est une distinction entre sujets d'une part (extraction en: *c'est...qui*), compléments et adverbiaux d'autre part (extraction en *c'est que*) qui est nécessaire. S'agissant des contraintes sur l'antécédent d'une anaphore pronominale, les distinctions pertinentes ne sont pas binaires et passent entre sujets, premiers compléments et deuxièmes compléments²².

C'est pourquoi des solutions alternatives, moins "géométriques", sont en général préférables. Ainsi, il y a eu progrès dans la détermination des contraintes sur les

19 La théorie de la prédication de E. Williams 1982 est un bon exemple de cette remotivation. Elle tend à "dédire" la notion de SV d'un principe sémantique plus général, laissant entier sa justification syntaxique (comme sa définition précise: on ne sait toujours pas quelle est la place exacte des adverbies).

20 Même si les structures syntagmatiques sont censées y être largement déduites des cadres lexicaux (principe de projection). La question de la sous-catégorisation des sujets reste d'ailleurs entière.

21 Si bien que l'on peut se demander à quoi sert SV comme sous-constituant de la phrase dans ces théories, particulièrement dans le cas de HPSG.

22 Au point que R. Larson 1988 propose pour l'anglais un SV pour le premier complément, chapeauté par un autre SV pour le deuxième complément: la structure syntagmatique qui devait justifier la contrainte se retrouve justifiée par celle-ci ! Ce n'est certainement pas, s'agissant de SV, le seul exemple de raisonnement circulaire.

antécédents des pronoms dans la mesure où l'on s'est éloigné de la structure syntagmatique pour considérer une structure d'abord plus abstraite (c-commande en forme logique) puis plus lexicale (f-commande sur la structure fonctionnelle en LFG, o-commande sur des cadres de sous-catégorisation définis comme listes ordonnées en HPSG). La situation est similaire pour les phénomènes d'extraction.

Les contraintes d'anaphore ont été peu décrites en TAG. Nous examinons en détail les phénomènes d'extraction, et leurs asymétries possibles au chapitre 5. Nous n'avons pas besoin de groupe SV pour en rendre compte.

2.2. Le statut des infinitives

L'utilisation d'une catégorie SV dans les cadres de sous-catégorisation des verbes, noms et adjectifs s'est popularisée en linguistique informatique. Des théories telles que GPSG et LFG considèrent en effet compléments ou sujets au gérondif (pour l'anglais) ou à l'infinitif comme étant de catégorie SV. C'est une conséquence directe de leur refus soit de tout élément non phonétiquement réalisé (donc des sujets vides) en structure syntagmatique (LFG), soit d'un élément vide aux propriétés particulières tel que le sujet PRO en grammaire chomskyenne (et du système d'assignation de CAS associé) (HPSG)²³. Ce refus ne semble pas empiriquement justifié pour le français. La discussion a été menée pour les compléments des verbes par M. Gross 1975 et nous l'étendons aux compléments de noms et d'adjectifs, en nous appuyant sur les recensements très complets effectués au LADL. Nous considérons ensuite les infinitives après conjonctions de subordination, avant de présenter dans ses grandes lignes la représentation que nous avons adoptée dans notre grammaire TAG.

2.2.1. Les infinitives sujets ou compléments de verbes

La plupart des compléments à l'infinitif commutent avec des compléments phrastiques et ont les mêmes propriétés²⁴. C'est la conclusion de M. Gross 1975 qui a étudié de façon systématique les verbes prenant des infinitives (notées *V-inf*, à *V-inf*, de *V-inf*) ou des complétives (*si P* ou *que P*) comme sujets ou comme compléments.

Aller à la pêche déplaît à Jean
Jean promet d'aller à l'école
Jean veut aller au Canada

Dans ces phrases, les groupes à l'infinitif commutent systématiquement avec des compléments phrastiques:

²³ Ces théories ont pu considérer les infinitives comme des SV et non comme des phrases à partir du moment où elles considéraient le passif comme "règle lexicale" et non comme transformation opérant sur une structure phrastique. Sinon, le lien entre infinitives actives et passives aurait fait défaut.

²⁴ H. Huot 1981, dont nous ne reprendrons pas ici les analyses, apporte d'autres arguments pour considérer les infinitives comme des phrases en leur assignant un complémenteur unique : *de*.

Que Marie soit partie déplaît à Jean
Jean promet qu'il ira à l'école
Jean veut que Marie aille au Canada

Il n'y a aucune raison d'attribuer une catégorie particulière aux infinitives et de supposer deux sous-catégorisations différentes pour ces verbes. En effet les verbes sont sémantiquement les mêmes dans les deux cas, et les infinitives ont les propriétés syntaxiques des complétives, non des groupes nominaux: pronominalisation en *cela* (et éventuellement en *le*), modification par des relatives en *ce* (*que* + *qui* + *dont*) :

Jean l'a promis, qu'il irait à l'école demain matin.
Jean l'a promis, d'aller à l'école demain matin.

Jean aime que l'on s'occupe de lui.
 = *Jean (* l'aime + aime cela)*
Jean aime aller chez Marie.
 = *Jean (* l'aime + aime cela)*

Si l'on supposait deux cadres syntaxiques différents pour chaque verbe, cette rencontre de propriétés serait totalement arbitraire. il faut donc considérer qu'on a affaire à un même complément phrastique dans les deux cas et analyser ces infinitives comme des phrases à sujet vide²⁵.

Pour les infinitives sujets, la commutation est régulière (il y a aussi commutation avec sujet SN). Pour les infinitives compléments, environ 200 verbes semblent faire exception, que M. Gross 1975 classe en verbes opérateurs (table 1), verbes de mouvement (table 2) et causatifs de mouvement (table 3). Pour les deux derniers groupes, qui sont les plus nombreux, l'infinitive alterne avec un complément locatif de type prépositionnel ou adverbial:

Jean court acheter du pain
Jean court à la boulangerie

Jean envoie Luc acheter du pain
Jean envoie Luc à la boulangerie

L'infinitive a les mêmes propriétés syntaxiques quant à la question et à la pronominalisation que le groupe prépositionnel de type locatif (ou directionnel) :

Jean y envoie Luc, acheter du pain
Jean y envoie Luc, à la boulangerie
Où Jean envoie-t-il Luc ? - Acheter du pain
Où Jean envoie-t-il Luc ? - A la boulangerie

²⁵ Et plus précisément comme des "réductions" de complétives selon M. Gross 1975 (effacement de: (*ce*) *que*, et du sujet).

Là non plus, il n'y a donc pas de raison de supposer deux sous-catégorisations différentes pour ces verbes, et l'on peut faire de l'infinitive un SP²⁶.

Pour les quelques dizaines de verbes restants (tables 1 et 17), plusieurs situations se présentent. Dans quelques cas, il y a commutation de l'infinitive avec un SN, comme avec *oser* ou *tenter de* :

Jean a (osé + tenté d') agresser Marie
Jean a (osé + tenté) cet acte horrible
Jean l'a (osé, + ? tenté, d') agresser Marie
Jean l'a (osé + tenté), cet acte horrible

Il n'y a donc pas lieu, là non plus, de faire de l'infinitive une catégorie syntaxique particulière, et l'on peut la considérer comme un SN.

Dans la majeure partie des autres cas, l'infinitive ne commute avec aucune autre catégorie. Elle n'apparaît pas comme ayant les propriétés d'un constituant: elle n'est ni déplaçable, ni sujette à question, ni pronominalisable²⁷:

Jean semble en vouloir à Marie = **Jean le semble*
La pluie commence à tomber = *La pluie (*le + *y) commence*

L'infinitive n'est pas directement sujette à question :

- * *Que semble-t-il ?* (vs *Que semble-t-il faire ?*)
- * *Que doit-il ?* (vs *Que doit-il faire ?*)
- * *(Que + De quoi) se dépêche-t-il ?* (vs *Que se dépêche-t-il de faire ?*)

Et l'extraction en *c'est que* ne s'applique pas:

- * *C'est aller à la pêche qu'il peut*
- * *C'est d'aller à la pêche qu'il se dépêche*
- * *C'est pleuvoir qu'il semble*

Deux solutions ont été proposées: l'analyse "à montée" du sujet (Cf N. Ruwet 1972, qui relie *Jean semble partir* à *Il semble que Jean part*, faisant de l'infinitive un groupe de type phrastique, relation qui s'observe sur très peu de verbes²⁸), et celle de M. Gross 1975

26 Comme le note M. Gross 1975, il y a en fait possibilité de cooccurrence des deux groupes:

Jean court à la boulangerie chercher du pain.
Jean envoie Marie à la boulangerie chercher du pain.

Dans ce cas, il faut analyser l'infinitive comme étant en apposition au complément locatif, ce qui est d'ailleurs cohérent avec le fait de la traiter comme un SP.

27 Les trois exceptions semblant être:

Il le peut
Que peut-il ?
? Il le doit.

28 Dans ce groupe, ce sont les verbes ayant un sujet phrastique (en général extraposé), tels *sembler*, *paraître* et *s'avérer*, qui motivent le mieux une analyse par "montée" du sujet de l'infinitive.

qui considère les verbes en question comme des auxiliaires (ou semi-auxiliaires). Cette deuxième solution ne suppose pas forcément que la sélection du sujet se fasse entièrement par le verbe à l'infinitif (et il semble que des verbes comme *se dépêcher de* ou *se hâter de* préfèrent des sujets humains, même s'ils sont par ailleurs semblables à *pouvoir* ou *devoir*). Dans les deux cas, l'infinitive n'est donc pas un SV complément du verbe mais est à rattacher directement au sujet (avec lequel elle forme la phrase simple de base).

2.2.2. Les infinitives compléments d'adjectifs

La situation est similaire pour les adjectifs. L. Picabia 1978 a observé une commutation assez générale entre complétives et infinitives compléments:

Marc est fier d'aller à Bilbao
Marc est fier (de ce) que l'on l'emmène à Bilbao.
Marc est prêt à aller à Bilbao
Marc est prêt à ce qu'on l'emmène à Bilbao.

L'adjectif a le même sens dans les deux cas, et les propriétés des complétives et des infinitives sont les mêmes (*De quoi Jean est-il fier ? A quoi Jean est-il prêt ?*) et l'on doit donc en faire deux variantes du même complément phrastique.

Il y a quelques exceptions. Mentionnons simplement un tour comme *Jean est difficile à croire*, où l'infinitive ne commute pas avec une complétive mais n'a pas non plus les caractéristiques d'un complément d'adjectif ni même d'un constituant autonome (d'où l'idée de la relier à: *Il est difficile de croire Jean* en grammaire transformationnelle).

Il reste quelques autres adjectifs (ou participes) comme *susceptible de*, *propre à*, *censé...* pour lesquels l'infinitive ne commute pas avec un complément phrastique. Les mêmes solutions que dans le cas des verbes s'appliquent. Avec *capable* (ou *impropre*), l'infinitive commute avec un SN et a les mêmes propriétés:

Jean est capable de progrès
Jean est capable de progresser
De quoi Jean est-il capable ?

Ce matériau est impropre à cette utilisation
Ce matériau est impropre à être utilisé de cette façon
A quoi est-il impropre ?

Avec d'autres adjectifs ou participes (*censé*, *susceptible*) l'infinitive n'a pas des propriétés de complément (ni de syntagme) et la combinaison *être + Adj* est mieux analysée comme un auxiliaire complexe (ou adjectif "à montée") :

Jean est censé s'amender
Jean est susceptible de s'amender
 * *Qu'est-il censé (vs Qu'est-il censé faire ?)*
 * *De quoi Jean est-il susceptible ? (vs Qu'est-il susceptible de faire ?)*
 ?* *Il en est susceptible*

Là non plus, il n'y a pas de raison de faire de l'infinitive un constituant.

2.2.3 Les infinitives compléments de nom

Les mêmes remarques s'appliquent aux compléments des noms. La commutation entre compléments infinitifs et phrastiques est assez générale (Cf C. Leclère 1971, J. Giry-Schneider 1978).

Jean a renoncé à son idée de venir
Jean a renoncé à l'idée que Marie vienne
Marc a donné l'ordre de déblayer la cour
Marc a donné l'ordre que la cour soit déblayée

Dans quelques cas, la commutation se fait avec un complément nominal:

Jean a une (certaine) tendance à mentir
Jean a une (certaine) tendance au mensonge
 ? *Jean y a une certaine tendance.*

Dans les cas où aucune commutation n'est possible, l'infinitive n'a pas des propriétés de complément (ni de syntagme). Ils se laissent analyser comme des cas où la combinaison verbe-nom (relativement figée) se comporte comme un auxiliaire:

Jean a tort de marcher pieds nus
Jean a coutume de marcher pieds nus
 * *De quoi Jean a-t-il tort ?*
 * *De quoi Jean a-t-il coutume ?*
 * *Jean (y + en) a tort*
 ?* *Jean en a coutume*

En général, ces combinaisons sont "neutres" pour le choix du sujet (Cf: *Il a coutume de pleuvoir*), sauf quelques cas comme *avoir tort*, *avoir raison*.

2.2.4. Les infinitives après conjonction de subordination

Si l'on considère enfin les prépositions ou les conjonctions (étudiées par M. Piot 1978), la commutation des infinitives avec des arguments phrastiques semble sans exception:

Jean est venu après (avoir été appelé + qu'on l'ait appelé)
Jean est venu pour (être embrassé + qu'on l'embrasse)
Jean est venu afin (d'être embrassé + afin qu'on l'embrasse)
Jean s'est usé à force (de travailler + qu'on le fasse travailler)

Les propriétés syntaxiques (question possible ou non, pronominalisation en *cela* ou non) sont les mêmes:

Après quoi Jean est-il venu?
Pour quoi Jean est-il venu ?
** Afin (de) quoi Jean est-il venu ?*
? A force (de) quoi Jean s'est-il usé ?*

Jean est venu après cela
Jean est venu pour cela
** Jean est venu afin (de) cela*
** Jean s'est usé à force (de) cela*

Si l'on supposait deux cadres syntaxiques différents pour chaque conjonction, cette rencontre de propriétés serait totalement arbitraire.

Qu'il s'agisse des arguments des verbes, des adjectifs, des noms ou des conjonctions, il y a donc globalement une bonne corrélation en français entre contextes des infinitives et contextes des complétives (et entre leurs propriétés). Il convient donc de considérer en général les infinitives comme des phrases à sujet vide. Il y a quelques exceptions, mais elles s'expliquent toujours mieux sans recours à une catégorie ad hoc comme SV.

2.2.5. Représentation en TAG

Dans notre grammaire TAG, nous avons donc considéré les infinitives comme des phrases. Nous n'avons pas de transformations explicites les analysant comme des réductions de complétives, mais nous avons regroupé dans une même "famille" élémentaire les arbres à compléments phrastiques introduits par *que* (ou *ce que*) et ceux à compléments infinitifs (ce qui revient au même ; Cf chapitre 3).

Pour les cas où l'infinitive commute avec un complément nominal ou prépositionnel (mais non avec un complément phrastique), par exemple après *oser* ou un verbe de mouvement, nous avons tout de même considéré les infinitives comme des phrases (non dominées par un noeud SN ou SP), afin de permettre l'extraction hors de ces groupes (extraction qui est en général interdite pour les compléments SN ou SP; voir chapitre 5 pour plus de détails sur les contraintes d'extraction):

Jean a osé aller au cinéma
Où Jean a-t-il osé aller ?
Jean a osé la suggestion d'aller au cinéma ?
** Où Jean a-t-il osé la suggestion d'aller ?*

Nous notons donc que ces verbes ont des complétives uniquement à l'infinitif. Il sera parfois nécessaire de préciser le blocage d'une propriété syntaxique avec de tels compléments, par exemple ici celui de l'extraction en *c'est que*: * *C'est aller au cinéma qu'il a osé*. Il ne s'agit pas là d'un cas exceptionnel et nous utilisons de telles contraintes pour l'ensemble de la grammaire (Cf chapitre 3, section 3).

Pour les cas restants, nous avons considéré les verbes en cause comme des auxiliaires qui s'adjoignent au verbe dans une structure phrastique initiale à l'infinitif (avec sujet réalisé; Cf chapitre 3):

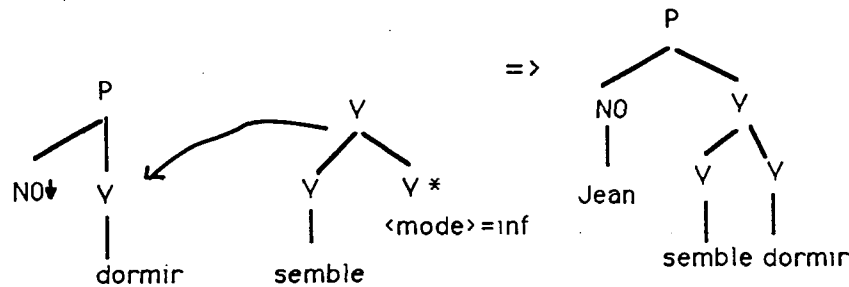


Figure 6

L'infinitif et ses compléments éventuels ne forment donc pas un constituant, et pronominalisation, question etc... sont systématiquement interdites. La sélection du sujet se fait en général entièrement par le verbe à l'infinitif (c'est-à-dire dans le cadre de la phrase simple). Il est à noter que le verbe à l'infinitif est obligatoire et que nous ne rendons pas compte pour l'instant des phénomènes d'ellipse éventuels (voir néanmoins chapitre 6 pour les ellipses en coordination).

Nous sommes également conduite, conformément aux distinctions faites notamment par N. Ruwet 1972 et au LADL, à distinguer deux verbes *promettre*: *N₀ promettre* (*N₁ + que P + de V-inf*) à *N₂* et *promettre de V-inf* auxiliaire, comme dans *La situation promet d'empirer* (complément en à *N₂* interdit); de même pour *risquer*, *menacer* et les aspectuels comme *finir*, *commencer*, *continuer*... qui sont systématiquement dédoublés.

Nous n'avons donc pas trouvé de justification syntaxique sérieuse à la notion de syntagme verbal, qui semble ne faire que perpétuer une vieille tradition logico-grammaticale distinguant en toute phrase un sujet et un prédicat. Au mieux, elle rencontre des phénomènes liés à l'ordre SVO observable en français (et en anglais). Nous l'avons donc supprimée pour notre description linguistique du français car cette catégorie superflue est en fait un obstacle à la compréhension d'un grand nombre de phénomènes syntaxiques.

3. Le syntagme nominal

Nous ne discutons pas l'existence de ce constituant qui est bien établie. En revanche, nous discutons de la pertinence d'une distinction catégorielle entre N et SN. Pour nous, il s'agit de la même catégorie N, dont seuls certains traits diffèrent.

Nous reprenons également les critiques de M. Gross 1975 et 1977 contre les représentations simplificatrices de SN en grammaire générative (critiques qui pourraient toujours s'appliquer aux représentations de SN dans les grammaires d'unification actuelles) :

- la notion de SN fait croire à une indépendance illusoire des niveaux syntaxiques, alors que de nombreux phénomènes indiquent que la composition de SN est étroitement déterminée par les autres éléments de la phrase et notamment le verbe (choix du déterminant, choix des compléments phrastiques éventuels, modifieur obligatoire ou non, alternance adverbe/adjectif, portée de la négation).
- la définition transformationnelle de SN est loin d'être claire car les différentes transformations ne s'appliquent pas forcément aux mêmes SN : devra-t-on définir autant de SN que de transformations ?
- il y a des SN à double extraction: un certain nombre de noms prédicatifs peuvent entrer dans une construction telle que leurs compléments ont toutes les propriétés de compléments de nom et en même temps toutes celles de compléments du verbe principal. C'est notamment le cas des phrases à verbe support. Comment analyser le SN dans ce cas ?

Nous allons ici reprendre les deux premières critiques, la troisième étant développée au chapitre 4 dans la section consacrée aux phrases à support. Nous préciserons d'abord certaines différences selon les SN acceptables dans différentes positions syntaxiques dans les phrases simples, puis d'autres différences entre les SN selon les transformations. Nous examinerons en particulier les contraintes pesant sur les SN que l'on peut extraire par *c'est que*. Nous montrerons à chaque fois le rôle essentiel du déterminant. Enfin, nous présenterons le traitement que nous avons adopté en TAG, qui consiste à considérer le déterminant comme adjoint et à redéfinir la distinction entre N et SN non comme une différence de catégories mais comme une différence de traits.

3.1. Une notion distributionnelle claire

J'ai vu (Jean + la plupart de ces gens + une jolie fille + quelqu'un que tu connais + la destruction soudaine de cette ville).

La commutation de la suite de catégories [(Prédéterminants) (Det) (Adj)* Nom (Adj)* (compléments) (relatives)*] avec un nom propre tend à assigner à celle-ci un statut de constituant, de type nominal.

Mais le fait de regrouper ces différentes combinaisons sous une même catégorie (SN)

<i>Pred</i>	exemples: <i>tous, environ...</i>
<i>Pred SN</i>	<i>tous les gâteaux</i>
* <i>Pred de SN</i>	* <i>tous de mes amis</i>
* <i>Pred N</i>	* <i>tous gâteaux</i>

Les SN introduits par un déterminant de type *Dadv* ou *Dadj* peuvent occuper toutes les fonctions syntaxiques²⁹. Pour ceux commençant par un déterminant de type *Dnom*, il est parfois difficile de les employer après une préposition:

Aux environs de 30 personnes ont vu Max

* *Max a été vu par aux environs de 30 personnes* (M. Gross 1977)

De même avec *autour de*, *en dessous de*, *de l'ordre de*, *dans les* ...

Mais ce sont les SN à prédéterminants (*Pred*) qui sont les plus contraints. Ceux introduits par *seul* ne peuvent être que sujets, ceux introduits par *jusqu'à* seulement sujets ou objets directs, et ceux introduits par *à moitié*, *entièrement*, *en entier*... seulement objets directs:

Seule sa femme a fait cela

* *J'ai vu seule sa femme*

* *Je pense à seule sa femme*

Jusqu'à ses amis ont battu Luc

Luc a vu jusqu'à dix personnes à la fois

* *Luc rêve de jusqu'à dix personnes*

* *Luc a été battu par jusqu'à ses amis*

Il s'agit là de contraintes (dont on peut se demander si elles sont syntaxiques ou sémantiques) entre des niveaux syntagmatiques différents, puisque le choix du déterminant est normalement du domaine du SN tandis que la fonction du SN est du domaine de la phrase. Dans le cadre d'une grammaire d'unification, des phénomènes de dépendance entre niveaux différents sont l'objet même de la grammaire, et il est relativement indifférent qu'ils soient de nature syntaxique ou sémantique.

En TAG, on peut a priori indiquer de telles contraintes soit dans les arbres SN eux-mêmes (c'est-à-dire dans les arbres des noms), soit au niveau des arbres phrastiques dans lesquels ceux-ci s'insèrent. Dans le premier cas, il faudrait également ajouter des traits fonctionnels distinguant les SN sujets, objets ou membres d'un groupe prépositionnel. Or, de tels traits sont contestables (puisque à une même fonction correspondent rarement des propriétés convergentes); de plus, le système résultant sera complexe puisqu'il s'agira d'implications du type: $\langle \text{Pred} \rangle = \text{seul} \Rightarrow \langle \text{sujet} \rangle = +$. Nous préférons indiquer ces contraintes dans les arbres phrastiques correspondants, quitte à avoir une certaine redondance (puisque par exemple tous les noeuds SN sujets dans tous les arbres phrastiques porteront ces mêmes traits). Il reste, comme pour les autres formalismes, à définir de tels traits avec précision. Le plus simple est de considérer des

²⁹ Plus précisément toutes les fonctions de base dans la phrase simple. Les restrictions sont nombreuses pour les autres positions: SN détaché, SN questionné ... (Cf chapitre 3)

traits lexicaux, sauf si des regroupements naturels apparaissent. Nous avons implémenté les restrictions sur les positions compléments prépositionnels, en excluant les SN à déterminants ou prédéterminants regroupés sous un trait <Det-Prep> (pour "déterminant commençant par une préposition").

3.2. Une notion transformationnelle moins claire

La situation est, en apparence, beaucoup plus favorable que dans le cas du syntagme verbal puisque la plupart des transformations (ou des métarègles des grammaires "sans transformations") font appel à SN. Il n'est cependant pas sûr qu'il s'agisse à chaque fois du même type de SN.

La passivation semble être celle qui présente à cet égard le moins de contraintes. A priori la plupart des SN sont passivables: SN à déterminant 0, SN définis, indéfinis, génériques, SN négatifs, SN *qu-* etc³⁰... Il s'agit d'une situation exceptionnelle. Les contraintes portant sur l'extraposition, l'inversion ou la dislocation ont été maintes fois mentionnées. Elles conduisent en général à des distinctions mettant en jeu des traits complexes et pas toujours faciles à manipuler: SN définis, SN indéfinis, SN quantifiés ... Ainsi, les extrapositions à partir de la position sujet portent sur des SN "indéfinis", la dislocation à gauche (avec reprise par le clitique *le, la, les* s'applique, elle, à des SN "définis":

(Un homme + Cet homme + Quelque chose) est arrivé
*Il est arrivé (un homme + quelque chose + *cet homme)*
*(*Un homme + *Quelque chose + Cet homme), il est arrivé hier.*

Nous avons plus particulièrement étudié les contraintes portant sur les SN susceptibles d'une extraction en *c'est ...que*, dans le cas des phrases simples d'abord, dans le cadre des phrases figées ensuite.

3.2.1. Extraction de SN en *c'est qu-* dans les phrases simples

Nous nous sommes appuyée sur la classification des déterminants de M. Gross 1977 qui est apparue comme particulièrement opératoire (Cf supra).

Nous avons testé d'abord l'extraction sur SN sujets et objets directs. Nous avons considéré les séquences: *Det N* mais aussi les combinaisons de déterminants du type: *Det1 de Det2 (de) N*, *Det1 de Det2 de Det3 (de) N*. Le résultat est que les cas interdits sont plus nombreux que les cas où l'extraction est possible. Il est également à noter que c'est toute la séquence des déterminants, et non le seul premier, qui entre en ligne de compte, ce qui rend l'expression de ces contraintes particulièrement délicate.

Nous allons examiner successivement l'extraction des SN introduits par les différents types de déterminants

30 Ce qui ne veut bien sûr pas dire que la passivation n'est pas autrement contrainte (voir chapitre 3).

a) SN introduits par *Dadv*

Pour tous les SN introduits par un déterminant de type *Dadv* l'extraction est impossible, que *Dadv* soit seul ou en combinaison avec d'autres déterminants³¹:

(*Dadv* de N)

Jean a mangé beaucoup de pain

Beaucoup de gens sont déjà partis

* *C'est beaucoup de pain qu'il a mangé*

* (*C'est + ce sont*) *beaucoup de gens qui sont déjà partis*

(*Dadv* de Det N)

Jean a mangé beaucoup de ce pain

Beaucoup de nos amis sont déjà partis

? * *C'est beaucoup de ce pain qu'il a mangé*

* (*C'est + ce sont*) *beaucoup de nos amis qui sont déjà partis*

(*Dadv Dadv* de N) et (*Dadv Dadv* de Det N)

Jean a mangé beaucoup trop de (E + ce) pain

Beaucoup trop de (E + ces) gens sont déjà partis

? * *C'est beaucoup trop de (E + ce) pain qu'il a mangé*

* (*C'est + ce sont*) *beaucoup trop de (E + ces) gens qui sont déjà partis*

b) SN introduits par *Dadj*

Pour les SN introduits par un déterminant de type *Dadj*, la situation est différente. Avec les numéraux et les déterminants formés sur *quel*, l'extraction est possible (même si un autre déterminant les suit ou les précède):

Ce sont trois filles que tu as vues

Ce sont trois de mes filles que tu as vues

Ce sont mes trois filles que tu as vues

C'est un groupe de trois filles que tu as vu

Avec les déterminants négatifs (*aucun*, *nul*), l'extraction est interdite, qu'ils soient ou non suivis d'un autre déterminant:

* (*Ce n'est + C'est*) *aucun touriste qui est venu*

* (*Ce n'est + C'est*) *aucun (de ces) touriste(s) que j'ai vu(s)*

Pour les *Dadj* restants, l'extraction est toujours possible lorsqu'on a la combinaison *Ddef Dadj N*:

³¹ En revanche, l'extraction de *de Det N* est parfois possible:

? *C'est de ce pain qu'il a beaucoup mangé*

Ce sont ces divers amis que tu as vus
Ce sont les quelques personnes que tu vois là qui sont venues

Dans le cas simple (*Dadj N*), les résultats sont souvent d'acceptabilité douteuse, mais il y a un contraste net entre *plusieurs*, *chaque*, *certain* et *quelques* d'une part (même si l'extraction est plus difficile avec les trois derniers), et *divers*, *force* ou *maint* d'autre part:

Ce sont plusieurs clients qui se sont plaints (et non quelques uns seulement).
? C'est chaque client que tu dois voir personnellement.
?? C'est (seulement) certains clients que tu as vus.
** (C'est + Ce sont) (diverses + maintes + force) glaces qu'il a mangées*

Il semble que les blocages soient les mêmes pour la combinaison *Dadj de Ddef N*:

Ce sont plusieurs de nos clients qui se sont plaints (et non quelques uns seulement)
?? Ce sont certains de nos clients qui se sont plaints
*? * Ce sont divers de mes amis qui sont venus*

Lorsque *Dadj N* (ou *Dadj de SN*) est à l'intérieur d'un groupe prépositionnel, l'extraction peut être plus facile:

C'est à divers de mes amis que tu as parlé
C'est à chaque client qu'il faut parler

c) SN introduits par *Dnom*

Pour les SN introduits par un déterminant de type *Dnom*, la situation est la suivante. L'extraction en *c'est .. qu-* n'est pas possible avec *la plupart*, *bon nombre*, *une foule*:

** C'est la plupart des gens qui sont partis*
? C'est la plupart des mes amis que tu as vus*

Elle n'est pas possible non plus avec les *Dnom* prépositionnels du type *dans les*, *aux alentours de*:

** (C'est + Ce sont) aux alentours de 100 personnes qui sont venues*

Pour les autres *Dnom*, l'extraction est souvent similaire pour la construction *Dnom de Det N* et la construction simple *Dnom de N*. Des contraintes supplémentaires peuvent également s'appliquer à l'accord en nombre:

*(C'est + *Ce sont) un groupe de trois personnes qui (est + *sont) venu*
*(C'est + *Ce sont) un groupe de (*personnes + médecins) qui (est + sont) venu(es)*

? *C'est plein de crème qu'il s'est mis*
 ? *C'est plein de cette crème qu'il s'est mis*

Comme dans le cas des SN introduits par *Dadj*, ces contraintes d'extraction ne s'appliquent pas toujours lorsque le SN fait partie d'un groupe prépositionnel:

? *C'est à une foule de gens qu'il a parlé*

En TAG, ces contraintes vont être exprimées au niveau des arbres phrastiques de type *c'est .. que* ou *c'est ... qui* qui sont des arbres élémentaires (voir chapitre suivant). Elles ne peuvent être exprimées directement au niveau des Det ou des SN car le contexte du SN (appartenance à un groupe prépositionnel ou non) est déterminant.

La situation est un peu plus favorable qu'en grammaire d'unification basée sur des règles de réécriture hors contexte. Il faudrait alors recopier les traits de Det sur SN et de SN sur SP, et distinguer selon qu'un SN fait partie d'un SP ou non. La finesse des contraintes en jeu montre bien qu'une métarègle comme celle que propose GPSG pour l'extraction: $S \rightarrow X^2 S/X^2$ est totalement insuffisante, puisqu'il faut distinguer selon le syntagme extrait (X^2) et selon le type d'extraction (question, *c'est que* ...)

Il reste, comme pour les autres formalismes, à définir les traits pertinents. Voir le système de traits retenus au chapitre 3 (section 1.5). Nous aboutissons en fait à des traits différents suivants les transformations: <def> pour la topicalisation (et reprise par *le*), <indef> pour l'extraposition à partir de la position sujet, <extrac> pour l'extraction en *c'est ... que*. Nous avons pris comme approximation que les SN négatifs ne pouvaient être clivés³².

3.2.2. Extraction de SN en *c'est qu-* dans les phrases figées

Il a été souvent noté que les phrases clivées supposent la possibilité d'une lecture contrastive (M. Gross 1977). L'extraction en *c'est que* devrait donc être difficile pour les expressions figées puisque le SN (ou le SP) figé ne doit pas, par définition, commuter avec un autre. En fait, comme le note N. Ruwet 1984, la situation est plus complexe.

Certains compléments figés contiennent un complément libre obligatoire de type *de N* (tables CDN et CPN de M. Gross 1989b). L'extraction est alors généralement possible puisqu'on peut toujours contraster avec un autre *de N* (ou un possessif):

Jean marche sur (les brisées de Jean + sur tes brisées)
Jean se paye (la tête de Luc + ta tête)
C'est la tête de Luc qu'il s'est payé (E + ce n'est pas la tienne)
C'est sur les brisées de Luc que Jean marche (E + ce n'est pas sur les tiennes)

³² Alors que *c'est que* est en général possible avec *ni ni*:
Ce n'est ni Paul ni Jeanne que tu as vu

Dans d'autres cas, le complément figé peut prendre un modifieur (adjectif ou relative). Là encore, le contraste avec un autre modifieur (et donc l'extraction) est généralement possible.

Nous avons particulièrement examiné l'extraction en *c'est que* des phrases à objet figé. M. Gross 1989b distingue 4 classes suivant la valeur du déterminant de N_1 :

- C1D (déterminant = Le) : *Jean demande la lune*
- C1G (déterminant possessif obligatoire) : *Jean a cassé sa pipe*
- C1I (déterminant = Un) : *Jean a pris une veste*
- C1R (autres déterminants) : *Jean a quitté ce monde*

Pour la classe C1I, l'objet figé peut toujours prendre des modifieurs et l'extraction en *c'est que* est toujours possible avec un modifieur.

Jean a pris une (E + sacrée) veste
*C'est une (*E + sacrée) veste que Jean a prise*

Elle est plus difficile avec N_1 sans modifieurs.

Pour les autres classes, la possibilité d'insertion d'un modifieur est généralement liée à l'alternance du déterminant du SN figé avec l'indéfini *Un*. Dans ce dernier cas l'extraction est généralement possible. On peut donc se reporter à la colonne *Autres Det* des tables de M. Gross.

Tables C1D et C1G; Autres Det = -

Jean a cassé sa pipe
Jean a mis le paquet
 * *C'est sa pipe que Jean a cassé*
 ?* *C'est le paquet que Jean a mis*

Tables C1D et C1G; Autres Det = +

Jean a fait (son entrée + une entrée remarquée) dans le monde
Jean tient (la forme + une forme éblouissante)
C'est une entrée remarquée que Jean a fait (E + ce n'est pas une entrée discrète qu'il a fait)
 ? *C'est une forme éblouissante que Jean tient (E + ? ce n'est pas une mauvaise forme qu'il tient)*

Il n'est donc pas étonnant que E. Laporte 1988 note une bonne corrélation entre déterminant figé et interdiction de l'extraction en *c'est .. que*, tout en remarquant que les exceptions sont nombreuses.

Une explication est donnée par I. Simatos 1986 qui, reprenant la distinction de J-C. Milner 1978, voit dans les SN à déterminants libres une capacité de référence "actuelle", qui est impossible pour les SN à déterminants figés qui n'ont eux qu'une référence

"virtuelle". Elle remarque également que la présence d'un modifieur peut permettre une pronominalisation autrement interdite, c'est-à-dire changer la capacité référentielle du SN:

*Marie donne le ton à toutes les soirées. (*Il + *Celui-ci) est désagréable.
Un ton nouveau a été donné à nos soirées. Il est désagréable.*

Mais l'insertion d'un modifieur ne suffit pas toujours:

*Jean a fait ((E + une grande) peur) + la peur de sa vie) à Marc
?* C'est ((E + une grande) peur) + la peur de sa vie) qu'il a fait à Marc*

Lorsque la modification est impossible, l'extraction est en général impossible. N. Ruwet indique certains cas particuliers et nous en avons noté quelques autres en C1D et C1G:

*C'est dans le ruisseau que Luc a trouvé Léa (Ruwet 1984)
? C'est la lune qu'elle demande
C'est sa carte maîtresse qu'il a abattue*

Selon N. Ruwet, le verbe a le même sens qu'en construction libre, et il y a donc commutation possible du SN figé:

*C'est (à la fac + dans le ruisseau) que Luc a trouvé Léa (Ruwet 1984)
C'est peu de chose qu'elle demande (ce n'est pas la lune)*

C'est un critère difficilement formalisable, et nous ignorons ces cas en TAG.

Enfin, quand seul le déterminant du SN est figé, le nom appartenant à une classe sémantiquement restreinte mais ouverte, le contraste, et *c'est que*, est en général possible:

*N₀ jouer DE LE Ninstrument := Paul joue du violon
C'est du violon qu'il joue (E + ce n'est pas du piano)*

*N₀ chasser LE Ngibier := On chasse la poule d'eau dans cette région
C'est la poule d'eau qu'on chasse dans cette région (E + ce n'est pas le colvert)*

Ce n'est donc pas ici la valeur du déterminant en soi qui contraint l'extraction, mais le caractère totalement figé ou non du couple Déterminant-Nom.

Pour notre représentation en TAG, nous avons retenu que l'extraction était interdite quand déterminant et nom étaient complètement figés (avec Det différent de *Un*). Pour les expressions à Ni figé, nous notons en fait la contrainte de modifieur obligatoire aux noeuds correspondants des arbres en *c'est que* ou *c'est qui* de leurs familles. Il suffit donc de se reporter à la propriété d'alternance du déterminant avec Un-modif pour savoir si l'extraction est possible, et ce n'est pas la peine de noter un trait *clivé-Ni* = + /- dans le

lexique.

Pour les expressions où seul le déterminant est figé, il s'agit d'une autre famille et leurs arbres en *c'est que* ou *c'est qui* ne porteront pas cette contrainte. Pour plus de détails sur notre représentation, voir chapitre 4 (sections 1.2.2. et 2.2.4.).

3.3. Représentation en TAG

Nous n'avons pas de catégorie SN à part entière. Nous utilisons N comme catégorie uniforme, et distinguons par des traits des noeuds N aux propriétés différentes.

3.3.1. Distinction N/SN

SN est donc une catégorie qui recouvre des syntagmes aux propriétés très différentes. Nous avons abandonné cette notion trop générale au profit de distinctions plus fines de propriétés selon la composition interne du syntagme. Le faisceau des propriétés résultantes nous apparaît comme plus important qu'une simple distinction catégorielle du type N/SN ou N/N'/N''/N'''... Une distinction du type de celle du trait <barre> en GPSG n'est pas non plus assez fine. En fait, comme le notent A. Kornai et G. Pullum 1990, la notation barre, chargée à l'origine de rendre explicite la tête des syntagmes (Cf Z. Harris 1951), a été peu à peu vidée de son sens si bien que des théories incompatibles peuvent s'en réclamer. Nous n'en avons pas besoin puisque les têtes sont dans notre formalisme directement apparentes dans les structures de base.

Nous utilisons l'opération d'adjonction pour introduire les déterminants car d'une part c'est la façon la plus naturelle de faire dépendre les propriétés du groupe résultant de celles du déterminant (puisque, comme nous l'avons vu sur l'exemple de l'extraction en *c'est que*, la nature du déterminant, et souvent la combinaison verbe-déterminant, détermine nombre de propriétés syntaxiques de Det-N). D'autre part, cela permet de faire apparaître un groupe *de N* à l'intérieur du groupe nominal dans le cas de déterminants complexes (sans avoir à prévoir ce morphème *de* par avance dans la structure du nom; Cf chapitre 3). Il en résulte, par la définition même de l'adjonction, que nom avec et sans déterminant auront une même catégorie. Leurs traits, en revanche pourront être différents.

Un nom comme *garçon* aura donc certains traits en commun avec un nom propre comme *Jean* (masculin, singulier, humain...), et en différera par d'autres qui seront en revanche communs à *Jean* et à un groupe comme *le garçon* (déterminé, défini..). Nous ne développons pas ici ce système de propriétés exposé plus en détail au chapitre 3. La propagation de tels traits doit être apparemment non monotone (puisque'un déterminant peut changer les propriétés du groupe auquel il s'ajoute, notamment dans le cas de combinaisons de plusieurs déterminants). Mais l'opération d'adjonction permet, en ajoutant un nouveau noeud à chaque fois, d'en donner une représentation formellement monotone, ce qui simplifie les calculs.

Le fait d'avoir une même catégorie pour N et SN, permet d'assigner aux noms communs un seul arbre élémentaire, qui pourra servir de N ou de SN suivant les contextes. Ainsi *La France* et *en France*, ou *Le président est sorti* et *On a nommé Luc Président* seront dérivés à partir des mêmes arbres élémentaires de *France* et *Président* respectivement.

A ces réserves près, nous avons une structure de SN assez classique. A cause de l'opération d'adjonction, qui introduit à chaque fois un niveau supplémentaire, nous distinguons un premier niveau qui domine directement la valeur lexicale du nom et où s'adjoignent les adjectifs, un deuxième avec les compléments et où s'adjoignent déterminants, modificateurs prépositionnels et relatives.

Ainsi, un SN comme *l'amour (E + passionné) de Jean pour Marie* sera représenté à l'aide des arbres élémentaires suivants:

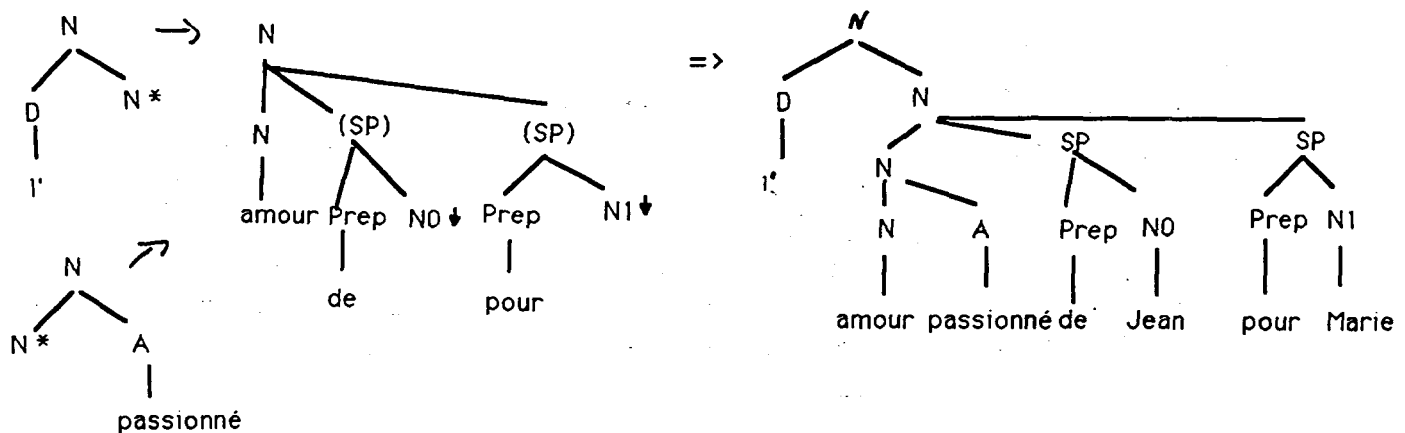


Figure 7

Il y a un parallélisme avec les structures phrastiques où les adverbes s'adjoignant en V jouent le même rôle que les adjectifs, les adverbes s'adjoignant à P (en tête ou en fin de phrase) jouant le même rôle que les déterminants et les relatives respectivement. Comme dans l'arbre phrastique, nous mettons sur le même plan, à la différence de J-C. Milner 1978 notamment, "sujets" et "compléments" sous-catégorisés du nom. Par exemple *Jean aime (E + passionnément) Marie* est représenté à l'aide des arbres élémentaires suivants:

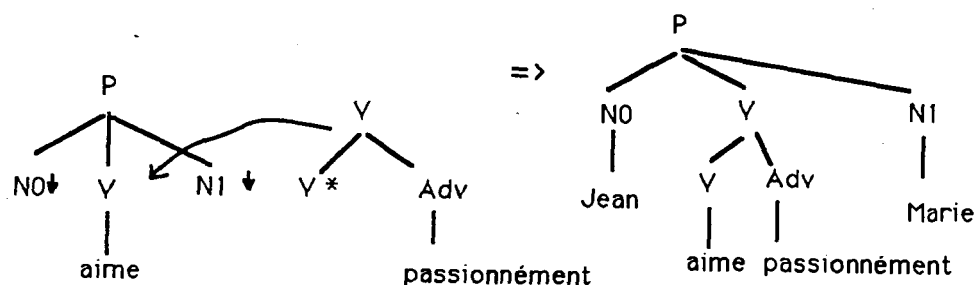


Figure 8

Dans les deux cas, prédicat et arguments forment l'arbre élémentaire (pour l'alternance entre déterminants possessifs et compléments en *de N*, voir chapitre 6).

3.3.2. Quelle tête du SN ?

Les grammaires d'unification comme LFG ou GPSG considèrent le nom comme la tête du syntagme nominal, c'est-à-dire comme ce qui détermine sa structure interne et la plupart de ses propriétés. Or, nous avons vu que le déterminant jouait un rôle au moins aussi important dans les propriétés syntaxiques du syntagme nominal.

Des discussions récentes en grammaire générative (S. Abney 1986, N. Fukui et M. Speas 1985) ont proposé de considérer le déterminant comme la tête de SN, qui devient par là même syntagme déterminatif ou SD. Mais leurs arguments sont surtout théoriques (généralisation de la notion de tête fonctionnelle à l'ensemble des catégories) et ont peu à voir avec les nôtres.

D'autres discussions en grammaire catégorielle ont pour objet de savoir lequel, du déterminant ou du nom, est le foncteur et lequel l'argument. Nous considérons que les déterminants s'adjoignent au nom, rejoignant en cela les grammaires qui le considèrent comme le foncteur, mais pour des raisons qui nous sont propres (voir chapitre 3).

La notion de tête n'a pour nous de sens que concernant les arbres élémentaires, et le déterminant est tête de son arbre comme le nom est tête du sien. Pour le SN final (arbre dérivé), parmi le faisceau de propriétés le caractérisant, certaines viennent du déterminant qu'il domine, certaines viennent d'un nom en dessous, certaines viennent du contexte phrastique qui le domine, sans qu'il soit utile d'établir une hiérarchie, ce qui est conforme à la propagation des traits par unification.

3.3.3. SN à double extraction

Il s'agit d'une confirmation que SN n'est pas, contrairement à ce qu'avaient affirmé des premiers travaux en grammaire générative, un îlot pour l'extraction. Nous en rendons compte seulement dans des contextes figés ou semi-figés (verbes support). Leurs propriétés se déduisent plus généralement des propriétés des syntagmes "étendus", c'est-à-dire des phénomènes de non compositionnalité (Cf chapitres 4 et 5).

4. Le syntagme prépositionnel

C'est un syntagme assez hétérogène dont la composition interne peut considérablement varier suivant la valeur de la préposition et suivant le contexte de SP.

4.1 Définition distributionnelle

Les tests de commutation font apparaître d'une part plusieurs paradigmes fort distincts, d'autre part que les "groupes prépositionnels" commutent souvent avec des groupes sans préposition réalisée.

Les groupes compléments (de verbes, de noms ou d'adjectifs) sont ceux où la situation est la plus favorable:

Jean est prêt à (tout + aller au cinéma + ce que Marie le quitte)

Même si la commutation peut se faire également avec une complétive :

Jean rêve (de Marie + d'aller au cirque + de ce que Marie parte + que Marie parte)

Pour régulariser la distribution, et rendre compte des propriétés particulières de cette complétive (pronominalisation en *en*, question en *de quoi*), il faut, dans une optique cette fois transformationnelle, l'analyser comme dérivée d'un groupe prépositionnel.

Pour les compléments locatifs, la situation est moins simple. Comme le font remarquer J-P. Boons, A. Guillet, C. Leclère 1976a, la préposition peut être optionnelle, ce qui fait hésiter entre complément nominal et complément prépositionnel:

Pierre habite (E + dans) cette maison

Pierre fouille (E + dans) ses poches

De plus, les groupes locatifs commutent avec des adverbes ou certains nominaux:

*Pierre est allé (*E + là-bas + près d'ici + à la maison + rue Racine)*

Pour les "circonstanciels", la commutation se fait aussi avec des adverbes et des conjonctives. Autre particularité, le SP peut être constitué de la suite *Prep Adv* et non seulement des suites *Prep SN* ou *Prep que P* :

Pierre est parti (hier + avant hier + avant le match + avant qu'il ne pleuve)

Pierre est parti (sans raison + pour son père + pour faire plaisir à Marie + parce qu'il en avait assez)

L'on peut analyser les conjonctives comme des groupes prépositionnels (certaines "prépositions" étant suivies uniquement d'une complétive).

Par ailleurs, il y a peu d'ajouts spécifiques à SP. La plupart sont également des ajouts à SN ou aux adverbes comme *bien*, *juste* ou *précisément*:

C'est (E + bien) (hier + (E + juste) avant le match) que cela s'est passé
C'est (E + précisément) de Marie qu'il rêve

Enfin, l'on a presque toujours besoin d'accéder à l'intérieur des SP compléments. La sous-catégorisation se fait par le verbe et non par la préposition. Ainsi avec un verbe comme N_0 *rêver de* N_1 , *de* peut être suivi d'un SN ou d'une complétive, mais avec un verbe comme N_0 *hériter de* N_1 , *de* est seulement suivi d'un SN. De même pour les restrictions de sélections, *de* est plutôt suivi d'un SN concret lorsqu'il est complément de N_0 *sortir de* N_1 , et plutôt d'un SN humain lorsqu'il est complément de N_0 *apprendre* N_1 de N_2 .

Il peut donc être tentant d'opérer une "mise à plat" de SP et de remplacer cette notion hétérogène par des suites *Prep SN* ou *Prep (ce que) P* apparemment plus pertinentes (M. Salkoff 1973). Mais les autres propriétés syntaxiques (et ces seules distributions) nous en empêchent, si l'on veut garder le principe que l'on ne déplace que des syntagmes.

4.2. Définition transformationnelle

C'est la définition la plus claire, car tous les groupes définis ci-dessus ont en commun de pouvoir être extraits ou déplaçables.

- (1) *C'est de Marie que Jean rêve*
- (2) *C'est à Marie que Jean pense*
- (3) *C'est près d'ici que Jean est allé*
- (4) *C'est avant le match que Jean est parti*

Mais, dans une langue comme le français, le statut d'ilot de SP est moins clair (Cf chapitre 5). L'on ne peut extraire par question les éléments qu'il contient (à la différence de *preposition stranding* en anglais par exemple)³³. En revanche, pour les SP compléments, une extraction du SN dominé par SP est souvent possible, si SN est repris par un pronom *qu-* ou *dont*. A côté des phrases (1) et (2), on a en effet:

- (1a) *C'est Marie dont il rêve*
- (2a) *C'est Marie à qui il pense*

Certains SP, étudiés par L. Danlos 1980 ont des propriétés particulières: construits comme attributs du verbe *être*, ils sont difficiles à extraire par *c'est que*, mais sont pronominalisables par *le* et dislocables:

Jean est de bonne humeur

³³ Comme le note A. Zribi-Hertz 1983, il y a certains phénomènes de "prépositions orphelines" en français (*La peine de mort, je suis contre*) mais ils ne s'analysent pas en termes d'extraction.

?* *C'est de bonne humeur qu'il est*
De bonne humeur, il l'est (assurément)

Il convient donc de les considérer comme des groupes, mais ils ont des propriétés de groupes adjectivaux autant que de groupes prépositionnels.

4.3. Représentation en TAG

On distingue plusieurs types de SP. Une première distinction passe entre SP arguments (de verbe, de nom ou d'adjectif) et SP adverbiaux. Les premiers n'ont pas d'existence indépendante, ce sont des noeuds intérieurs des arbres de leurs prédicats: la préposition, simple "marqueur d'argument" forme avec son prédicat la tête multiple d'un arbre phrastique, nominal ou adjectival. Les seconds sont des arbres auxiliaires (de la catégorie de ce qu'il modifie) dont la tête lexicale est la seule préposition (+ *que* dans le cas d'une conjonctive).

Dans les deux cas, le syntagme prépositionnel est "étendu", c'est-à-dire que SP est toujours un noeud intérieur, jamais un noeud feuille: sa composition est donc entièrement déterminée par la structure globale de l'arbre et y est apparente. Il est donc aisé de faire des distinctions entre, par exemple, la composition des modificateurs de nom introduits par *de* et celle des modificateurs de verbe introduits par la même préposition car il s'agit de deux arbres auxiliaires différents (la possibilité d'avoir un déterminant zéro après *de* sera différente). De même pour les compléments de verbe introduits par *en* et les modificateurs introduits par la même préposition.

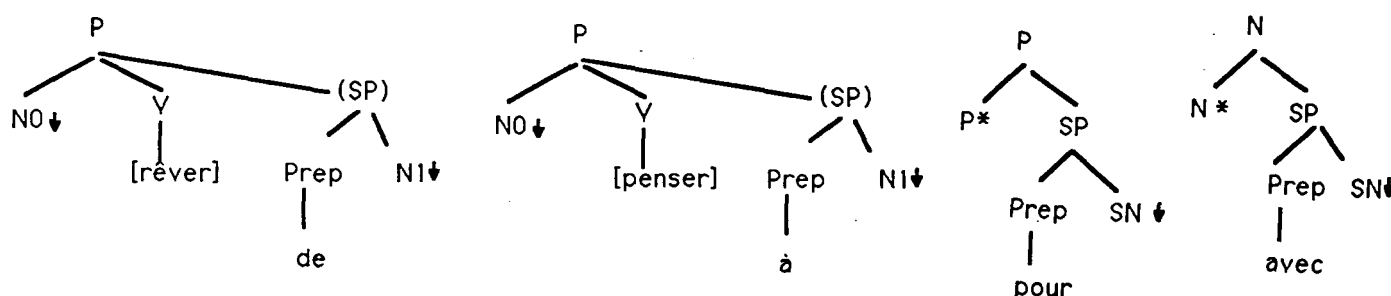


Figure 9

Nombre d'ambiguïtés concernant l'attachement des groupes prépositionnels peuvent être ainsi levées (Cf chapitre 7). Il est également aisé de rendre compte des doubles extractions mentionnées ci-dessus (Cf chapitre 5).

Un cas particulier est celui des compléments locatifs, obligatoires par exemple avec *N₀ aller Loc N₁* ou *N₀ mettre N₁ Loc N₂*. Ils ne peuvent être étendus puisqu'ils alternent avec des adverbes (*ici, là*). Nous les représentons donc comme des arbres initiaux (de racine SP) qui sont substitués aux noeuds feuilles correspondant des arbres de leur

prédicat³⁴. Les prédicats à arguments locatifs sont donc les seules structures où le SP n'est pas "étendu".

La notion de SP "étendu" a aussi des conséquences sémantiques en français: c'est l'unité Verbe (ou Nom ou Adjectif tête) - Prep qui est pertinente du point de vue sémantique (Cf chapitre 3).

5. Autres syntagmes

Comme pour le syntagme nominal, nous ne nous appuyons pas sur une différence de niveaux de barre et nous ne faisons pas du groupe adjectival ou du groupe adverbial des catégories distinctes de celles de l'adjectif ou de l'adverbe. Une telle différence ne peut distinguer pour nous le groupe [Adjectif modifieurs] ou le groupe [Adverbe modifieurs], puisque l'adjonction d'un modifieur ne change pas, par définition de l'adjonction, la catégorie de ce à quoi il s'ajoute. Elle n'est pas non plus apparue utile pour distinguer l'adjectif avec et sans compléments, ou l'adverbe avec et sans compléments.

5.1. Le Syntagme adjectival

Nous ne faisons pas du syntagme adjectival une catégorie particulière. Les adjectifs ont toujours la catégorie A avec et sans compléments, avec et sans modifieurs.

Ainsi un adjectif sans complément comme *rouge* a la même catégorie maximale qu'un adjectif à complément obligatoire comme *propre à*. Nous n'avons pas de traits distingués entre le noeud A préterminal et celui qui surplombe l'adjectif et ses compléments (car ils se comportent de la même façon eu égard aux modifieurs)³⁵.

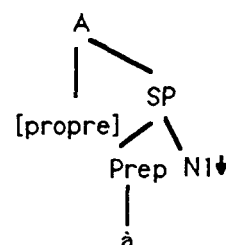
En revanche des traits peuvent être différents entre le noeud A supérieur et un noeud A inférieur en cas d'adjonction de certains modifieurs. En effet, les modifieurs d'adjectifs ne sont pas librement combinables entre eux:

(très + assez + plutôt) rouge
 très (très + *assez + *plutôt) rouge
 assez (*très + *assez + *plutôt) rouge
 plutôt (très + ?assez + *plutôt) rouge

Il faut donc un trait distinctif entre l'adjectif *rouge* (qui peut recevoir n'importe lequel de ces adverbes), et le groupe *très rouge* par exemple, qui ne peut prendre que *très* et *plutôt*,

34 Les adverbes locatifs ont la même représentation.

35 En fait, si nous ne nous limitons, par commodité, à avoir des préterminaux non branchants, nous devrions avoir une structure à un seul noeud A dans tous les cas:



ou le groupe *plutôt rouge* qui ne peut en prendre aucun³⁶.

Il pourra donc y avoir des différences de traits entre adjectifs avec et sans modificateurs, et c'est là une justification du niveau syntagmatique supplémentaire apporté par l'opération d'adjonction.

5.2. Le syntagme adverbial

Nous n'avons pas non plus de catégorie particulière pour le syntagme adverbial. Les adverbes ont toujours la même catégorie (Adv), avec et sans compléments, avec et sans modificateurs, même s'il y a des différences de niveaux syntagmatiques (les noeuds Adv peuvent être enchâssés comme les noeuds A). On peut faire la même remarque que pour les adjectifs sur les éventuelles différences de traits entre ces niveaux: il n'y en a pas entre Adverbe avec et sans complément, mais il peut y en avoir dues à des adjonctions de modificateurs, puisque ceux-ci ne sont pas toujours compatibles entre eux:

(très + assez + plutôt) rapidement
*très (très + *assez + *plutôt) rapidement*
*assez (*très + *assez + *plutôt) rapidement*
*plutôt (très + ?assez + *plutôt) rapidement*

En conclusion, nous avons donc reconsidéré la plupart des syntagmes traditionnels d'un point de vue purement linguistique. Nous avons redéfini les notions de syntagme nominal ou syntagme prépositionnel et éliminé des notions qui nous ont semblé inutiles comme celle de syntagme verbal, de syntagme adjectival ou de syntagme adverbial.

Pour le syntagme nominal, le syntagme adjectival et le syntagme adverbial, la définition de noeuds distincts (mais de même catégorie) auxquels peuvent être associés des faisceaux de traits différents nous a semblé plus importante que les distinctions habituelles entre sous-catégories (N et SN (ou N'), A et SA, Adv et SAdv).

Dans tous les cas nous tirons parti de la notion de syntagme étendu, propre à notre formalisme, pour faire apparaître dans la même structure élémentaire une tête lexicale et ses différents arguments. Pour les syntagmes prépositionnels, nous faisons apparaître de la même façon la structure interne du syntagme.

Il est à noter que ces remises en cause sont relativement indépendantes de notre formalisme (ou plutôt dépendent avant tout des contraintes linguistiques que nous y avons ajoutées: la lexicalisation).

³⁶ On pourrait se demander si ces restrictions ne sont pas uniquement le fait des adverbes. Ce n'est pas le cas car elles mettent en jeu l'adjectif comme le montre le contraste suivant:

*(E + *assez) bien rouge*

(E + assez) bien fourni

(à moins de considérer qu'il s'agit de deux adverbes *bien* différents)

Chapitre III

Les structures élémentaires de la grammaire d'arbres adjoints du français

Dans le cadre d'un lexique-grammaire, la seule unité est la phrase simple, c'est-à-dire le verbe et ses arguments essentiels (M. Gross 1981). Les phrases complexes sont obtenues par transformations. Ainsi, une phrase avec un adverbe, par exemple *Jean a vu Marie hier*, est analysée comme la composition des deux phrases simples: *Jean a vu Marie* + *Cela a eu lieu hier*, qui donne le résultat souhaité après effacement des pronoms anaphoriques et du verbe support *avoir lieu*¹. Les adjectifs et les noms doivent en toute logique être intégrés de la même façon et une phrase comme : *Un garçon adroit est venu* devrait être obtenue à partir des trois phrases simples (en omettant le problème de l'insertion du déterminant): *Quelqu'un est venu* + *Ce quelqu'un est un garçon* + *Ce quelqu'un est adroit*.

Dans notre système, nous n'avons pas de mécanisme transformationnel de ce type. Nous avons donc besoin de considérer d'autres catégories que la phrase comme unités. Nous avons en effet des arbres dont les racines correspondent à différents types de syntagmes. Mais la plupart des items lexicaux ont aussi une structure élémentaire de racine Phrase. Ainsi nous opérons une mise à plat de certaines de ces transformations, puisque les items lexicaux se voient associer plusieurs structures : celle correspondant à la phrase simple, et celle(s) correspondant aux formes réduites.

Les structures qui, dans un cadre transformationnel, seraient reliées par transformations, sont regroupées dans des classes d'équivalence appelées "familles d'arbres". Ces familles d'arbres élémentaires sont les catégories des verbes dans le dictionnaire TAG.

1. Les différents types d'arbres élémentaires

Les arbres initiaux sont utilisés pour les noms, les verbes à arguments nominaux ou prépositionnels. Les arbres auxiliaires sont utilisés pour les modificateurs (relatives, adjectifs, adverbes) et plus généralement les "ajouts" au sens harrissien du terme (verbe auxiliaires ou opérateurs, verbes à complétives, déterminants...). De plus, un arbre élémentaire doit correspondre à une structure argumentale complète. A chaque terme prédicatif est donc associée une structure arborescente où tous les noeuds correspondant à ses arguments sont présents (sous forme de noeuds à substituer ou de noeud pied).

La différence entre arbres substitués et arbres adjoints correspond en général à la distinction linguistique entre arguments et ajouts:

- la substitution est obligatoire, et s'applique une seule fois, aux noeuds feuilles marqués comme tels.
- l'adjonction est en général facultative et réitérable aux noeuds intérieurs ou racines.

¹ Ces transformations ne sont en fait pas orientées.

Contrairement aux autres grammaires d'unification, nous encodons directement dans la structure du prédicat tous ses arguments. Nous n'utilisons pas de traits <sous-cat> comme GPSG, ou de niveau spécifique de représentation comme LFG (structure fonctionnelle).

Nous allons examiner les différents types d'arbres élémentaires correspondant à des phrases simples, en les distinguant selon la catégorie de leur tête lexicale. Nous présenterons successivement les structures des verbes, des clitiques, des prépositions, des déterminants, des noms, des adjectifs et des adverbes.

1.1 Les verbes

L'on distingue les auxiliaires (ou verbes à montée du sujet) des autres verbes. Les premiers sont traités comme des ajouts au verbe principal, les seconds sélectionnent des arbres de racine phrastique dont les feuilles portent leurs différents arguments. Ils sélectionnent des arbres initiaux s'ils prennent des arguments nominaux ou prépositionnels, et des arbres auxiliaires s'ils prennent des arguments phrastiques (cette distinction est fondée sur les possibilités d'extraction hors des compléments: voir chapitre 5)².

Voici par exemple des arbres élémentaires associés à N_0 *dormir*, N_0 *aimer* N_1 , N_0 *donner* N_1 à N_2 , N_0 *penser* *que* P ³:

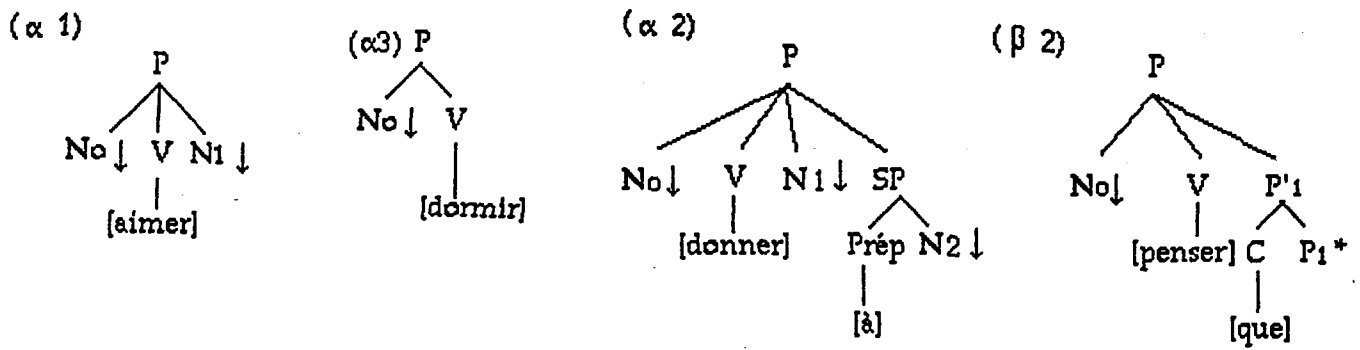


Figure 1

Nous nous sommes appuyée sur les classifications du LADL pour distinguer les verbes selon leurs emplois. A une forme verbale ambiguë comme *voler* correspondent deux structures (en fait deux entrées lexicales et deux "familles" d'arbres) différentes:

² Le verbe principal (tête de l'arbre phrastique) est donc bien pour nous celui qui sous-catégorise les arguments de la phrase, et non forcément le verbe conjugué (a contrario voir M. Salkoff 1973).

³ $\downarrow$ marque les noeuds à substituer, et $*$ les noeuds pieds. $[xx]$ représente n'importe quelle forme fléchie (ou élidée) du lemme correspondant. (X) marque que le noeud X (et le sous-arbre qu'il domine) est optionnel. / marque la disjonction. Ces arbres sont simplifiés puisque les traits associés à chaque noeud n'y ont pas été portés (voir section suivante).

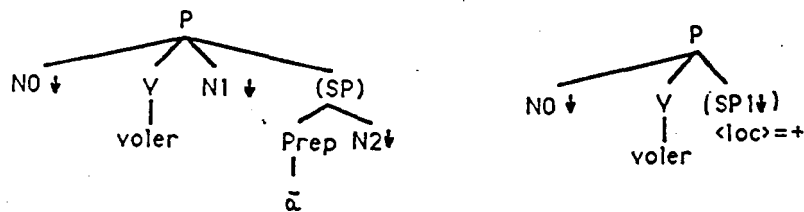


Figure 2

La catégorie des arguments est donc définitoire de la structure d'une entrée verbale⁴. Ce n'est pas par unification (ni par des principes de bonne formation des structures de traits comme en LFG) que sont représentés en TAG les phénomènes de sous-catégorisation ou de sélection des prépositions. En revanche, la plupart des autres contraintes imposées par les verbes à leurs arguments sont représentées sous formes de traits.

Nous allons présenter les structures des auxiliaires puis celles associées aux verbes ordinaires. Pour ces derniers, nous traiterons d'abord les contraintes portant sur les sujets puis celles portant sur les compléments (nominaux, prépositionnels ou phrastiques).

1.1.1. Auxiliaires, modaux et verbes "à montée du sujet"

Ils sont tous traités comme des ajouts au verbe, c'est-à-dire par des arbres auxiliaires de racine V.

a) Modaux et aspectuels: *venir de*, *pouvoir*...

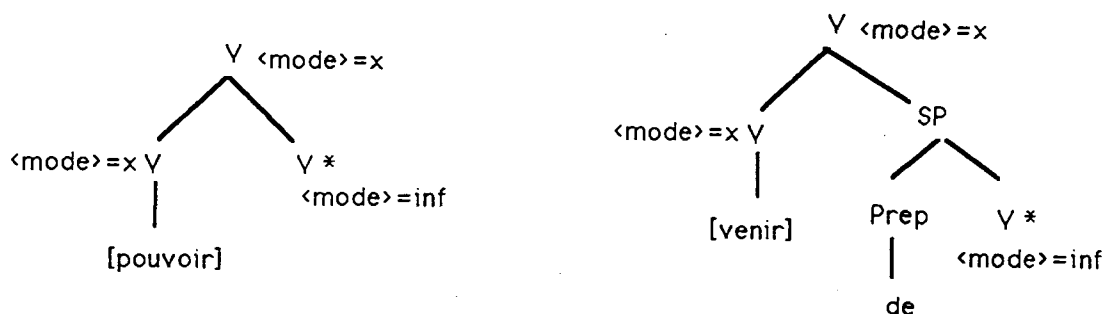


Figure 3

Il peut y avoir adjonctions en chaîne:

Jean semble devoir pouvoir partir.

⁴ Les critères utilisés pour distinguer entre compléments circonstanciels (modifieurs) et arguments du verbe sont expliqués dans J-P. Boons, A. Guillet, C. Leclère 1976a. Ils sont classiques: caractère obligatoire ou non, forme de la question, antéposition possible (sans inversion du sujet).

verbe dans l'entrée correspondante du dictionnaire syntaxique. Par exemple, N_0 remonter à N_1 :

L'accident remonte à 10 ans

* *L'accident (a + est) remonté à 10 ans*

Quand un verbe n'existe qu'au participe passé (par exemple, le verbe à second complément figé : N_0 perclu N_1 de rhumatismes, c'est une contrainte générale sur le mode des phrases bien formées qui forcera l'adjonction de l'auxiliaire :

Le travail a perclu Max de rhumatismes⁵

Max est perclu de rhumatismes

L'auxiliaire *être* du passif fait partie de la structure du verbe passif. Il est considéré comme sémantiquement vide et ne forme donc pas une structure autonome. (Cf infra p. 155).

1.1.2 Contraintes sur les sujets

Nous considérons que la sélection des sujets et celle des compléments se fait de la même façon (Cf chapitre 2), c'est-à-dire qu'ils apparaissent de la même façon dans la structure élémentaire du verbe. En revanche, leurs contraintes ne sont pas toujours exactement les mêmes.

a) Verbes à sujets phrastiques

La plupart des verbes à sujets phrastiques acceptent également des noms abstraits comme sujets. Nous incluons donc des structures à sujets nominaux dans les familles regroupant les arbres des verbes à sujet phrastique.

Les sujets phrastiques sont structurellement restreints quant à leur mode: subjonctif ou infinitif, et nous marquons cette indication directement dans la structure arborescente. En général, les verbes ne seront donc pas lexicalement marqués pour le mode de leur sujet phrastique.

De façon limitée, des contraintes lexicales sur le mode du verbe du sujet phrastique analogues à celles qui contraignent le mode du verbe des compléments phrastiques peuvent apparaître:

*(Que Marie soit amoureuse + *Aimer Bob) a (percé + transpiré) dans la presse.*

(?? Que Marie aille à Paris + Aller à Paris) dit à Jean.

Ces verbes porteront donc cette contrainte supplémentaire dans leur entrée. Cette contrainte, en s'unifiant avec celle déjà présente dans l'arbre, remplacera celle-ci.

⁵ Cette phrase n'est pas acceptée par tous les locuteurs.

b) Verbes à sujet obligatoirement pluriel

Cette contrainte est étudiée dans J-P. Boons, A. Guillet, C. Leclère 1976b, dont nous reprenons les exemples:

*(*Cette abeille + les abeilles) grouillent dans le jardin*

En fait, ces verbes acceptent aussi des sujets collectifs ou génériques:

La foule grouille dans les rues

L'or pullule dans cette région

Nous notons donc la contrainte sur N_0 comme une disjonction de ces trois traits⁶.

Pour les verbes à sujets figés, comme pour les verbes impersonnels, voir chapitre 4.

c) Contraintes sémantiques

Nous utilisons un système de traits sémantiques assez rudimentaire pour l'instant: <hum> pour les humains ou les animés, <conc> pour les concrets et <abs> pour les abstraits.

Les verbes à sujet phrastique ont toujours le trait $N_0:<abs> = +$ dans l'arbre à sujet nominal qui leur correspond. Pour les autres, c'est le verbe qui précise les éventuelles restrictions sur le sujet : *N0 admirer N1* ($N_0:<hum> = +$), *N0 se rapporter à N1* ($N_0:<abs> = +$).

Les noms eux-mêmes sont souvent non marqués pour ces traits, car ils sont susceptibles de multiples interprétations métaphoriques. Le verbe imposera éventuellement le trait qui convient dans la phrase. Les noms de pays par exemple pourront se comporter comme des noms humains ou comme des concrets:

La France admire M. Gorbatchev

Ils ont découpé le Koweït en deux

Ils seront donc marqués <abs> = -.

M. Salkoff 1979 note qu'il est préférable d'interdire les sujets (ou les objets) inacceptables (ce qu'il appelle "méthode d'exclusion") plutôt que de tester si un sujet (ou un objet) donné est acceptable ("méthode de permission") : il est plus aisé de définir les classes de sujets (ou d'objets) inacceptables que les classes de sujets (ou d'objets) acceptables, et la non exhaustivité des descriptions est moins dommageable dans le

⁶ Nous ne tenons pas compte des restrictions supplémentaires sur les sujets coordonnés:

((Jean et Marie) + (Les abeilles et les mouches) + (le mouton et le lièvre)) pullulent dans cette région.*

Chaque membre de la coordination doit pouvoir en fait être admissible en N_0 .

premier cas. Si l'on formule ces contraintes par unification de traits (et non par appartenance à des sous-catégories prédéfinies), cette différence méthodologique perd beaucoup de son sens. Nous exprimons des contraintes aussi bien positives que négatives. Par exemple, $N0: \langle \text{hum} \rangle = +$ ou $N0: \langle \text{conc} \rangle = -$ (ce qui inclut les humains, les abstraits et tous les noms qui ne sont pas explicitement marqués $\langle \text{conc} \rangle = +$).

1.1.3. Accord sujet-verbe

Nous considérons qu'il s'agit d'une contrainte existant indépendamment de la valeur des items lexicaux en jeu⁷. Nous représentons donc l'équation forçant l'unification des traits correspondants dans les arbres élémentaires et non au niveau de leurs têtes lexicales:

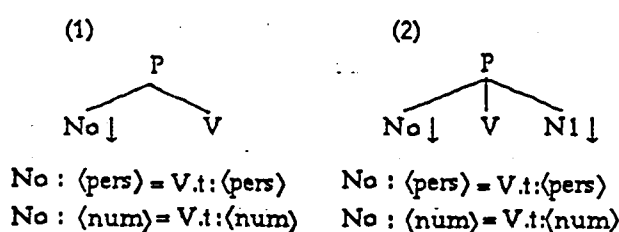


Figure 5

L'unification se fait en (1) et en (2) entre des traits de type top. En effet, les noeuds à substituer héritent leur valeur du top du syntagme nominal qui leur est substitué (par définition de l'unification en cas de substitution). Ainsi, les traits de genre et de nombre ne dépendront pas uniquement du N tête (pour qui ce sont des traits bottom) mais aussi de son déterminant et de ses modifieurs éventuels (voir section suivante).

Pour le verbe, il s'agit également d'un trait top car il faut tenir compte des éventuelles adjonctions d'auxiliaire⁸. S'il s'agissait de traits bottom, le verbe tête à l'infinitif pourrait s'unifier avec n'importe quel sujet puisque ne portant pas de trait nombre ni genre, et rien ne contraindrait la forme des auxiliaires éventuels. En tant que trait top, en revanche, l'accord se fera toujours avec le verbe le plus dominant dans l'arbre, qui est celui qui est toujours pertinent en français pour l'accord en personne nombre:

⁷ Ceci est seulement une approximation puisque certains verbes (généralement à sujet phrastique) et certaines conjonctions autorisent des cas où le sujet ne s'accorde pas avec le verbe:

Jean et ses bavardages continuel m' (a + ont) toujours déplu.

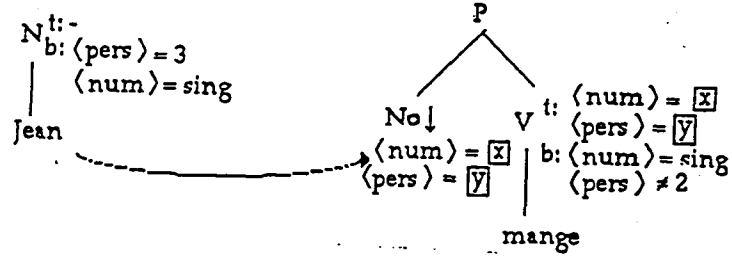
*De la bière dans toutes les cantines serait (excellent + *excellente) pour la santé des travailleurs.*

*Jean (comme + aussi bien que) Luc (n'a + *n'ont) pas pu venir.*

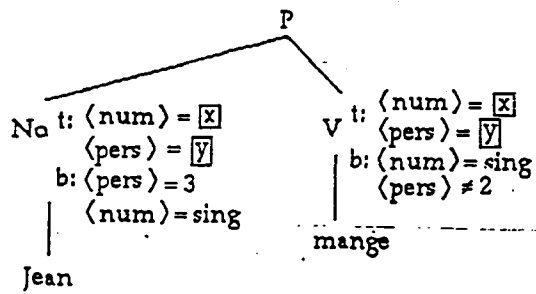
⁸ Nous ne considérons pas ici les cas d'accord du participe passé. Pour l'accord avec l'objet, il faut ajouter des traits $\langle \text{num-obj} \rangle$ et $\langle \text{genre-obj} \rangle$ qui seront activés uniquement en structure à N_1 antéposé (question, relative, clivées) ou en cas d'adjonction d'un clitique correspondant à N_1 .

Jean mange

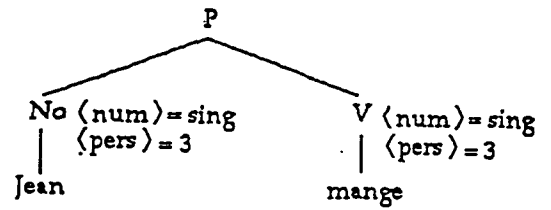
Jean mange



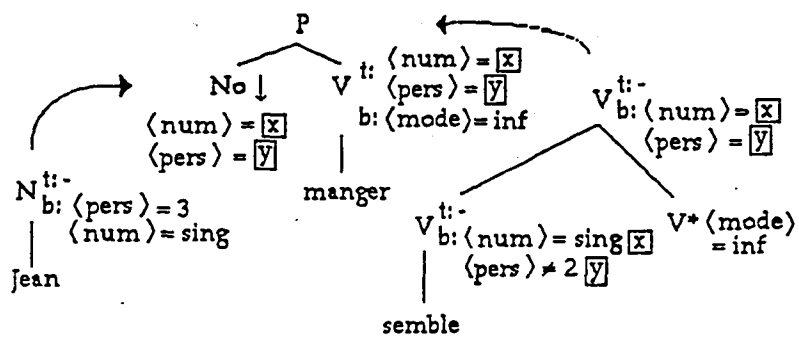
après substitution



après unification:



Jean semble manger



après adjonction et substitution

après unification:

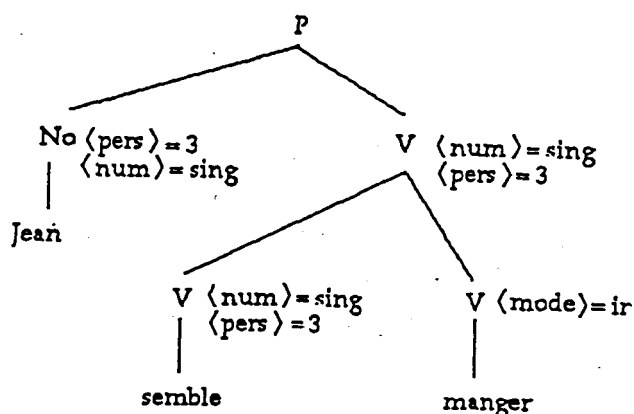
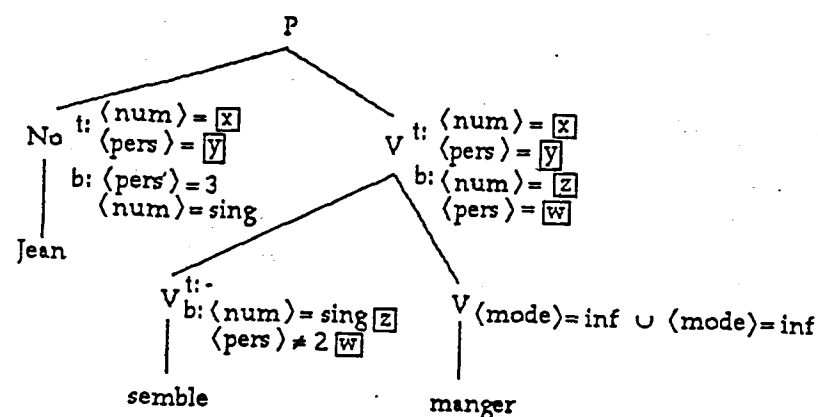


Figure 6

Lorsque le verbe prend un sujet phrastique, son nombre et sa personne (masculin singulier) sont directement indiqués dans l'arbre (en top)⁹:

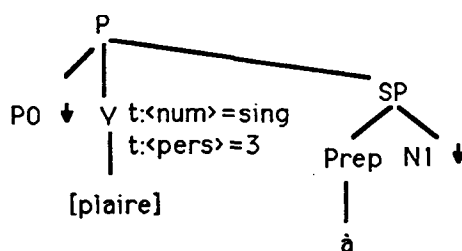


Figure 7

D'autres phénomènes d'accord apparaissent dans les phrases où un déterminant figé est un possessif obligatoirement coréférent à un argument de la phrase (voir section 4 sur les expressions figées).

1.1.4. Contraintes sur les compléments

a) Compléments pluriels

Les verbes à complément pluriel ont été étudiés par J-P. Boons, A. Guillet, C. Leclère 1976a (tables 32PL et 38PL):

*On a entassé (*Jean + les gens) dans cette pièce*
*On a divisé le gâteau en (*un morceau + deux morceaux)*

Certains de ces verbes prennent des compléments collectifs également:

⁹ Si l'on ajoutait la coordination entre phrases, il faudrait peut-être considérer que le sujet phrastique peut être pluriel:

Aller à la montagne et visiter des abbayes plaisent également à Marie.

On a entassé la foule dans cette pièce
 * *On a divisé le gâteau en une masse*

b) Contraintes sémantiques

Nous utilisons les mêmes traits que pour les sujets: <hum>, <abs> et <conc>, plus un trait <quant> pour les compléments obligatoirement quantifiés. C'est le cas du verbe *peser* dans son emploi sans passif: *N0 peser N1*: <quant> = +. Le trait <quant> peut venir du nom tête (si c'est un nom de mesure) ou du déterminant: *Il pèse un certain poids*, *Il pèse trois kilos*.

c) Clitiques figés

Pour les verbes à complément clitique figé (*en vouloir*, *s'évanouir*), on considère que l'unité lexicale est formée des deux éléments -verbe et clitique - et que lui correspond un arbre où tous deux figurent:

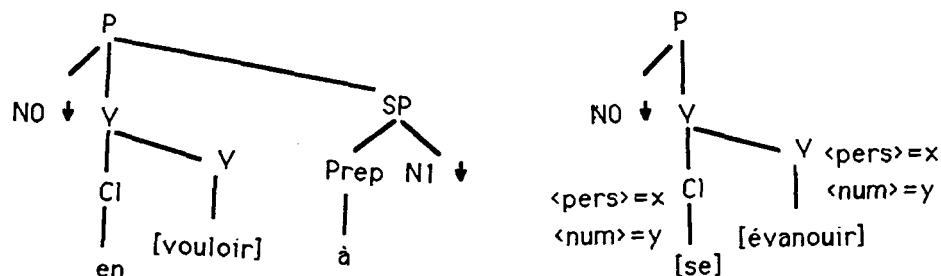


Figure 8

Ainsi *N0 acquitter N1 (de N2)* et *N0 s'acquitter de N1* seront deux verbes différents, de même pour *N0 apercevoir N1* et *N0 s'apercevoir de N1*.

d) Compléments phrastiques

Conformément aux positions de Z. Harris 1976, nous considérons les complétives comme des phrases de base et les 'principales' comme des ajouts. Les verbes à complétives sont donc représentés en TAG par des arbres auxiliaires de racine et noeud-pied phrastique et s'adjoignent sur des arbres phrastiques initiaux ou sur d'autres arbres à complétives:

Jean pense que Marie promettra à Paul que Bob partira

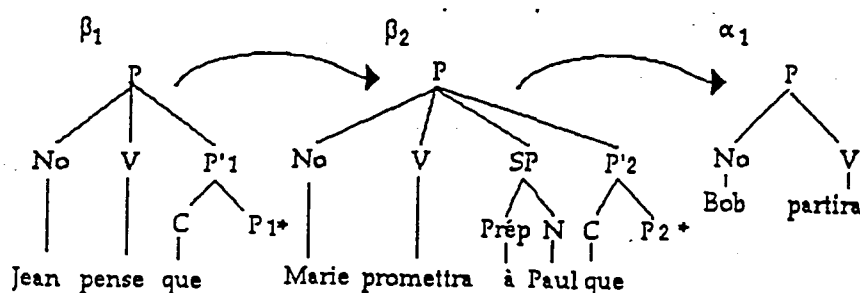


Figure 9

Cette représentation est motivée sémantiquement par la parenté souvent constatée entre "principales" et adverbes, et la transformation des "principales" en incises:

- Jean pense que Marie est une fripouille*
- = *Selon Jean, Marie est une fripouille*
- = *Marie, pense Jean, est une fripouille.*
- Jean a promis à Marie qu'il ira à Bilbao*
- = *Jean ira à Bilbao, selon sa promesse à Marie*
- = *Jean, il l'a promis à Marie, ira à Bilbao.*

Mais son intérêt essentiel est syntaxique puisqu'elle permet un traitement original des dépendances non bornées (voir chapitre 5).

Les verbes à complément phrastique portent sur le noeud pied de leur arbre auxiliaire un trait contraignant le mode de leur complétive. L'on représente ainsi N_0 craindre que P_{subj} , N_0 penser que P_{ind} , ou N_0 promettre (de P_{inf} + que P_{ind})¹⁰:

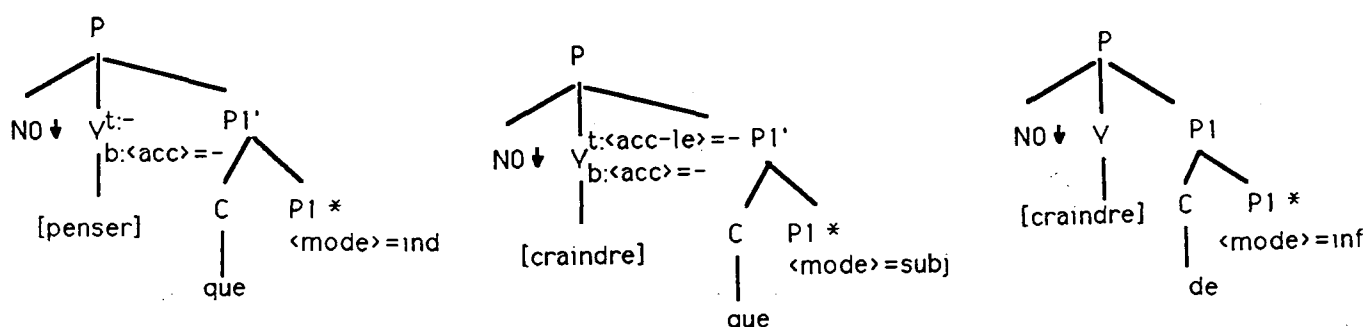


Figure 10

¹⁰ On ne considère pas les cas d'induction du subjonctif (contextes négatif, interrogatif...) Cf M. Gross 1968, H. Huot 1986.

Puisque le noeud correspondant à la complétive est directement accessible dans la structure du verbe, on peut le marquer sans avoir, comme dans une grammaire basée sur des règles d'écriture hors contexte, à faire d'abord passer la contrainte par le noeud du verbe "principal". On peut donc utiliser le même trait <mode> que celui associé aux formes verbales, à la différence d'une grammaire hors contexte où il faudrait distinguer deux traits (par exemple *mode* et *mode-compl*, l'un pour le mode du verbe lui-même, l'autre pour celui de sa complétive) pour pouvoir "faire passer" l'information de la règle lexicale:

$V[\text{mode}=\text{ind}] \rightarrow \text{craint}[\text{mode}=\text{ind}, \text{mode-compl}=\text{subj}]$

à la règle suivante:

$VP[\text{mode}=x] \rightarrow V[\text{mode}=x, \text{mode-compl}=y] S[\text{mode-compl}=y]$

sans la confondre avec la valeur du mode du verbe principal.

L'on considère les infinitives comme des phrases alternant avec les complétives (Cf chapitre 2; pour plus de détails sur la représentation des infinitives voir chapitres 5 et 6).

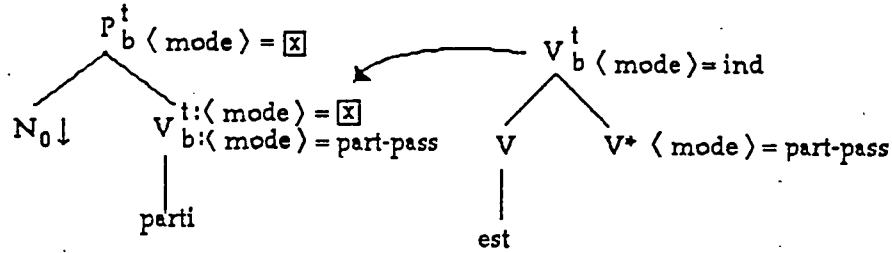
Les contraintes sur le mode des complétives sont toujours dynamiques, c'est-à-dire indiquent le résultat attendu et non le noeud exact ni l'arbre auxiliaire dont l'adjonction est attendue. Ainsi, si l'on veut exprimer le fait qu'une phrase bien formée doit être à l'indicatif (ou à l'impératif), il suffit de placer cette équation <mode> = ind (ou imp) en top du noeud racine P^{11} . Si le verbe tête de la structure est à l'indicatif (ou à l'impératif) le trait mode du bottom du noeud P pourra s'unifier avec ce trait top et l'arbre sera considéré comme bien formé. Rappelons qu'un arbre est considéré comme bien formé si à la fin de sa dérivation les traits top et bottom s'unifient à tous les noeuds. Si le verbe tête a un mode d'une autre valeur, par exemple participe passé, l'unification échouera au noeud P dont le top et le bottom auront des traits <mode> incompatibles et il en résultera une contrainte d'adjonction obligatoire. Mais il faut noter que cette contrainte vaut au niveau de l'arbre tout en entier et non au niveau d'un noeud particulier. De fait, elle pourra être satisfaite par des adjonctions différentes. Partons de la phrase rejetée *Jean parti* :

1) si l'on adjoint directement un auxiliaire à l'indicatif, par exemple *est*, le trait top mode de V sera maintenant indicatif et par conséquent le trait mode bottom au noeud P aussi, et l'unification réussira à tous les noeuds. De fait la séquence obtenue *Jean est parti* est correcte.

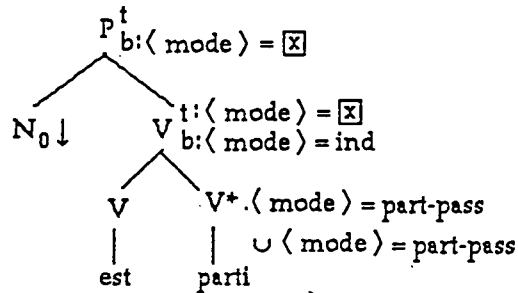
2) si l'on adjoint un auxiliaire à un autre mode, par exemple *être*, l'unification échoue toujours en P; en revanche elle réussit à partir du moment où l'on continue d'adjoindre

¹¹ Si l'on veut tenir compte des cas où cette phrase sera substituée comme argument phrastique (Cf chapitre 5), il faudrait indiquer en fait un rapport d'implication: <princ> = + => <mode> = ind/imp.

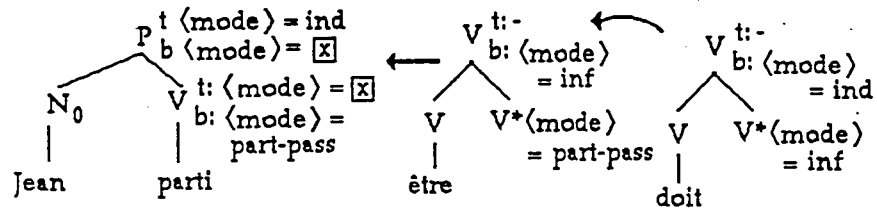
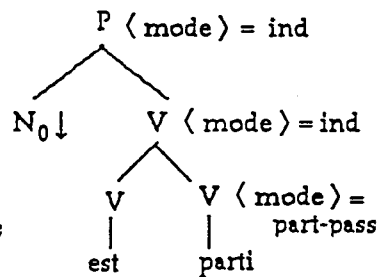
des auxiliaires et que le dernier adjoint est à l'indicatif, autorisant par exemple la séquence: *Jean doit être parti*.



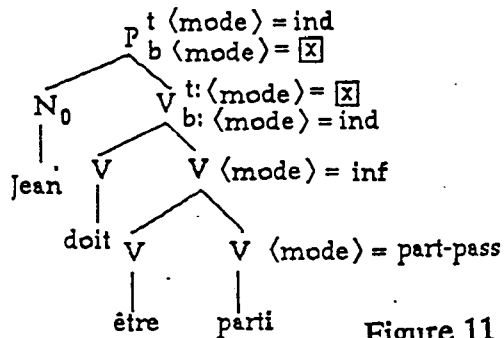
après adjonction:



après unification:



après adjonction:



après unification:

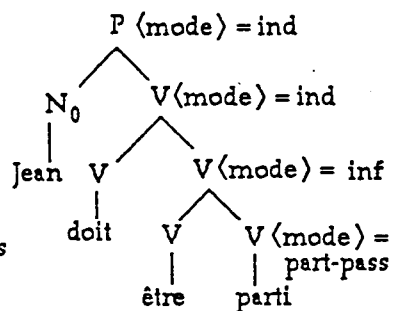
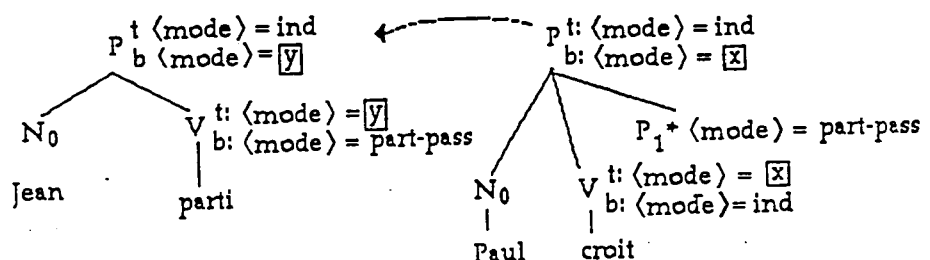
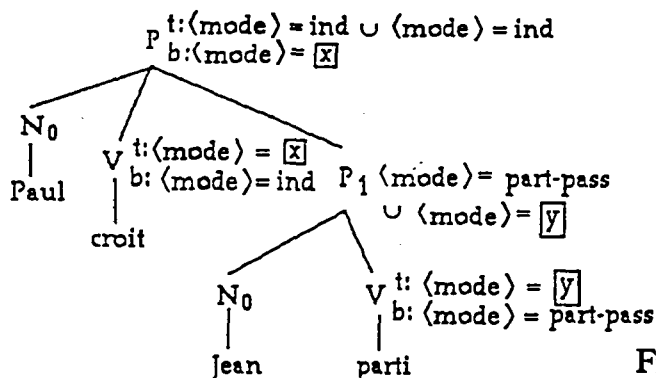


Figure 11

3) si l'on adjoint un arbre auxiliaire phrastique (verbe à complétive), le trait mode = ind de P en tant que trait top s'unifie avec le noeud racine de ce nouvel arbre phrastique, la contrainte est donc déplacée au niveau du verbe de la principale. Ainsi, *Paul croit Jean parti* sera accepté, de même que *Paul semble croire Jean parti ...*



après adjonction:



après unification:

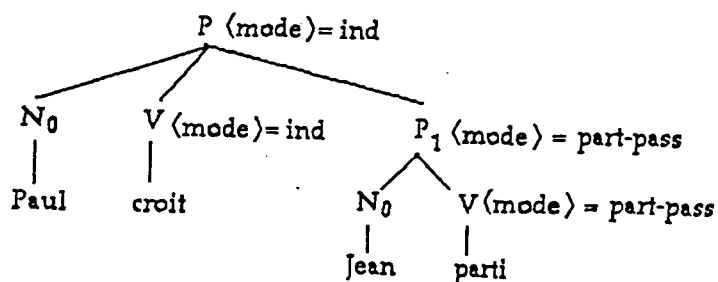


Figure 12

C'est ainsi que l'unification permet un mode d'expression déclaratif particulièrement commode, là où des contraintes d'adjonction statiques associées à un noeud obligerait à prévoir autant d'arbres différents qu'il peut y avoir de cas d'adjonction obligatoires, ce qui est impossible dans un cas de propagation récursive d'une contrainte comme celui-ci.

Les compléments n'ont pas d'autonomie sémantique et font partie de l'entrée lexicale du verbe principal. On peut exprimer, sous forme de disjonction sur les formes lexicales, les cas où deux compléments sont possibles:

Jean craint (que Marie parte + de voir Marie partir)

Des principes généraux de corestrictions de traits doivent contraindre la valeur du complément et celle du mode de la complétive¹². Une solution serait d'avoir, à tout noeud P, les implications suivantes :

$\langle \text{comp} \rangle = \text{de}/\grave{\text{a}} \Rightarrow \langle \text{mode} \rangle = \text{inf}$
 $\langle \text{comp} \rangle = \text{que}/\text{ce que} \Rightarrow \langle \text{mode} \rangle = \text{ind}/\text{subj}$

Par commodité, les rapports d'implications étant difficiles à gérer pour l'analyseur, et pour éviter d'avoir à introduire un trait $\langle \text{comp} \rangle$ supplémentaire pour le complément, nous avons préféré dédoubler les arbres dans les familles des verbes à complément phrastique et faire porter par ces arbres eux-mêmes les restrictions de mode désirées:

¹² Nous avons en fait très peu de cas de complément $\grave{\text{a}}$. Il s'agit de verbes dont les infinitives ont des propriétés de compléments directs et ne commutent pas avec $\grave{\text{a}}$ ce que P:

Jean apprend à nager

*Jean apprend (cela + *à cela)*

* *Jean apprend à ce que P.*

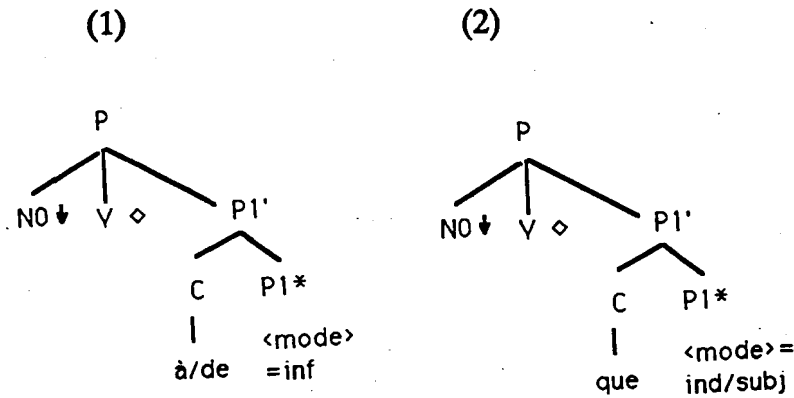


Figure 13

Selon la tête lexicale du verbe, ces structures porteront des contraintes supplémentaires (e.g. $\langle \text{mode} \rangle = \text{subj}$ pour le noeud pied de la structure (2) du verbe *craindre*).

Les complétives en *ce que* sont traitées de la même façon que les complétives en *que*, essentiellement en raison des possibilités d'extraction (voir chapitre 5)¹³.

Lorsque les infinitives se construisent sans complémenteur après le verbe, il faut prévoir de nouveaux arbres pour la principale¹⁴. L'on doit donc dédoubler les structures élémentaires (avec et sans complémenteur) pour des verbes comme *vouloir* ou *espérer*. Ce sont toujours des arbres auxiliaires:

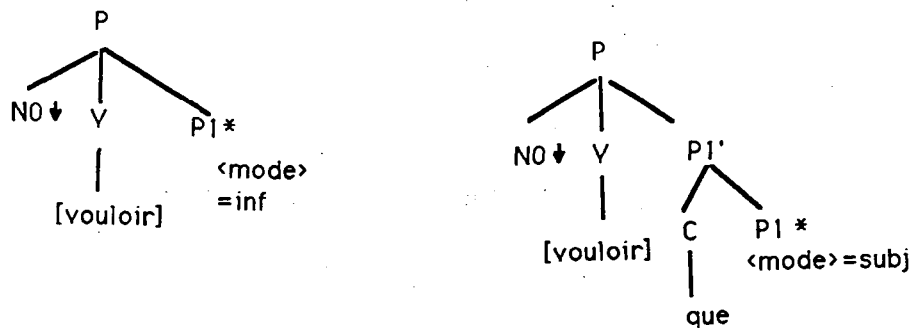


Figure 14

Les interrogatives indirectes, compléments de verbes comme *demander* ou *savoir*, sont des feuilles P de trait $\langle \text{qu} \rangle = +$, les éléments interrogatifs faisant partie de l'arbre de la

¹³ Pour un traitement différent qui les considère comme des SN à tête vide, voir H. Huot 1981.

¹⁴ A moins de noter une valeur vide du complémenteur (en disjonction avec *que*) dans la tête lexicale de l'expression. C'est le traitement adopté par exemple pour l'anglais (A. Abeillé et al. 1990). Précisons que, pour le français, il s'agit d'arbres élémentaires différents mais qu'ils sont regroupés dans une même "famille", et que nous n'avons pas à dédoubler pour autant les entrées lexicales des verbes considérés (Cf infra section 3.2).

complétive, et non de celui de la principale¹⁵. Les verbes à interrogative indirecte ont donc les mêmes structures que pour les verbes à complément infinitif, sauf que ce sont des arbres initiaux et non auxiliaires, en raison des différences d'extraction (voir chapitre 5).

Pour les contraintes sur le sujet implicite des infinitives (contrôleurs) voir chapitre 6.

e) Compléments prépositionnels

Les compléments prépositionnels sont en général "étendus" c'est-à-dire qu'on distingue les structures où ils sont constitués d'une séquence Préposition-nom ou groupe nominal de celles où ils sont formés de Préposition-complétive. Cette distinction est en effet le fait du verbe principal:

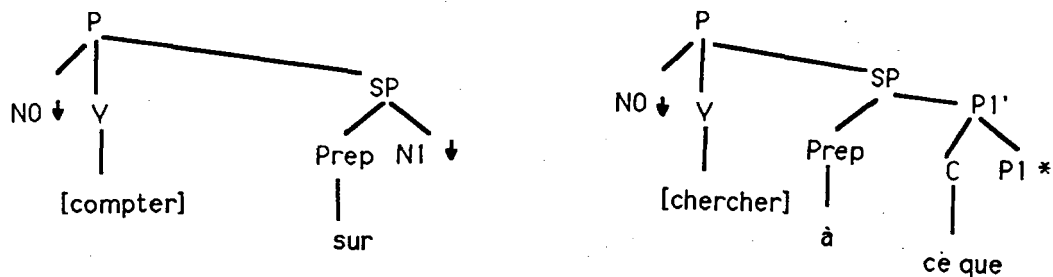


Figure 15

La préposition forme avec le verbe principal la tête de l'arbre (Cf infra section 1.3).

Si l'on considère la structure de ces arbres, il s'agit bien à chaque fois de compléments prépositionnels. Mais si l'on considère la catégorie des noeuds à substituer, en interprétant l'arbre phrastique comme une fonction sur d'autres arbres élémentaires, il faut parler d'arguments nominaux ou phrastiques, le verbe et la préposition formant ensemble le prédicat (sans pour autant former un syntagme au sens syntaxique du terme). La notion de syntagme étendu permet donc de rendre compte naturellement de cas où groupements syntaxiques et groupements sémantiques ne se recoupent pas.

Les contraintes sur les arguments nominaux ou phrastiques à l'intérieur du groupe prépositionnel s'expriment directement, comme dans le cas de compléments directs. Ainsi N_1 n'est pas contraint dans N_0 *penser à* N_1 mais doit être obligatoirement abstrait dans N_0 *s'appliquer à* N_1 (avec N_0 humain).

Le mode des complétives est également contraint: il est infinitif ou subjonctif pour N_0 *demander à* (*ce que*) P , il est obligatoirement indicatif pour N_0 : *<abs> = + provenir de ce que* P . L'alternance entre *que* P et *Prep ce que* P (*Jean rêve (de ce + E) que Marie parte*) est prévue par un arbre supplémentaire dans la famille des verbes à complément prépositionnel phrastique. Dans ce cas, la complétive est forcément au subjonctif. Comme le note H. Huot 1981, ces complétives ont les propriétés syntaxiques de compléments prépositionnels.

¹⁵ Si est un ajout à la phrase qui contraint son mode à être à l'indicatif.

1.2. Les clitiques

L'on distingue les clitiques (ou particules préverbaux, notées Cl) des autres pronoms qui sont traités comme des noms (N). Les clitiques sont considérés comme des ajouts au verbe. Ils forment donc, conformément à M. Gross 1968 ou R. Kayne 1976 un groupe avec le verbe conjugué, ou à l'infinitif. Nous avons seulement commencé leur étude et leur traitement est loin d'être achevé. Nous présentons ici les restrictions sur les combinaisons de plusieurs clitiques ainsi que certaines contraintes lexicales portant sur les clitiques arguments. Un traitement possible des corestrictions entre clitiques et arguments du verbe, ainsi que le cas particulier des constructions disloquées sont examinés au chapitre 6. Les clitiques figés ont été vus supra (voir aussi chapitre 4).

1.2.1 Combinaison des clitiques entre eux

Ces combinaisons, très fixes, maintes fois décrites, ont fait l'objet d'implémentations variées: par exemple, M. Silberztein 1989 avec un automate à état fini et K. Bashung et al. 1987 ou G. Bès 1988 avec un système de traits.

L'ordre canonique est le suivant:

je, tu, il(s), elle(s), nous, vous, on < *me, te, se, nous, vous* < *le, la, les* < *lui, leur* < *y* < *en*.

Il y a de plus incompatibilité entre *me, te, se, nous, vous* (compléments directs) et *lui, leur*.

Nous utilisons des traits distinctifs selon les fonctions: <nom> (nominatif), <acc> (accusatif), <datif> (pour les compléments en *à*) et <gen> (génitif; pour les compléments en *de*), plus les traits <refl> (réfléchi), <acc-le> (qui distingue *en* et *le, la, les*) et <datif-h> (qui distingue *y* de *lui, leur*). Chaque clitique porte à son noeud pied les traits (= -) des clitiques impossibles avant lui.

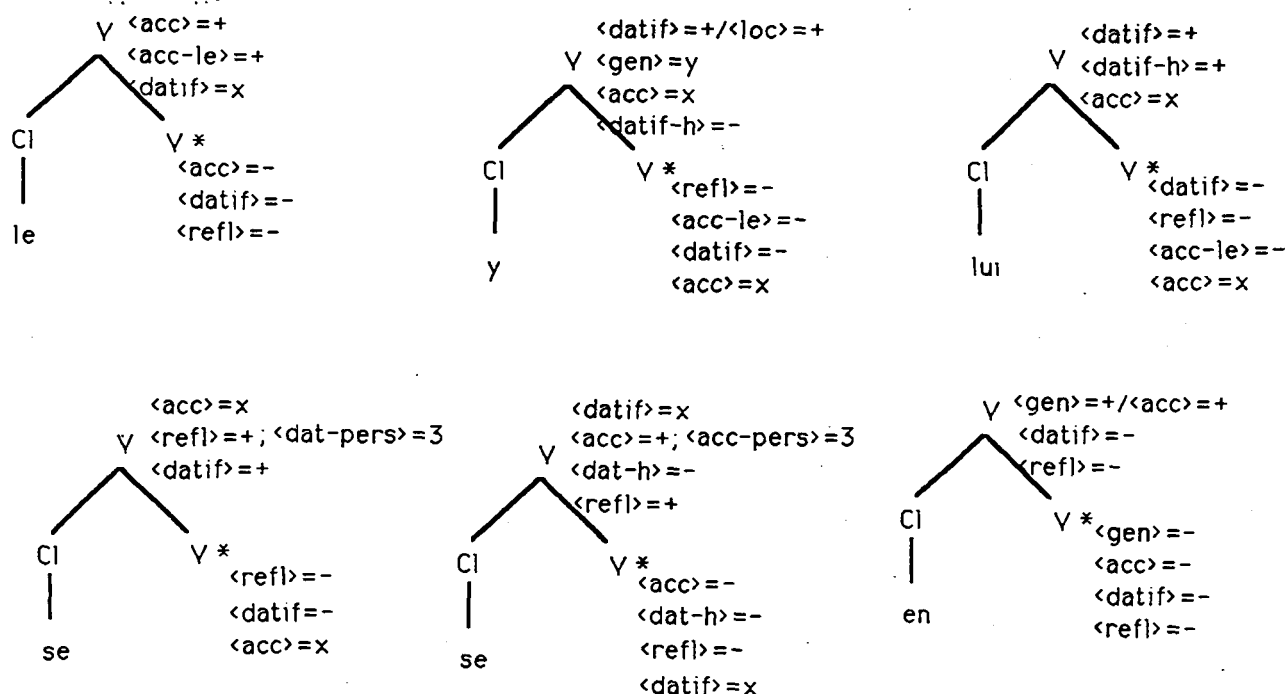


Figure 16

Les clitiques sujets inversés, comme les clitiques compléments après les verbes à l'impératif, sont représentés par des arbres différents:

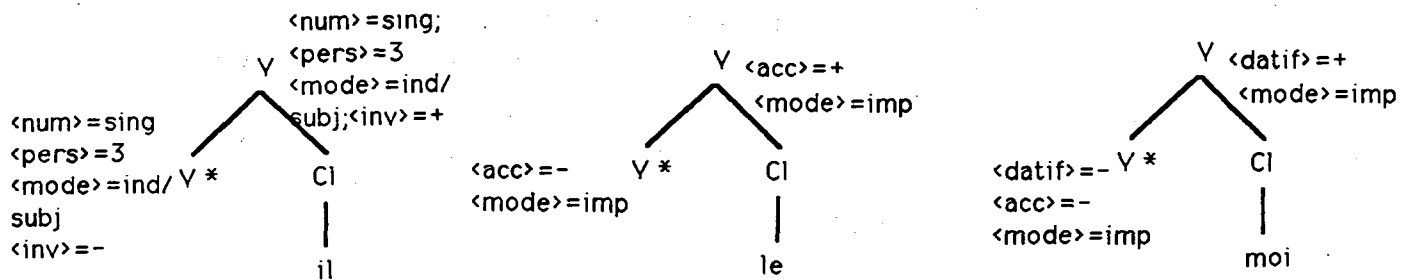


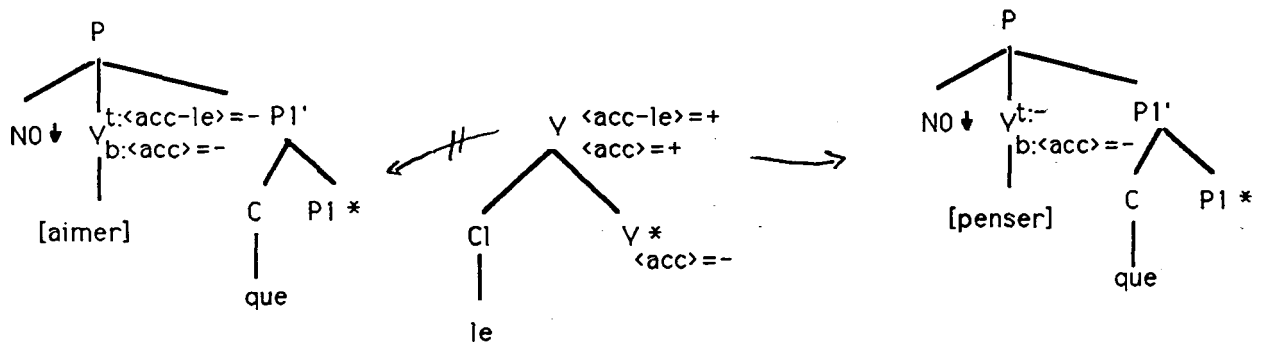
Figure 17

A l'impératif, les contraintes d'ordre entre compléments sont inversées:

en < y < le, la, les < lui, leur, moi, toi, nous, vous

1.2.2. Contraintes lexicales

Nous utilisons les mêmes traits pour contraindre la forme du clitique selon certains verbes. Ainsi, les verbes à complétives directes qui n'admettent pas la cliticisation de leur complétive ont le trait <acc-le> = - en top:



Je pense que P = je le pense
*J'aime que P = * je l'aime*

Figure 18

Les verbes à complément en à admettent généralement le clitique y sauf certains verbes à infinitive : N_0 *persister* à P_{inf} = * N_0 y *persister*. Dans ce cas, le verbe porte une contrainte top <datif> = -, qui empêche l'adjonction du clitique. Les contraintes les plus fréquentes concernent *lui* ou *leur*. Quand il s'agit d'une interdiction liée au caractère

obligatoirement non humain du complément, il n'y a pas besoin de contrainte particulière; quand le complément peut être humain, le verbe porte le trait top <datif-h> = -¹⁶.

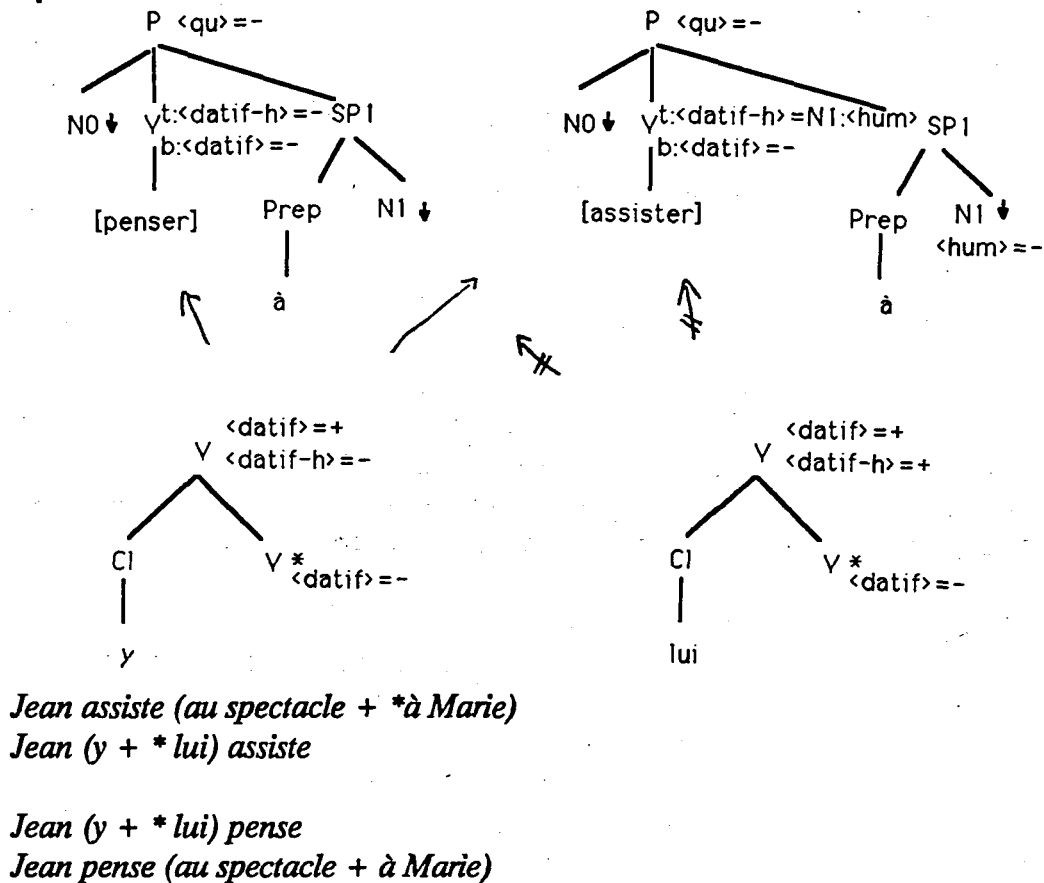


Figure 19

1.2.3. Les réfléchis et l'auxiliaire être

Lors de l'adjonction du clitique *se*, ou de *me, te, nous, vous* en emploi réfléchi, l'auxiliaire du verbe devient *être*. Nous ajoutons cette contrainte dans l'arbre de l'auxiliaires *être* (et sa négation dans l'arbre de *avoir*) :

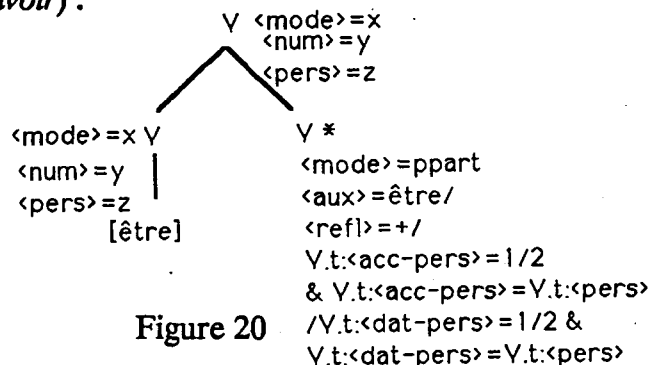


Figure 20

¹⁶ L'insertion de SN pronoms est également empêchée par le trait <datif-h> = - (**Jean ressemble à lui*). Notre représentation n'est pas très éloignée de celle qui distingue deux types de compléments introduits par *à*: les datifs (*N₀ ressemble à N₁*) et les compléments véritablement prépositionnels (*N₀ pense à N₁*) (Cf R. Kayne 1975, J-C. Milner 1982). La différence est que nous considérons *à* dans tous les cas comme une préposition.

Le trait <refl> ne note pour nous que la forme *se*. Pour *me, te, nous, vous*, nous devons donc préciser qu'il doit s'agir d'un de ces clitiques en emploi réfléchi. C'est pourquoi, nous ajoutons la contrainte sur l'accord avec le sujet.

1.3. Les prépositions

On range dans cette catégorie les prépositions traditionnelles et les conjonctions, qui sont traitées comme des prépositions à argument phrastique. Les conjonctions sont toujours têtes d'ajouts. Les prépositions sont têtes d'ajouts, mais elles peuvent aussi, en tant que marqueurs d'arguments, appartenir à la tête d'une structure phrastique, nominale ou adjectivale. On distingue ainsi par leurs structures des prépositions comme *selon* ou *malgré*, qui sont toujours têtes d'adverbiaux, de prépositions comme *de*, *à* ou *sur* qui peuvent être également marqueurs d'arguments¹⁷. Ces dernières appartiennent donc à plusieurs structures ce qui correspond bien à leur caractère ambigu.

1.3.1 Les prépositions têtes d'ajouts

Les prépositions (ou conjonctions) têtes d'adverbiaux sont représentées par des arbres auxiliaires (de la catégorie de ce qu'il modifient) avec leurs arguments éventuels.

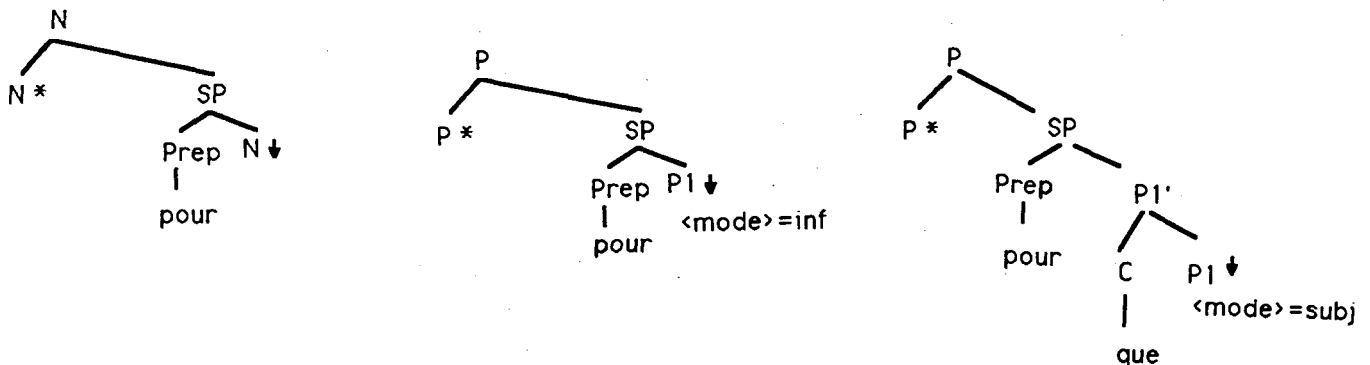


Figure 21

Les contraintes portant sur les arguments des prépositions s'expriment de la même façon que pour les arguments des verbes. Ce sont des contraintes de mode pour les arguments phrastiques, des contraintes sémantiques et de déterminant pour les arguments nominaux.

¹⁷ Une telle distinction recoupe celle, plus ancienne, entre prépositions "vides" et prépositions "pleines" (Cf notamment E. Spang- Hanssen 1963, et pour une discussion critique G. Gougenheim 1961).

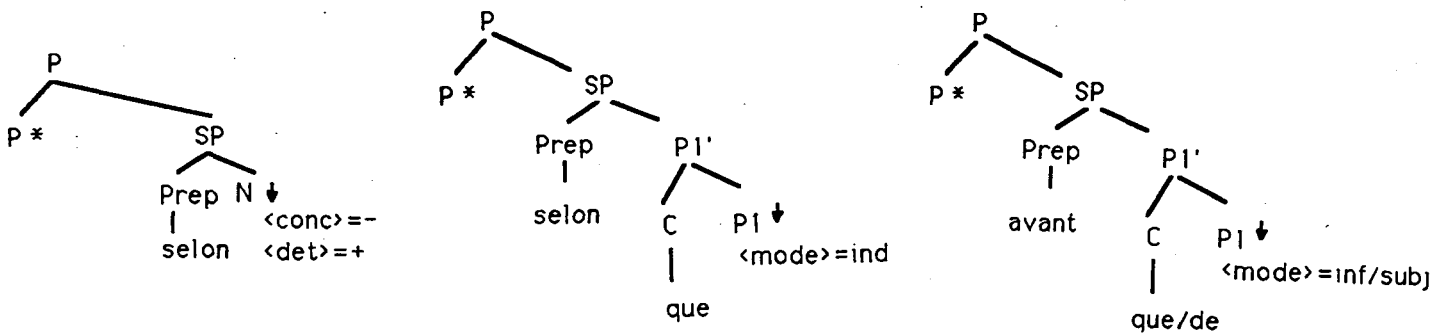


Figure 22

Il est à noter qu'une telle représentation permet de résoudre certains problèmes classiques d'attachement de syntagme prépositionnel. Chaque préposition aura autant de structures qu'il y a de catégories qu'elle peut modifier, et l'on pourra à chaque fois spécifier la structure interne du SP qui est souvent différente. Ainsi *de* suivi d'un nom concret sans déterminant sera plutôt un modifieur de nom (ou de groupe nominal; ex: *un cheval de bois*, *un anneau de jade*), tandis que *de* suivi d'un nom abstrait (de sentiment) sans déterminant plutôt un modifieur de verbe ou de phrase (*De rage*, *il est parti*; *Il pleure de honte*). Nous n'avons pas fait l'étude détaillée qui s'impose pour chaque préposition mais ce type de représentation devrait considérablement réduire les ambiguïtés superflues classiquement produites par les analyseurs syntaxiques (voir chapitre 7).

1.3.2 Les prépositions marqueurs d'arguments

Les prépositions marqueurs d'arguments n'ont pas d'existence indépendante, elles font partie de l'entrée lexicale de leur prédicat avec lequel elles forment une tête multiple¹⁸.

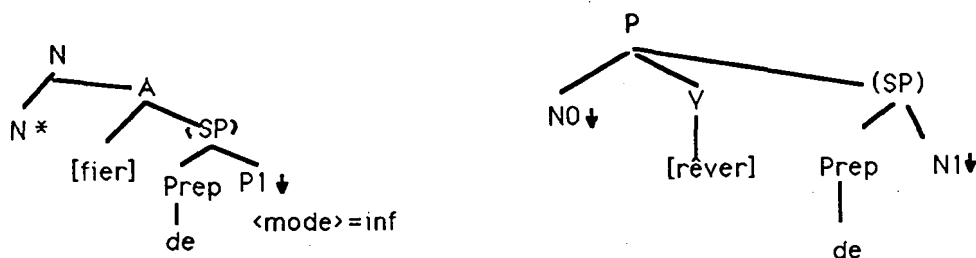


Figure 23

¹⁸ Comme pour les autres têtes multiples (Cf chapitre 4), on utilise la disjonction quand plusieurs valeurs lexicales de la préposition sont possibles (à propriétés syntaxiques et sémantiques inchangées):

Jean a rempli son camion (de + avec) ces pommes.

Ainsi, la question de leur sens ne se pose pas, puisque la seule unité sémantique est la combinaison verbe-préposition (ou adjectif-préposition ou nom-préposition) considérée comme non compositionnelle.

La plupart des contraintes imposées à leur argument (catégorie nominale ou phrastique, traits sémantiques, mode...) le sont par la tête principale de la structure. Ainsi *de* peut être suivi d'une complétive avec *fier de N/P*, mais pas avec *amoureux de N/*P*.

La plupart des prépositions marqueurs d'arguments peuvent être en même temps têtes d'adverbiaux (et c'est le contexte qui lèvera éventuellement les ambiguïtés) mais l'inverse n'est pas vrai.

Les prépositions locatives ont, outre leurs structures adverbiales, une structure particulière en tant que compléments obligatoires de certains verbes comme *N0 aller Loc N1* ou *N0 mettre N1 Loc N2* (les locatifs adverbiaux étant traités comme les autres adverbiaux). Elles ont même sens et même construction que dans leur emploi adverbial; leur valeur lexicale n'est pas fixée par le verbe et est susceptible de variations. Elles alternent avec des adverbes locatifs (*ici, là*). On ne peut donc pas utiliser un SP étendu et les traiter comme partie de la tête du verbe.

Nous les représentons comme têtes d'arbres initiaux (de racine SP) qui sont substitués aux noeuds feuilles correspondant des arbres de leur prédicat¹⁹. La composition interne du SP est toujours la même: un SN (ou un adverbe locatif), et ne dépend pas du verbe principal. Les prédicats à arguments locatifs sont donc les seules structures où le SP n'est pas "étendu".

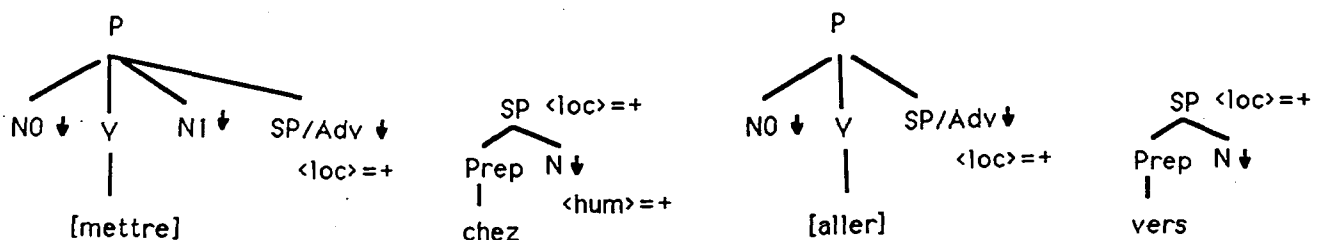


Figure 24

Nous utilisons un seul trait *<loc>* pour distinguer ces groupes prépositionnels et ces adverbes des adverbiaux de manière obligatoires également substitués (*N0 agit bien/comme il faut/de façon satisfaisante*). Il est clair qu'il faudrait des sous-distinctions, par exemple selon qu'il y a mouvement ou non (Cf A. Guillet 1990). Il faudrait aussi marquer parallèlement les noms (que nous ne marquons pour l'instant qu'avec *<hum>* : *chez Nhum*, ou *<conc>* : *dans Nconc*).

Un cas particulier est celui des SP prédicatifs mentionnés au chapitre 2, étudiés par L. Danlos 1980, qui ont la plupart des propriétés syntaxiques des adjectifs. Ils se

¹⁹ Les adverbes locatifs ont la même représentation.

construisent avec *être*, se pronominalisent en *le*, peuvent répondre à la question en *comment*: *être de bonne humeur*, *être en forme*. L'unité lexicale retenue est en fait composée des trois têtes *être*, préposition et nom, et peut s'analyser soit comme construction à verbe support, soit comme expression figée (voir chapitre 4).

1.4. Les noms

Les noms sont têtes d'arbres initiaux de racine N, comportant en feuilles leurs éventuels arguments:

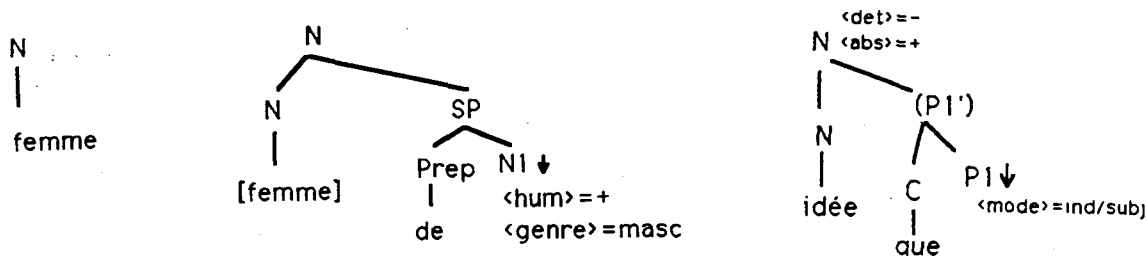


Figure 25

Il est possible de lever ainsi certaines ambiguïtés des noms. Ainsi *femme* (opposé à *homme*) aura un arbre différent de *femme (de N)* (signifiant *épouse*). Les contraintes portant sur les différents arguments s'expriment de la même façon que pour les verbes. Ainsi *femme de N* prend un complément humain (masculin), alors que le complément de *forme de N* sera concret et celui de *fondement de N* abstrait.

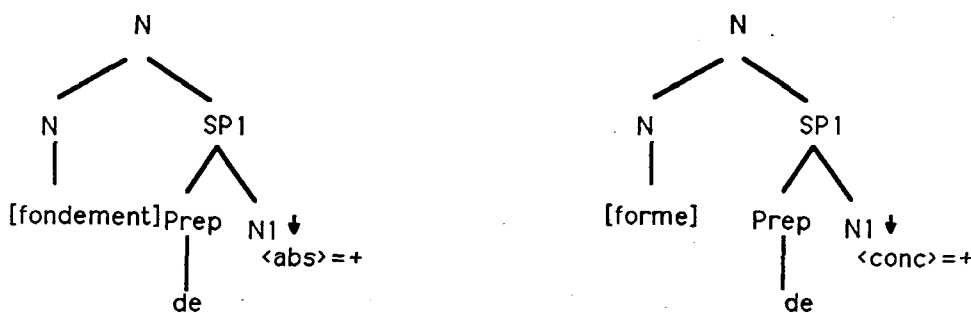


Figure 26

La "désambiguation" des noms se fait également par des traits sémantiques. Par exemple, le trait *<mass>*, qui contraint l'article partitif, permet de distinguer trois entrées pour le nom *glace* :

- glace*, N <mass> = + (=eau gelée)
- glace*, N <mass> = - (=miroir)
- glace*, N (=crème glacée)

La troisième entrée n'est pas spécifiée pour ce trait, puisqu'on a aussi bien *de la glace (à la vanille)* que *trois glaces (à la vanille)*.

Un autre moyen de lever les ambiguïtés est la représentation par phrases à verbe support. Le nom *éclat* au sens de *scandale* prend *faire* comme verbe support, mais *avoir* quand il a le sens de *brillant*²⁰.

Jean a fait un éclat hier soir

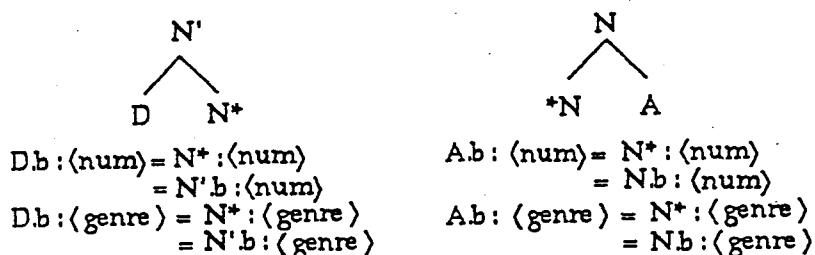
Cette pierre a (de l'éclat + un éclat remarquable)

Ces constructions étant considérées comme semi (ou non-) compositionnelles sont étudiées au chapitre 4, de même que les structures nominales qui leur sont liées. La distinction entre arguments et modifieurs du nom se fait, sur la base de la construction à verbe support correspondante, suivant les mêmes critères que la différence entre compléments circonstanciels et arguments pour les verbes.

1.4.1. Phénomènes d'accord

Le genre ou le nombre du groupe nominal est déterminé à la fois par le nom tête, le déterminant ou les modifieurs éventuels. L'unification, qui ne se contente pas de recopier des valeurs d'une catégorie à une autre mais modifie dynamiquement les valeurs en jeu, se prête particulièrement bien à une formulation déclarative de tels phénomènes. Considérons des noms morphologiquement ambigus du type *corps* (singulier ou pluriel) ou *élève* (masculin ou féminin). Le nombre du groupe *un corps* est singulier à cause de *un*, le genre de *une élève* est féminin à cause de *une*, celui de *des élèves curieux* masculin à cause de *curieux*²¹.

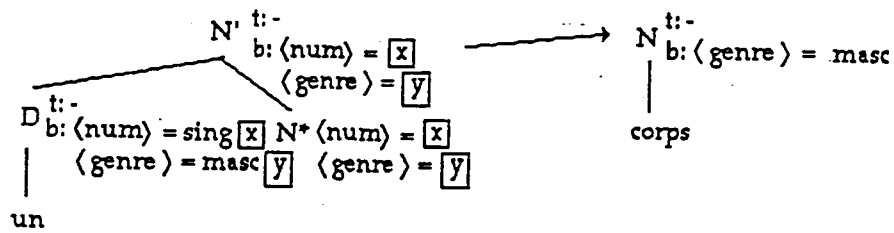
contraintes générales



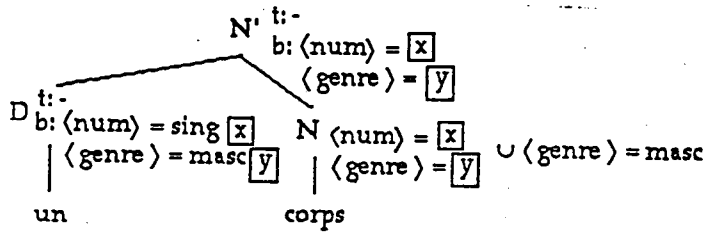
20 Pour les groupes nominaux régulièrement associés aux constructions à verbe support (Cf chapitre 4), c'est la valeur de la préposition et les traits portés par les arguments qui lèvent l'ambiguïté : *l'éclat de Jean* vs *l'éclat de cette pierre*.

21 L'unification est commutative et les dépendances ne sont pas orientées. La question ne se pose pas de savoir si c'est le déterminant qui s'accorde avec le nom ou l'inverse. Nous disons simplement qu'une forme peut être plus déterminée qu'une autre pour un trait donné: *une* est déterminée pour le trait $\langle \text{num} \rangle$, *corps* ne l'est pas ($\langle \text{num} \rangle = \text{sing/plur}$).

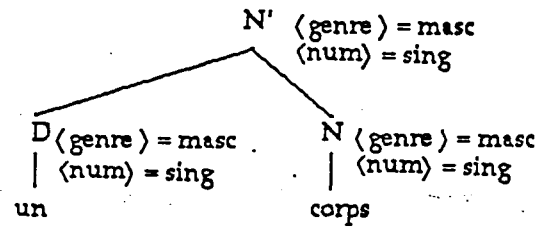
un corps



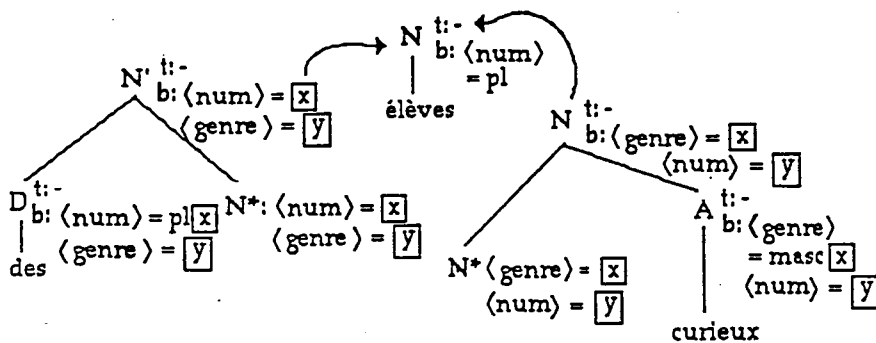
après adjonction:



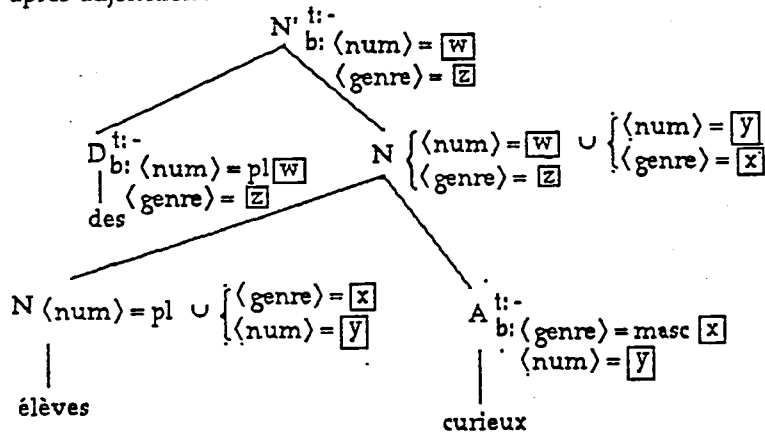
après unification:



des élèves curieux



après adjonction :



après unification:

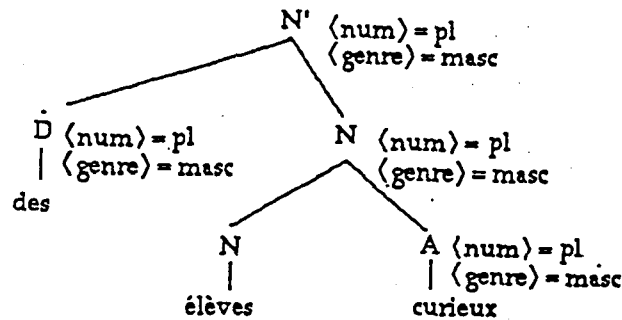
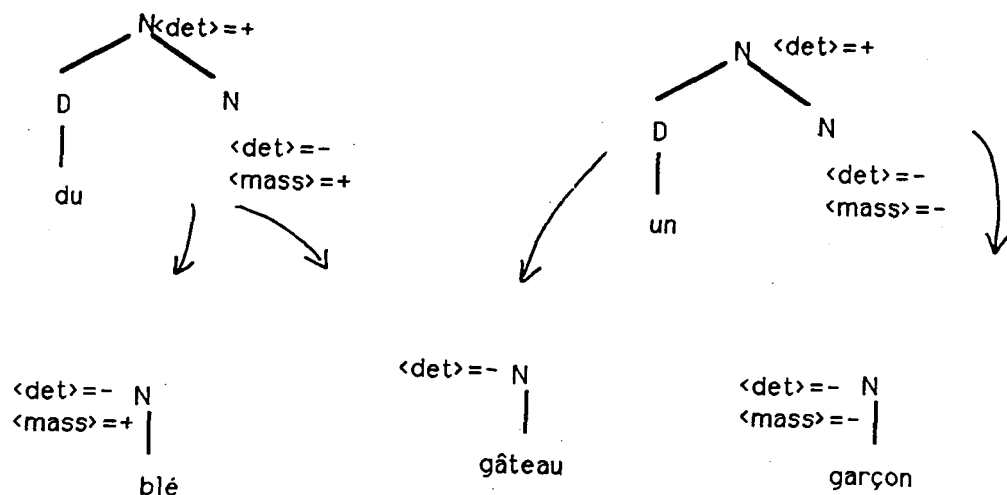


Figure 27

1.4.2 Contraintes sur les déterminants

En général, les rapports de dépendance entre noms et déterminants sont exprimés dans la cadre du groupe nominal considéré comme domaine autonome. Ainsi, suivant la valeur du nom tête, le déterminant du syntagme nominal pourra être indéfini, partitif, zéro etc...

On utilise ainsi le trait <mass> pour contraindre l'adjonction de *un* ou *du*. Ainsi, on autorise *un garçon*, *du blé*, *un gâteau* et *du gâteau* et on interdit **du garçon* et **un blé*²² :



(un + du) gâteau (un + *du) garçon (*un + du) blé

Figure 28

²² Dans ce dernier cas, on pourrait en fait permettre l'adjonction de *un* à condition qu'un modifieur (adjectif ou relative) s'adjoigne également, ce qui s'indique <mass>=+ => <modif>=+ au noeud pied N dans la structure du déterminant: *un blé* (*E + de bonne qualité); de même pour le déterminant défini pluriel, voir F. Corblin 1988.

Un cas particulier est celui des déterminants figés qui sont considérés comme appartenant directement à l'arbre du nom (qui comprend donc un noeud feuille en plus):

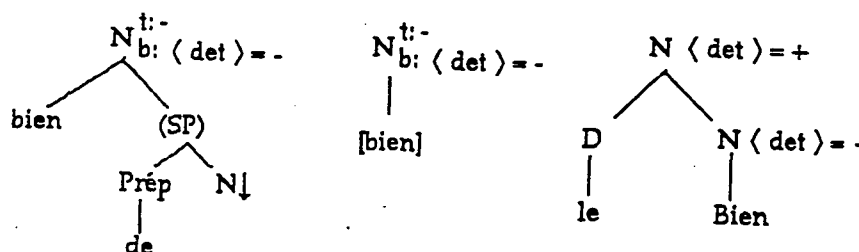


Figure 29

Ainsi le groupe nominal *le Bien* est considéré comme sans lien avec les groupes nominaux ayant le même nom tête *bien* mais à déterminants plus variables. De fait, on veut clairement distinguer le sens d'une expression comme *le Bien* (ou *le Mal*) d'au moins deux types de constructions: (*le bien de N + son bien*) (toujours singulier, sens de bien-être) d'une part, et *un bien, des biens* (variable en nombre, sens de possession) d'autre part.

Il en est de même pour les déterminants figés dans une expression verbale, exemple N_0 *chasser le Ngibier*. Dans ce cas nous considérons que nous avons affaire à une expression figée: [*chasser*] [*le*] *N*, et nous utilisons seulement des traits d'accord pour le déterminant (voir chapitre 4 sur les expressions figées).

Mais ce n'est pas seulement le nom qui contraint la valeur du déterminant. La structure de la phrase où s'insère le syntagme nominal, et notamment le choix du verbe, la conditionne aussi. C'est pourquoi nous utilisons également des traits contraignant la détermination aux noeuds *N* des arbres phrastiques.

Ainsi, c'est le verbe principal qui en général imposera la présence d'un déterminant à ses arguments nominaux, et dans certains cas l'interdira, par exemple avec N_0 *transformer N_1 en N_2* :

*Jean a transformé Jeanne en (E + *une + *cette) sorcière*

Prenons l'exemple du déterminant zéro, auquel nous assimilons ici par commodité l'absence de déterminant. Les noms sont marqués $\langle \text{det} \rangle = -$ en leur bottom (sauf la plupart des noms propres et des pronoms qui sont par eux-mêmes déterminés). S'ils sont marqués $\langle \text{det} \rangle = +$ en leur top, l'adjonction d'un déterminant sera obligatoire pour la plupart des noms communs et certains noms propres (*la France*). Mais le caractère obligatoire du déterminant varie aussi suivant les contextes, et l'on préfère marquer

<det> = + au niveau du noeud dans l'arbre où le N sera substitué. Ainsi, *terre* demande en général un déterminant, sauf dans des positions telles que second complément d'un verbe comme *N₀ couvrir N₁ de N₂* ou *N₀ recouvrir N₁ de N₂*:

Il a recouvert Jean de (E + cette) terre.

De même après une préposition comme *en*, le déterminant sera en général interdit, lorsque la préposition est tête d'un groupe adverbial. Alors que lorsque *en* introduit le complément d'un verbe, le déterminant peut être obligatoire: *Je crois en ((*E + cette) divinité (à moins que le nom soit déjà déterminé: Je crois en (Dieu + lui)).*

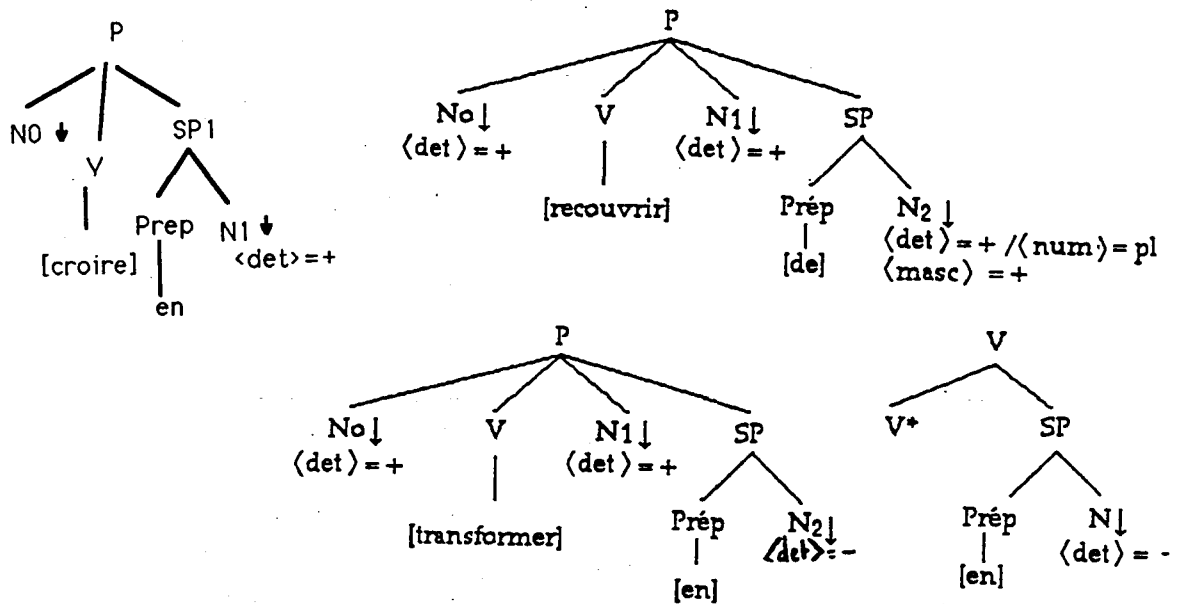
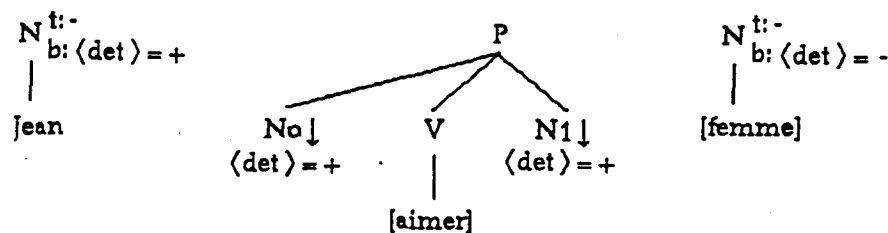


Figure 30

C'est donc au niveau de l'arbre du verbe ou de la préposition et non au niveau de celui du nom que sera marquée cette contrainte et que l'adjonction d'un déterminant devient obligatoire:

Jean aime une femme



après substitution:

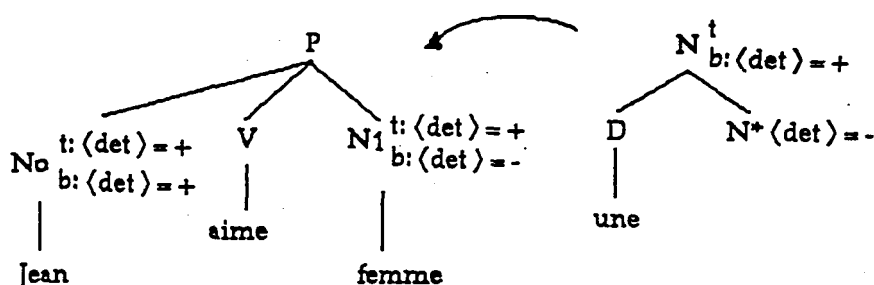
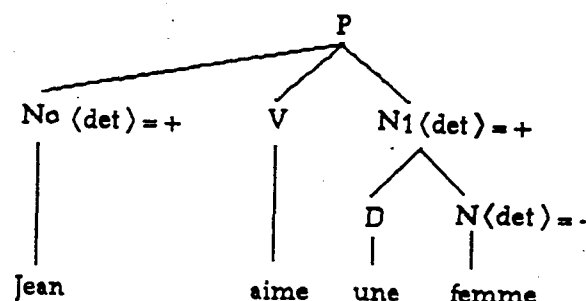
après adjonction
et unification:

Figure 31

Seul le cas du déterminant zéro a été implémenté pour l'instant. Des restrictions similaires existent pour les autres déterminants, par exemple:

Je veux (l'argent + de l'argent)
*J'aime (l'argent + * de l'argent)*

On exige (de l' + cette) information
On apprécie (de l' + cette) information*

Mais si l'on change le mode du verbe principal, la contrainte tombe:

J'aimerais (l'argent + de l'argent)
On apprécierait (de l' + cette) information)

Nous n'entrons pas ici plus avant dans le détail des contraintes gouvernant les déterminants, et leur interaction avec la structure prédicative d'une part, la présence d'autres modificateurs d'autre part.

1.5 Les déterminants

Conformément à l'étude esquissée au chapitre 2, nous traitons les déterminants complexes sur le même plan que les déterminants simples. Nous allons d'abord expliquer pourquoi nous considérons les déterminants comme des ajouts, puis exposer brièvement certains traits distinctifs que nous utilisons pour contraindre les combinaisons de déterminants entre eux²³.

²³ Nous ne traitons pas des enclises. Il y en fait un prétraitement qui décompose *au* en *à le* et, de façon optionnelle, *du* en *de le*. La forme *le* est marquée spécialement comme ne pouvant être un clitique, afin de ne pas créer d'ambiguïtés superflues.

1.5.1. Déterminants complexes et prédéterminants.

Substituer les déterminants simples à des noeuds prévus comme tels dans les arbres des noms ne poserait pas de problèmes particuliers, et c'est ce que nous avons fait dans des publications antérieures (Cf A. Abeillé 1988b). En revanche, il est plus difficile de substituer les déterminants complexes et les prédéterminants. C'est pourquoi, nous proposons, dans la perspective d'un traitement unifié, d'adjoindre tous les déterminants.

Considérons l'introduction des déterminants complexes par substitution, c'est-à-dire comme frères du nom (ou de N'):

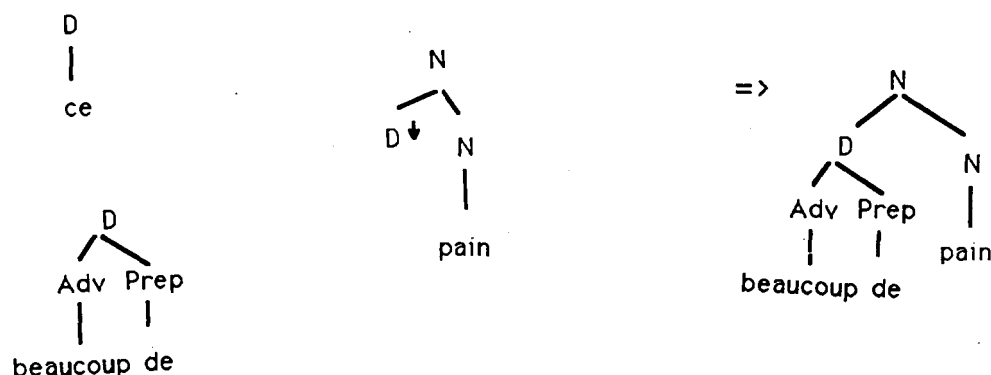


Figure 32

Le résultat empêche de faire apparaître une structure de groupe prépositionnel ou de SN complexe pourtant suggérée par la séquence des catégories, l'insertion régulière de modifieurs : *la (E+ grande) majorité de ces gens*, et la pronominalisation éventuelle : *j'en mange beaucoup (E + ,de ce pain)*, *j'en ai lu la plupart (E, de ces romans)*²⁴.

L'on pourrait penser à la solution inverse, à savoir substituer les noms (ou les SN) dans les structures des déterminants complexes (ou des prédéterminants). L'emboîtement des structures serait celui désiré, mais il devient alors impossible d'insérer des déterminants entre verbe et nom dans des expressions à verbe et nom figés et déterminant du nom assez libre (typiquement les phrases à verbe support et nom prédicatif: *N₀ pousser un cri/ un grand nombre de cris ...*). En effet, le nom fait partie de l'arbre élémentaire phrastique et la substitution ne peut se faire à un noeud intérieur (Cf chapitre 4).

²⁴ Nous souhaitons ici simplement pouvoir considérer *de N* comme un syntagme dans des tours comme *trois de mes amis* ou *beaucoup de pain*. Nous l'appelons SP par commodité mais il est clair qu'il faudrait une autre catégorie (ou des traits particuliers) si l'on voulait incorporer à notre grammaire les différences de propriétés éventuelles de ces syntagmes avec les syntagmes prépositionnels ordinaires (Cf J-C. Milner 1978).

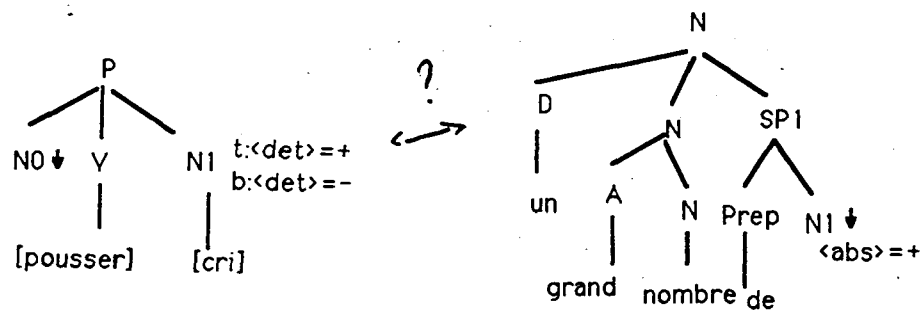


Figure 33

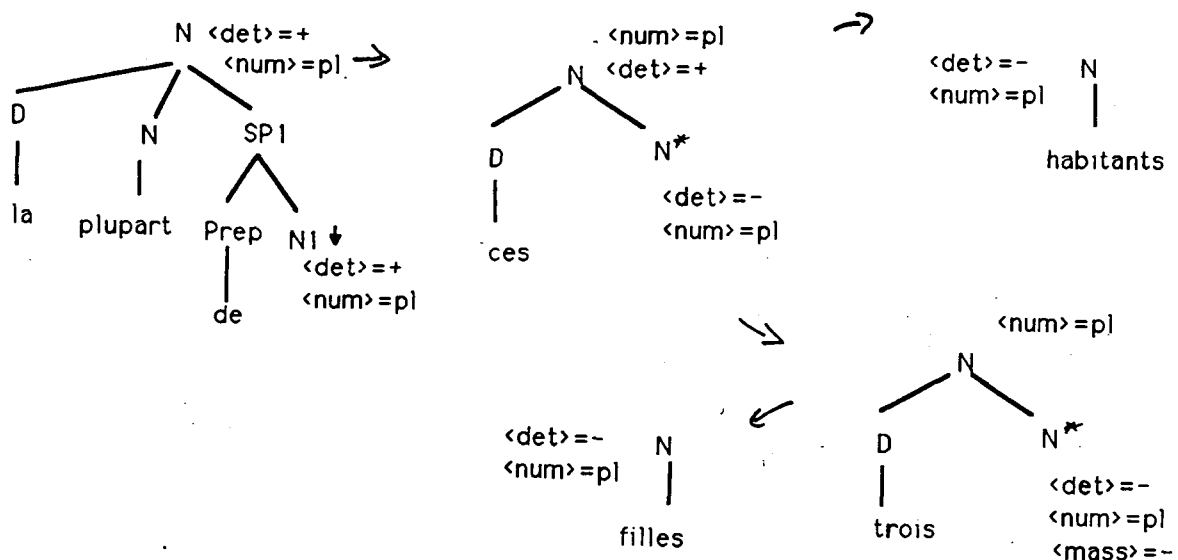
Nous représentons donc les déterminants comme des arbres auxiliaires qui s'adjoignent au nom.

1.5.2 Les combinaisons de déterminants

L'adjonction autorise des combinaisons éventuelles entre plusieurs déterminants:

Ces habitants sont venus

La plupart de ces habitants sont venus



Trois filles sont venues

Ces trois filles sont venues

Figure 34

Il s'agit de contraindre les combinaisons interdites. Ainsi, on ne pourra avoir **Trois ces filles* puisque *ces filles* porte à sa racine le trait $\langle \text{det} \rangle = +$. En revanche, on aura *Trois filles* ou *Ces trois filles* puisque *trois filles*, ne portant pas de trait $\langle \text{det} \rangle$, peut se comporter comme N' ou N'' .

Nous n'avons pas décrit toutes les contraintes en jeu. Nous avons simplement distingué les déterminants s'adjoignant toujours à un nom déterminé (certains D_{adv} , les Prédéterminants et certains D_{nom} de M. Gross 1977), ceux s'adjoignant toujours à un nom non déterminé (les D_{adj}), et ceux pouvant s'adjoindre indifféremment dans les deux cas (certains D_{adv} , certains $D_{adj} + de$ et certains D_{nom}). Cette contrainte s'exprime par un trait sur le noeud N du pied de l'arbre (respectivement : $\langle \text{det} \rangle = +$, $\langle \text{det} \rangle = -$ ou rien)

*la plupart de (*E + nos) amis*
beaucoup de (E + nos) amis

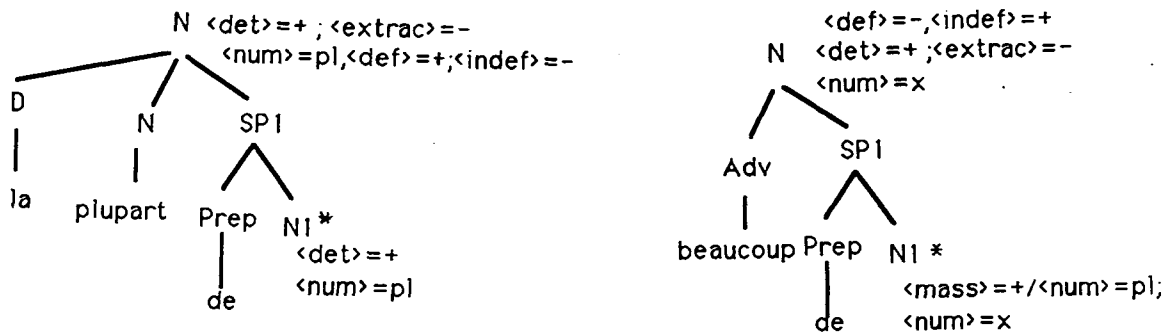


Figure 35

1.5.3 Propagation des traits associés

Un problème qui se pose est celui de la propagation des traits selon la combinaison des déterminants. En effet, selon les traits et selon les déterminants, il y aura blocage ou propagation jusqu'à la racine du groupe nominal.

Le trait nombre se propage différemment avec *la plupart de N* ou *la majorité de N* :

*La plupart des gens (*est + sont) là.*
La majorité des gens (est + sont) là.

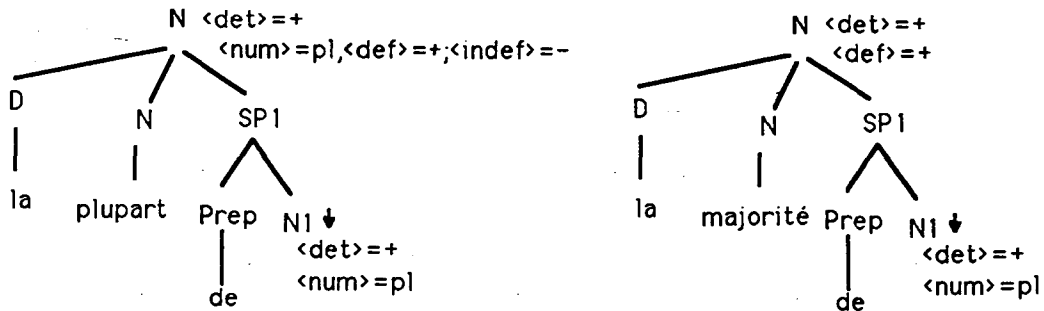


Figure 36

Un cas difficile est celui du caractère défini ou indéfini du groupe résultant. Nous avons pris pour critère des groupes indéfinis la possibilité d'extraposition à partir de la position sujet, et pour les définis, la dislocation à gauche (avec clitique *le*). Il est à noter que ces deux propriétés ne sont pas forcément en distribution complémentaire. Nous utilisons donc deux traits <def> et <indef> distincts, puisque selon nos critères, certains groupes nominaux pourront être à la fois <def>=+ et <indef>=+, et certains <def>=- et <indef>=-.

La différence est nette entre *un* et *le*:

(Un + Le) garçon est venu
*Il est venu (un + *le) garçon.*
*(*Un + Le) garçon, je l'ai vu.*

Les Dadj sont en général "opaques": ils déterminent le trait <indef> de tout le groupe, quel que soit le déterminant qui les suit. Ainsi, si *Dadj N* est indéfini, la séquence *Dadj de Detdef* est également indéfinie.

Il est arrivé certaines personnes
** Il est arrivé ces personnes*
Il est arrivé hier certaines (E + de ces) personnes

Certaines séquences Dadj N sont à la fois <def>=- et <indef>=-:

** Il est arrivé chaque personne*
** Chaque personne, je l'ai vue*

Les Dadv sont en général opaques également (<indef>=+ et <def>=- quel que soit le déterminant qui les suit):

Il est venu beaucoup de (E + ?ce genre de) gens avant hier
*Beaucoup de (*E + ?*ces) gens, je les ai (déjà) vus*

Les Dnom transmettent ou non les traits des déterminants qui les suivent:

Plein (d' + de nos) amis sont arrivés hier.

Il est arrivé hier plein (d'amis + ? de nos amis).*

** Plein (d' + de nos amis), je les connais*

** Il est arrivé hier tes amis*

Il est arrivé hier une partie de tes amis

Tes amis, je les connais

** Une partie de tes amis, je les connais*

Plein de N transmet du noeud pied au noeud racine le trait *<indef>*, mais pas le trait *<def>*. *Une partie de N* est toujours *<def>=-* et *<indef>=+* quel que soit le déterminant qui les suit.

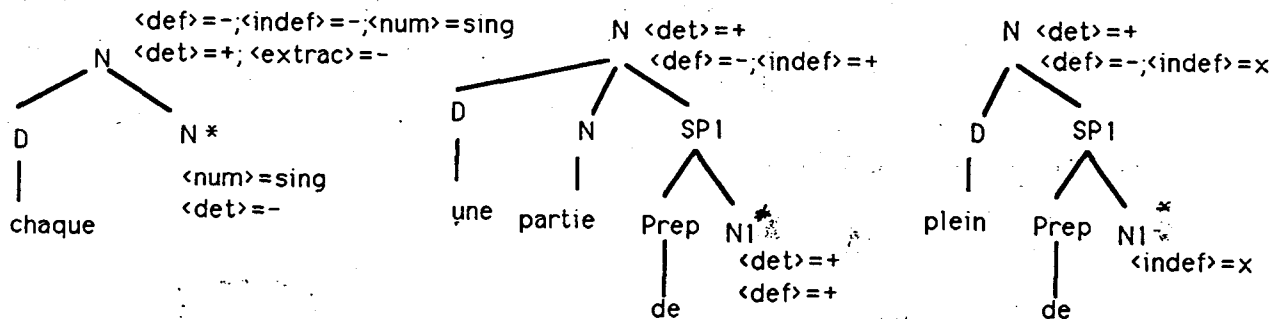


Figure 37

Les traits *<def>* et *<indef>* sont également utilisés pour régler les combinaisons de déterminants entre eux. Ainsi, beaucoup de N porte à son noeud pied *<def>=+* / *<det>=-* (pour *beaucoup de (ce + E) pain*). Ils ne peuvent, en revanche, être directement utilisés pour les contraintes d'extraction (groupe nominal acceptable en position topicalisée ou non). Ainsi, on ne peut utiliser le trait *<def>=+* ou *<indef>=-* puisque ces deux valeurs s'appliquent à un groupe comme *la plupart de ces gens*, alors qu'il est difficilement topicalisable par *c'est que* (Cf chapitre 2) :

La plupart de ces gens, je les connais

** Il est venu la plupart de ces gens*

? C'est la plupart de ces gens qui sont venus.*

Dans ce cas, c'est un trait spécifique *<extrac>*, qu'il faut employer.

1.6 Les adjectifs

Les adjectifs sont représentés avec leurs compléments éventuels. Nous notons A le groupe ainsi formé, pour ne pas introduire de distinction catégorielle superflue entre adjectifs sans compléments et adjectifs avec compléments (Cf chapitre 2).

En tant que modificateurs de noms, les adjectifs sont représentés par des arbres auxiliaires de racine N²⁵:

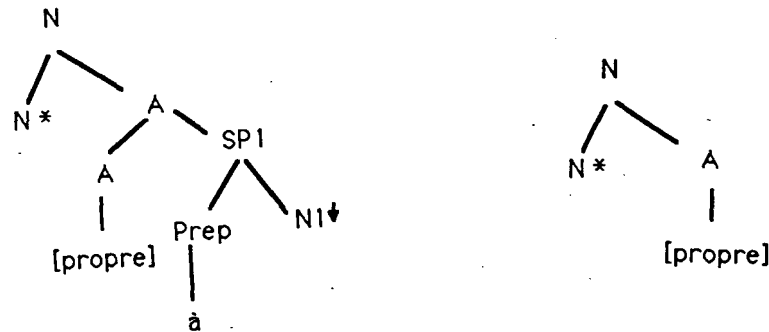


Figure 38

L'on peut donc rendre non ambigu (en lui assignant deux structures différentes) un adjectif comme *propre* qui, sans complément, signifie le contraire de *sale* et, avec un complément en *à*, signifie *spécifique* ou *approprié*. De même pour l'adjectif *soucieux* et son homonyme *soucieux de* Pin²⁶.

La "désambiguation" se fait également par les traits associés au nom modifié. L'adjectif *clair* modifiant un nom concret est l'opposé de *foncé*, mais est l'opposé de *obscur*, *complexe* lorsqu'il modifie un nom abstrait. Les contraintes sur le nom modifié (portées sur le noeud pied de la structure de l'adjectif) sont surtout sémantiques. Ainsi, un adjectif comme *violet* s'adjoit seulement à des noms concrets.

Certaines contraintes portent aussi sur la détermination du nom modifié. Le nom modifié doit être non déterminé, et le groupe Adjectif-Nom sera aussi non déterminé, sauf avec quelques adjectifs comme *semblable* ou *pareil* qui permettent des groupes nominaux sans déterminant:

25 Nous n'avons pas implémenté les adjectifs en apposition. Ils ont les mêmes structures élémentaires mais des traits sur N différents (<det> = +).

26 Quand un adjectif a la même sens (et est épithète des mêmes noms) avec et sans complément, il est représenté par une seule structure, avec un complément optionnel (e.g. *content* et *content de N*).

*Je n'ai jamais lu (*E + cette) aventure*
Je n'ai jamais lu (une + E) semblable aventure

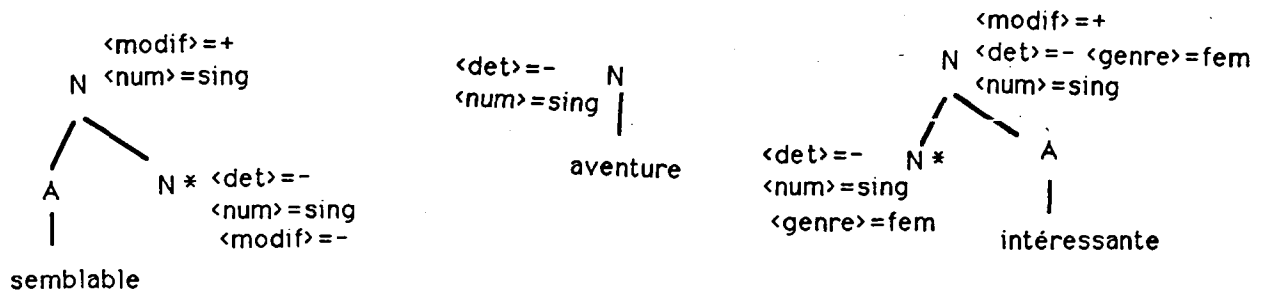


Figure 39

Les adjectifs, comme les autres modifieurs de noms, ajoutent au nom le trait $\langle \text{modif} \rangle = +$ qui interfère dans certains cas avec les déterminants possibles du groupe nominal (Cf supra).

Les contraintes sur les compléments des adjectifs s'expriment de la même façon que pour les compléments de noms ou de verbe. Ainsi *amoureux de N* prend en général un complément humain, tandis que le complément de *content de N* ne sera pas contraint.

Les adjectifs ont aussi des structures différentes selon qu'ils précèdent ou suivent le nom qu'ils modifient. L'on distinguera donc un *petit garçon* (=jeune) d'un *garçon petit* (petit de taille).

1.6.2. Adjectifs attributs

Les adjectifs attributs sont représentés par des arbres initiaux de racine adjectivale, substitués dans l'arbre phrastique du verbe principal. Ils doivent aussi comporter des traits sémantiques restreignant le type de noms dont ils peuvent être attributs. Selon le verbe une telle contrainte s'exerce sur le premier complément du verbe (N_0 *rendre* N_1 *A*) ou sur son sujet (N_0 *devenir* *A*).

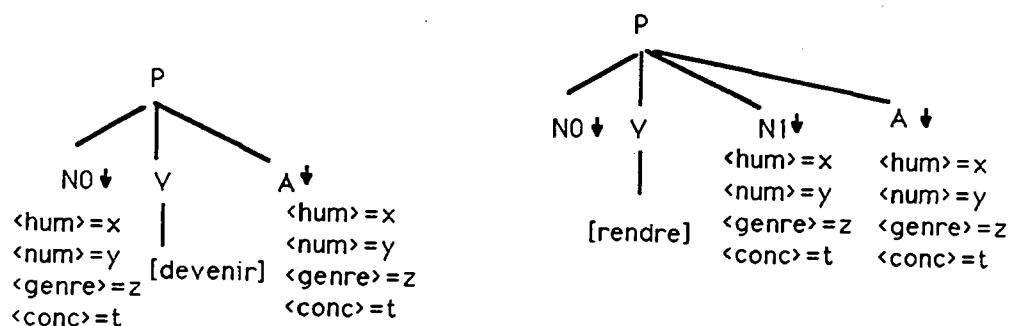


Figure 40

Les phrases copulatives sont représentées comme des structures phrastiques à tête multiples (*être* et l'adjectif) essentiellement à cause des possibilités d'extraction (voir chapitre 5)²⁷.

Tous les adjectifs n'entrent pas dans les deux types de structure. Les adjectifs "relationnels" par exemple, ne peuvent être utilisés comme arguments (ni entrer dans la construction copulative) et seront donc représentés uniquement par arbres auxiliaires²⁸.

1.7 Les adverbes

Les adverbes sont têtes d'arbres auxiliaires qui portent à leur racine la catégorie de ce qu'ils peuvent modifier. En théorie, toute catégorie a un adverbe qui lui correspond. Nous distinguons ainsi les adverbes selon qu'ils modifient un verbe (*rapidement*), une phrase (*probablement*), un autre adverbe, un adjectif (*très*), un nom (*hier*), ou un groupe prépositionnel.

On distingue ainsi deux adverbes *heureusement*: l'un qui modifie une phrase, l'autre qui modifie un verbe ou un adjectif. Une phrase comme *Jean est heureusement marié* sera ambiguë, selon que Jean est heureux dans son ménage ou qu'on juge avantageux le fait qu'il soit marié. Un adverbe comme *vraiment*, qui est très ambigu, modifie toutes les catégories (sauf les noms)²⁹. Les adverbes discontinus comme *ne...pas* se représentent de la même façon:

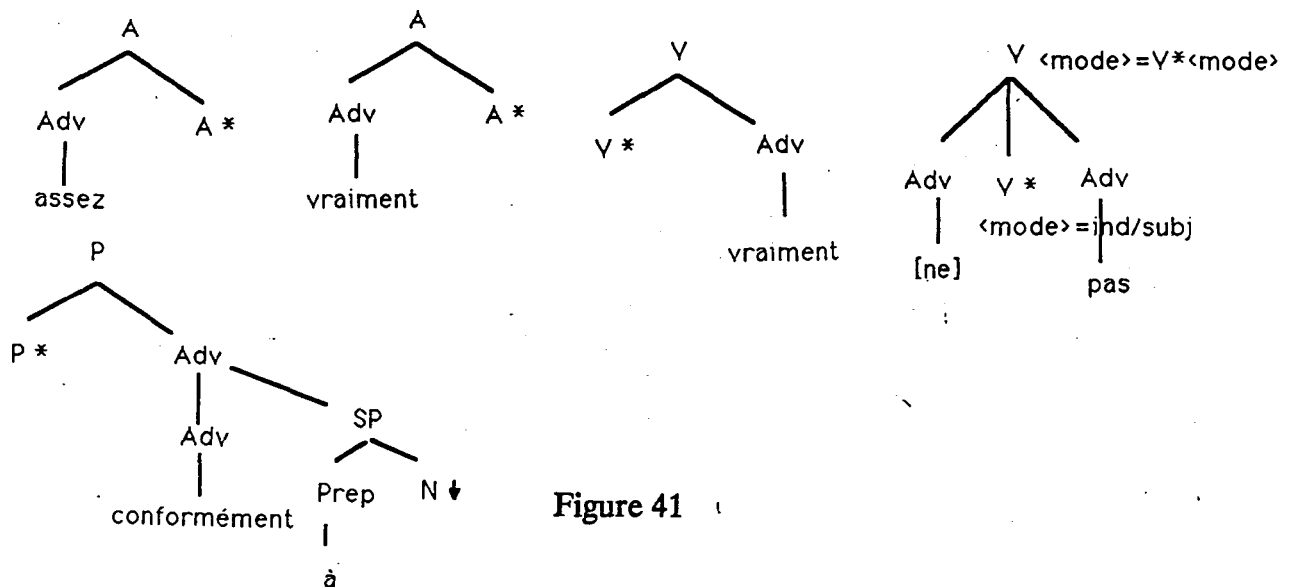


Figure 41

27 L'on pourrait de la même façon définir des "clausettes", c'est-à-dire des arbres initiaux de racine phrastique à tête adjectivale, pour tous les cas d'arguments adjectivaux, le verbe s'adjoignant (*sembler*, *rester*) ou prenant un complément phrastique (*croire*).

28 Ces adjectifs sont à traiter soit comme partie de noms composés de structure N Adj, soit comme la nominalisation d'une phrase à verbe support: *un voyage présidentiel* = *Le président fait un voyage* (A. Guyon 1991).

29 Quand un adverbe modifie plusieurs catégories sans être pour autant ambigu, on note une seule structure avec disjonction sur les catégories. Ainsi pour *assez* qui modifie des adjectifs ou des adverbes.

La structure argumentale des adverbes est représentée par des noeuds à substituer: conformément à *N*, contrairement à *N*, la catégorie du groupe étant toujours Adverbe.

Certains adverbes peuvent être arguments obligatoires de verbes ou de noms et sont alors représentés également par une structure initiale de catégorie Adv qui sera substituée dans celle de leur prédicat. C'est le cas en général des locatifs ou des adverbes de manière avec des verbes comme: *N₀ agit bien/comme il faut/de façon satisfaisante, N₀ mettre N₁ Loc2* (Loc=Adv/SP).

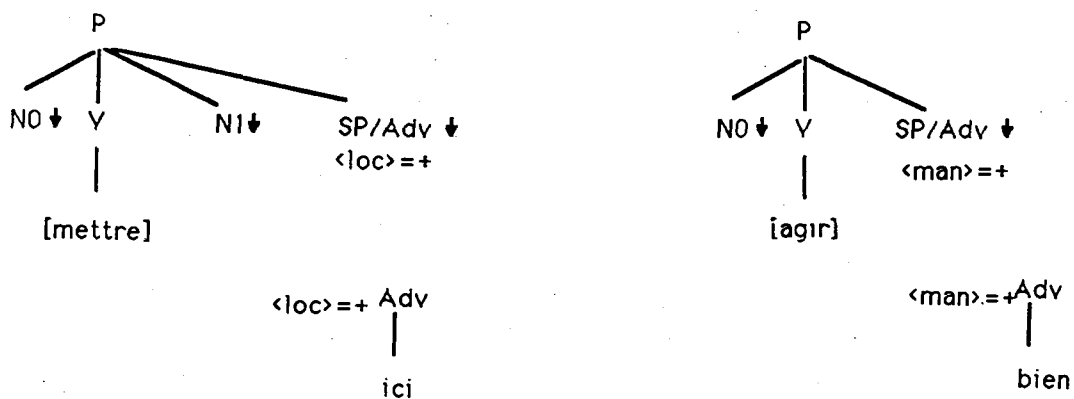


Figure 42

Nous n'avons pas abordé la question de la restriction entre adverbes et phrases. Le type de traits à retenir (temps, aspect...) est très vaste. Se pose aussi le problème de la localité de ces traits, puisque les adverbes (qui s'adjoignent à V ou à P) peuvent aussi imposer des contraintes aux arguments de la phrase. Par exemple, des adverbes comme *intelligemment* ou *soigneusement* supposent un sujet (ou un complément d'agent) humain; des adverbes comme *conjointement* ou *concurrentement* supposent dans la phrase au moins un argument pluriel (nombrable).

Nous envisageons d'ajouter des phrases à verbe support d'adverbe (M. Gross 1990), qui permettent d'éclairer ces contraintes, et qui seront représentées selon les mêmes principes que celles pour les noms prédicatifs (voir chapitre 4).

2. Dictionnaire morphologique et dictionnaire syntaxique

La grammaire TAG du français se présente comme un dictionnaire qui à chaque mot (ou groupe de mots) associe un arbre élémentaire ou une famille d'arbres élémentaires. Nous utilisons un dictionnaire de formes fléchies indépendant (sur le modèle du DELAF du LADL) qui regroupe les formes fléchies sous leurs lemmes, notés avec des crochets³⁰.

Une entrée du dictionnaire syntaxique (qui est le dictionnaire TAG proprement dit) se présente ainsi:

[lemme],X {traits} : arbres élémentaires {traits}.

où X est une catégorie préterminale.

exemple: [casser],V : n0Vn1{N1:<conc> = +}.

Si l'on veut tenir compte du fait que de nombreuses entrées lexicales comportent plusieurs mots (verbe et préposition par exemple), le schéma général est en fait:

[lemme1],X {traits};[lemme2],Y {traits}... : arbres élémentaires {traits}.

exemple: [penser],V;à,P : n0Vpn1{N0:<hum> = +, N1:<det> = +, V.t:<datif-h> = -}.

Les traits associés aux lemmes sont des traits morphologiques, les traits associés aux arbres sont syntaxiques et sémantiques.

Les mots qui n'ont pas de formes fléchies (par exemple les noms propres) apparaissent directement dans le dictionnaire syntaxique. Les autres y apparaissent par l'intermédiaire de leur lemme.

Un lemme donné a autant d'entrées dans le dictionnaire syntaxique qu'il entre dans des structures (familles d'arbres + traits syntaxiques et sémantiques) différentes ou appartient à des expressions figées différentes.

Le dictionnaire syntaxique permet donc de lever certaines ambiguïtés, tout en respectant l'unité morphologique des lemmes ambigus. En effet, la plupart des formes fléchies d'un adjectif ou d'un verbe, par exemple, se retrouvent quelle que soit sa structure syntaxique. Ainsi, les deux adjectifs [*propre*] et [*propre*] à N/Pinf, dont l'un est l'opposé de *sale* et l'autre signifie *spécifique*, ont le même ensemble de formes fléchies épiciques: *propres* et *propre*. Il y aura donc une seule entrée [*propre*] dans le dictionnaire morphologique, auquel correspondront deux entrées dans le dictionnaire syntaxique. De même un verbe comme [arriver] a toujours les mêmes formes, qu'il s'agisse de: N0

30 Au lieu d'un dictionnaire de formes fléchies, on pourrait utiliser un analyseur morphologique pour retrouver le (ou les) lemme(s) associé(s).

[arriver] Loc N_1 (verbe de mouvement), N_0 [arriver] à Pinf (=réussir) ou N_0 [arriver] à Nhum.

Mais, comme le note M. Gross 1989a, il y a également de nombreux exemples où propriétés morphologiques et propriétés syntaxiques interfèrent.

Pour les verbes, l'invariabilité du participe passé peut être une caractéristique de la construction intransitive d'un verbe (s'il prend *avoir* comme auxiliaire) par rapport à sa construction transitive. Par exemple, les formes *volée*, *volés* et *volées* ne peuvent correspondre qu' à N_0 *voler* N_1 à N_2 et non au verbe de mouvement N_0 *voler* Loc N_1 .

Pour les noms, le genre est le point commun qui permet de définir le lemme (et de distinguer deux lemmes pour *voile*, l'un masculin, l'autre féminin). Le nombre est souvent contraint par l'entrée syntaxique, notamment lorsque le nom appartient à une expression figée. Ainsi l'expression N_0 *prendre le taureau par les cornes* ne permet aucune variation: N_0 *prendre le taureau par la corne* ou N_0 *prendre les taureaux par les cornes* n'ont plus le même sens idiomatique. Néanmoins, afin d'incorporer le principe général selon lequel les expressions figées sont formées à partir du même matériel morphologique que les expressions libres, nous les avons représentées par les mêmes lemmes que celles-ci (avec des traits pour les contraindre) et non directement comme formes fléchies (Cf chapitre 4).

En théorie, il y a donc deux possibilités, dont aucune n'est entièrement satisfaisante: soit privilégier l'indépendance entre propriétés morphologiques et propriétés syntaxiques, et avoir deux dictionnaires (ou deux modules d'analyse) séparés; soit mettre l'accent sur les phénomènes de dépendance et utiliser uniquement un dictionnaire syntaxique qui aura comme entrées directement les formes fléchies.

Nous avons préféré avoir deux dictionnaires, quitte à regrouper dans l'étape morphologique des formes que l'analyse syntaxique amène ensuite à distinguer, et ce pour deux raisons. D'une part, cela permet, en prenant comme postulat de départ l'indépendance des propriétés morphologiques et syntaxiques, de cerner plus précisément les zones de dépendance, qui sont assez limitées. D'autre part, il était plus commode, pour la construction et la maintenance du dictionnaire TAG, de travailler sur des lemmes que sur les formes fléchies, ce qui permet un compactage important. Il est de plus possible de générer automatiquement un dictionnaire TAG de formes fléchies à partir de nos deux dictionnaires, alors que l'inverse ne serait pas possible. Il resterait donc à comparer les performances pratiques de l'analyseur dans les deux cas.

Voir des extraits des deux dictionnaires en annexe 3.

3. Les "familles" d'arbres

Nous n'avons pas de transformations au sens générativiste du terme. Il y a mise à plat et simple regroupement des structures résultantes dans des "familles" d'arbres. Ces familles d'arbres élémentaires définissent des classes d'équivalence³¹ au sens de Z. Harris 1976. Les familles d'arbres correspondent aux différentes structures syntaxiques pour un prédicat à structure argumentale donnée. Ces familles intègrent dans leur composition certaines transformations de la grammaire générative plus les relations (pas forcément transformationnelles) mises à jour au LADL pour les phrases simples du français, et toutes leurs combinaisons possibles: pour un verbe transitif, on aura ainsi un arbre pour le passif simple, un pour le passif avec question, un pour le passif avec extraposition... Ces familles sont définies pour les structures phrastiques à tête verbale (simples ou figées) ainsi que pour certains ajouts (tête adverbiale ou prépositionnelle).

Nous allons d'abord présenter le modèle formel que nous utilisons pour définir des liens entre arbres élémentaires appartenant à une même famille (et en déduire des dérivations de phrases sémantiquement équivalentes: les TAG "autosynchrones"). Puis, nous comparons notre représentation tant aux systèmes de métarègles des grammaires d'unification qu'aux représentations par tables du LADL dont nous nous sommes inspirée. Enfin, nous détaillons la composition des familles pour les verbes simples et certains ajouts.

3.1 TAG "auto-synchrones"

Les TAG synchrones (*synchronous TAG*) ont été définies comme moyen de relier deux "grammaires", et notamment une grammaire TAG et une représentation logique sous forme d'arbres sémantiques associés (Y. Schabes et S. Shieber 1990). Le même système peut être utilisé pour relier les grammaires TAG correspondant à deux langues différentes et nous explorons cette possibilité dans une perspective de traduction automatique (A. Abeillé et al. 1990). Mais il est également possible de considérer une TAG "auto-synchrone", c'est-à-dire reliée à elle-même. Nous examinerons ici cette représentation essentiellement du point de vue des règles syntaxiques (pour quelques développements concernant une utilisation plus sémantique de ce modèle, voir chapitres 4 et 7).

Il s'agit de pouvoir relier deux phrases possiblement complexes sur la base des liens existant entre les phrases dont elles sont issues, par exemple : *Jean observe la fille qui parle* et *La fille qui parle est observée par Jean*.

31 En fait, si l'on poussait davantage l'analyse sémantique, il faudrait considérer non seulement des relations symétriques entre arbres mais aussi des liens asymétriques (i.e. des liens orientés).

Comme il a été souvent remarqué dans la littérature transformationnelle, de *C'est Marie qui aime Luc*, on peut déduire *Marie aime Luc* mais non le contraire puisque la première phrase ajoute que seule Marie aime Luc.

Une dérivation "synchrone" procède comme suit. Considérant deux arbres (élémentaires ou dérivés) reliés entre eux (1 et 2), choisir un des liens existants entre eux (c'est-à-dire une paire de noeuds $a \in 1$ et $b \in 2$ reliés entre eux), choisir ensuite une deuxième paire d'arbres élémentaires 3 et 4 tels que leurs noeuds racines soient de la même catégorie que a et b respectivement, et combiner 3 avec 1 en a et 4 avec 2 en b :

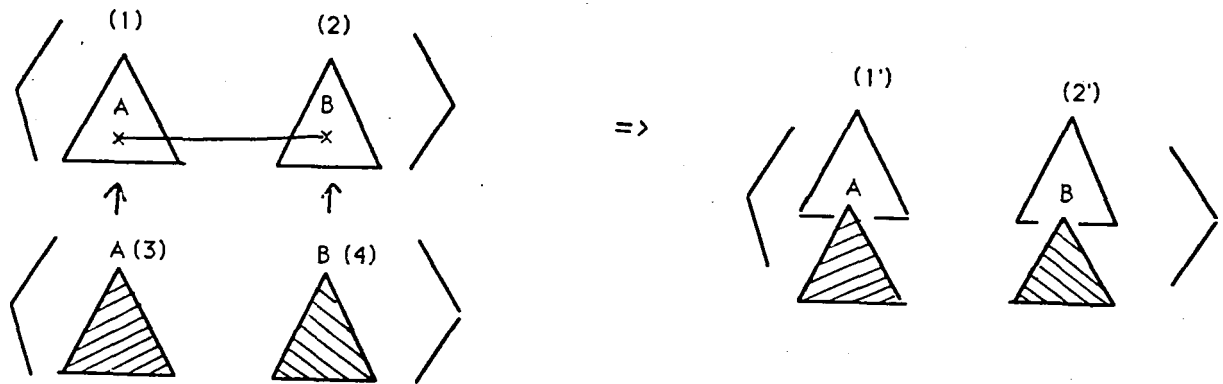


Figure 43

Une dérivation synchrone est une suite de telles dérivations. Les arbres 1, 2, 3 et 4 peuvent être initiaux ou auxiliaires. Il n'est pas nécessaire qu'ils soient de même type. Les paires d'arbres sont définies indépendamment des liens entre noeuds: les liens contraignent les adjonctions ou substitutions d'autres arbres avec la paire concernée.

Dans la définition originelle de la synchronicité, il s'agit de relier deux TAG quelconques. Dans notre optique, il s'agit de relier entre elles deux TAG exactement identiques. Deux arbres identiques (même topologie, même(s) tête(s) lexicale(s)) forment toujours une paire. De plus, leurs noeuds identiques (même catégorie, même position) sont toujours reliés. Mais l'on peut aussi définir des paires d'arbres différents. Des liens seront alors spécialement spécifiés entre eux. Par défaut, les noeuds de même catégorie en même position sont reliés. Les noeuds liés peuvent être de type différent: ainsi un noeud à substituer peut être lié à un noeud intérieur ou à un noeud racine: en cas de substitution à ce noeud, il devra y avoir adjonction au noeud correspondant. Lorsqu'aucune combinaison d'arbres ne peut advenir à un noeud (noeud racine ou intérieur avec contrainte d'adjonction nulle), ce noeud ne sera pas relié à d'autres.

Reprenons le cas du couple actif-passif déjà mentionné au chapitre 1. La phrase *Jean aime Marie* est analysée comme synchrone avec *Marie est aimée par Jean* car :

- les arbres élémentaires (1) (N_0 aimer N_1) et 2 (N_0 être aimé par N_1) sont reliés entre eux
- l'arbre de *Jean* forme paire avec lui-même et l'arbre de *Marie* aussi
- les deux combinaisons observées préservent ces liens.

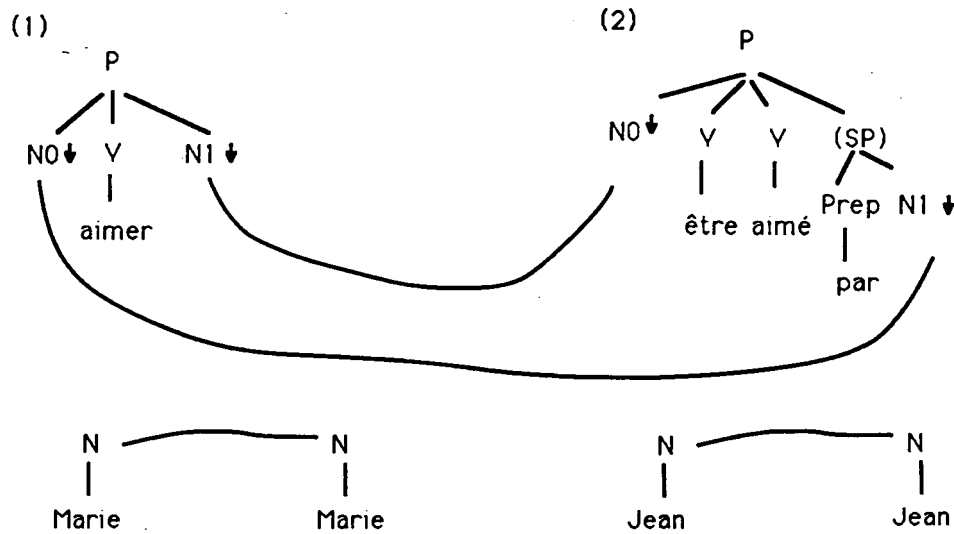


Figure 44

Il n'y a pas synchronie en revanche entre *Marie est aimée par Jean* et *Marie aime Jean*, puisqu'à des noeuds liés (N_1 dans 1 et N_0 dans 2) seraient substitués des arbres élémentaires non équivalents (*Jean* en 1 et *Marie* en 2). Seraient également analysées comme synchrones des phrases comme *Jean, que Paul déteste, aime Marie* et *Marie est aimée par Jean, que Paul déteste* ou *Jean, qui est détesté par Paul, aime Marie*.

Ce système permet de relier des phrases quelconques à partir de liens définis sur les arbres élémentaires (c'est-à-dire les entrées lexicales) dont elles sont formées. Les relations transformationnelles entre phrases peuvent donc être exprimées au niveau des arbres élémentaires (et non comme en grammaire générative au niveau des arbres phrastiques dérivés).

Le fait qu'il y ait un lien entre eux n'empêche pas des noeuds de porter des traits éventuellement différents. Il est clair qu'un SN foyer de clivé ou questionné est plus contraint qu'un SN complément en général. Nous aurons donc un trait $\langle \text{qu} \rangle = +$ dans l'arbre de la question tandis que ce trait sera en général non spécifié aux autres noeuds SN. Bien entendu, une phrase dérivée de l'arbre de la question ne sera synchrone qu'avec une phrase comportant un SN $\langle \text{qu} \rangle = +$ dans une position liée.

A quelle femme pense Jean ?

Jean pense à quelle femme ?

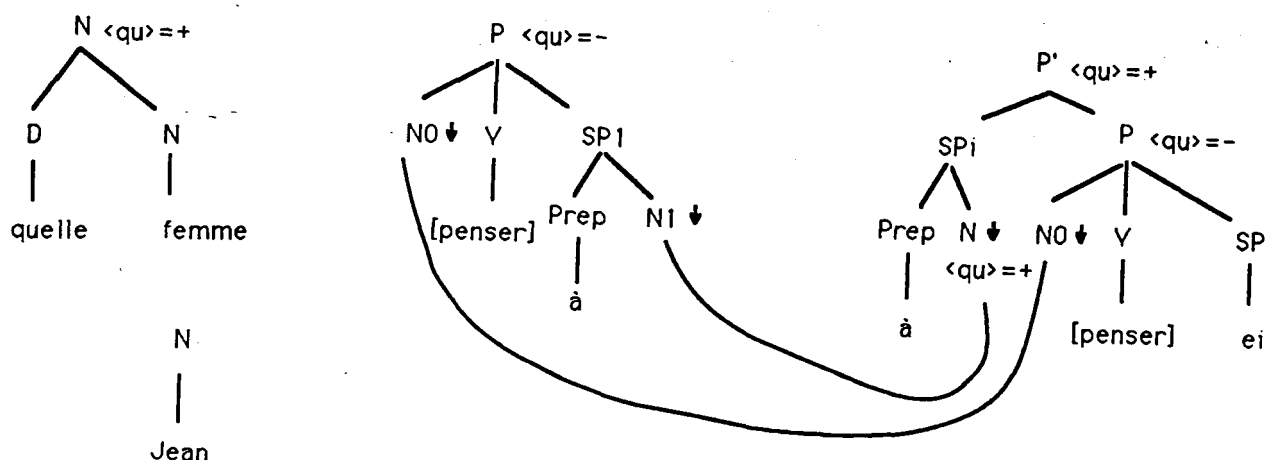


Figure 45

Par défaut, deux noeuds reliés partagent les mêmes traits, et il suffit d'indiquer les restrictions de sélection sur un arbre de la famille (par défaut, celui de la phrase déclarative active) dans l'entrée lexicale du verbe.

De même, un noeud peut être optionnel dans un arbre alors que le noeud auquel il est relié dans un autre arbre est, lui, obligatoire. Ces deux arbres ne donneront lieu à des dérivations synchrones que si le noeud optionnel est bien réalisé. Ainsi, *Marie est aimée* ne sera reliée à rien, seule *Marie est aimée par Jean* pourra l'être³².

Il y a des arbres qui, tout en étant reliés à d'autres, ne permettront jamais, en l'état actuel des choses, de dérivations synchrones. C'est le cas par exemple des arbres "neutres" dans les familles de verbes transitifs, puisqu'aucun arbre de la même famille n'a un sujet optionnel: *La branche casse* ne peut être dérivé de façon synchrone avec: *(*E + Jean) casse la branche*.

Nous avons donc redéfini nos familles d'arbres sur ce modèle, c'est-à-dire comme des classes d'arbres élémentaires avec des liens explicites entre leurs noeuds. A partir d'une paire quelconque d'arbres de la même famille, on pourra en général trouver une dérivation synchrone.

Au départ, nous avons des classes disjointes d'arbres élémentaires liés entre eux implicitement par coindexation de certains de leurs noeuds (A. Abeillé 1988b).

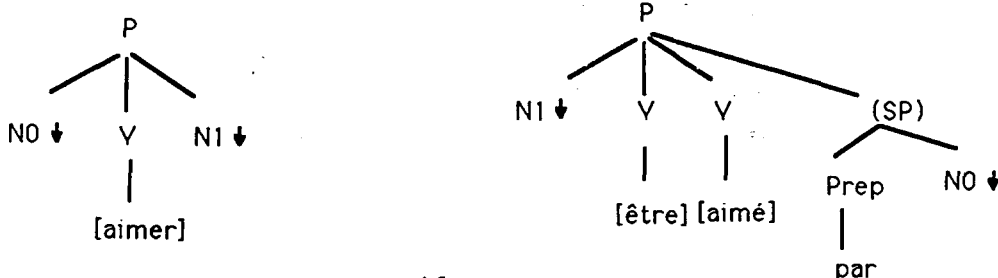


Figure 46

³² La phrase *Marie est aimée* n'est donc pas reliée à *(On + Quelqu'un) aime Marie*. Dans la définition "stricte" de la synchronicité, chaque dérivation doit comprendre le même nombre d'arbres élémentaires. Pour un assouplissement possible de la notion, voir chapitre 7.

Il y avait plusieurs inconvénients à cela. Il n'y avait pas de moyens de relier directement les phrases résultantes, puisque leurs arbres dérivés et leurs arbres de dérivation étaient fort différents (le seul moyen aurait été de passer par une représentation abstraite).

Il fallait aussi indexer N_1 le sujet du passif et du neutre, ce qui équivalait à une règle orientée et obligeait à avoir des traits supplémentaires pour la fonction sujet par exemple (pour avoir *C'est Marie qui est aimée* et non * *C'est Marie que est aimée*, ou la bonne forme du clitique). Avec les liens, l'indice de tous les sujets peut être le même N_0 , et cela suffit à les définir.

Désormais, plusieurs arbres élémentaires (par exemple, l'arbre intransitif pour les verbes intransitifs et les neutres) peuvent également être partagés par la même famille, ce qui est économique et justifié du point de vue linguistique³³.

3.2. Comparaison avec les métarègles ou règles lexicales de GPSG, HPSG et LFG

3.2.1 Contre la distinction règle syntaxique/ règle lexicale

La plupart des théories distinguent depuis N. Chomsky 1971 et T. Wasow 1977 entre règles syntaxiques et règles lexicales: N. Chomsky 1971 distinguait ainsi deux types de nominalisation et T. Wasow 1977 deux types de passif.

Les premières sont supposées s'appliquer régulièrement et ne changer ni le sens de l'expression ni la sous-catégorisation du verbe; tandis que les secondes sont soumises à des restrictions lexico-sémantiques nombreuses et peuvent altérer le sens ou la sous-catégorisation du verbe. Cette distinction reprend en partie des distinctions plus anciennes: entre transformations cycliques et non cycliques, entre règles sémantiquement gouvernées ou non. La plupart des discussions visent à savoir si telle règle (le passif par exemple) appartient à un groupe ou à un autre. Mais c'est le principe même d'une telle distinction binaire qu'il convient de mettre en question.

Certes, certaines règles sont plus locales que d'autres, certaines donnent naissance à des dépendances non bornées, d'autres non, certaines "laissent des traces", d'autres non. Mais ces distinctions n'isolent pas le même sous-ensemble de règles à chaque fois et il n'y a pas de raisons de privilégier un critère de distinction plutôt qu'un autre. De plus, toutes les règles, si l'on y regarde de près et si l'on prend en compte leur application aux expressions idiomatiques, sont soumises à des restrictions lexico-sémantiques.

Des grammaires comme LFG ou GPSG doivent adopter une représentation différente pour les règles lexicales et les règles syntaxiques, essentiellement parce qu'au niveau des

33 D'autres arguments nous ont aussi été fournis par notre travail parallèle sur l'anglais (A. Abeillé et al. 1990). Il semble peu justifié de représenter par coindexation "dative-shift" ou "particle hopping". Pour dative-shift, à côté des verbes comme *give* qui ont les deux structures il existe de nombreux verbes qui ont seulement la construction à complément *to* (*explain, donate...*) et de nombreux autres qui ont seulement la construction à deux SN (*spare, cost, envy...*). Il est impossible de dire quelle est la forme de base et lequel des N_i sera N_1 lequel sera N_2 (La situation est la même pour les verbes transitifs à particule, où il ne serait pas plus justifié de prendre le cas où la particule suit le verbe que celui où elle en est séparée par le nom pour forme de base).

items lexicaux elles n'ont pas accès à des structures syntaxiques complètes. Ainsi LFG et GPSG représentent le passif comme une règle "lexicale" reliant deux "formes lexicales" ou deux règles de réécriture à tête verbale respectivement, tandis que la question est représentée comme une règle de réécriture du noeud phrase, sans lien direct avec la tête verbale.

Ce type de distinction est inutile en TAG, puisque les structures de base sur lesquelles opèrent les règles sont toujours lexicales.

Nous regroupons actuellement dans nos familles la plupart des règles lexicales et syntaxiques de ces formalismes, sauf certaines règles lexicales de LFG qui changent la catégorie syntaxique (règles dérivationnelles), mais il serait possible de les intégrer également (Cf chapitre 7).

3.2.2. Sur la notion de métarègle

En GPSG, la plupart des règles lexicales sont traitées par un système de métarègles. Celles-ci ont pour fonction d'alléger le composant syntagmatique de la grammaire et d'exprimer à un niveau de généralité supérieur les conditions de bonne formation des structures de la langue.

Elles se distinguent des transformations des grammaires chomskyennes en ce que ces dernières portent sur des structures phrastiques complètes (de surface ou "dérivées"), alors que les premières permettent simplement de générer une règle de réécriture (un sous-arbre de profondeur 1) à partir d'une autre. Par exemple, la métarègle du passif s'écrit comme suit en GPSG:

VP -> V NP X => VP -> V[passif] X (PP[par])

X étant une métavariable, la métarègle condense les règles suivantes:

VP -> V NP => VP -> V[passif] (PP[par])
 VP -> V NP PP => VP -> V[passif] PP (PP[par])
 VP -> V NP AP => VP -> V[passif] AP (PP[par])

.....

T. Becker 1990 examine la définition de métarègles pour les TAG, essentiellement sous l'angle de leur capacité générative. La plupart des autres formalismes (GPSG, grammaires catégorielles) voient leur capacité générative augmenter de façon incontrôlée à partir du moment où l'on introduit des métarègles. Diverses techniques ont été proposées pour y remédier, mais aucune n'est linguistiquement satisfaisante³⁴ (Cf S. Shieber et al. 1983). Pour les TAG, la solution semble plus simple: les unités de départ (les arbres élémentaires) sont toujours de taille bornée. Les métavariables utilisées pour définir le schéma d'arbres sur lequel s'applique telle ou telle métarègle seront donc bornées. Le nombre d'arguments de chaque verbe (c'est-à-dire le nombre

34 Par exemple l'application unique de chaque règle en GPSG se heurte au fait que la dislocation peut concerner plusieurs éléments en français:

Mon vélo, Jean, je le lui ai montré.

De même pour les langues à questions multiples.

maximum d'utilisations d'une règle donnée) est également fini. La capacité générative des TAG avec métarègles devrait donc rester la même.

La métarègle du passif serait par exemple (avec X et W comme métavariabes):

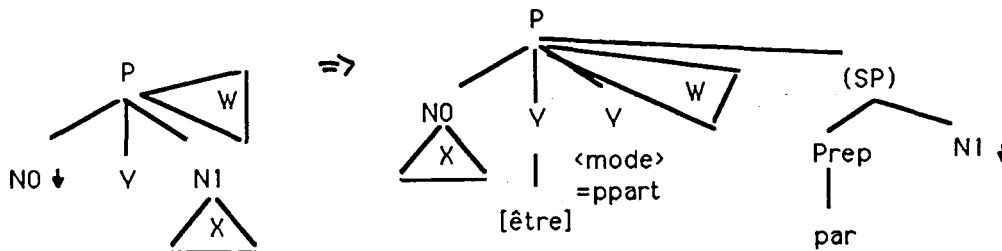


Figure 47

Il est intéressant de noter que ce type de métarègles s'applique sur des domaines plus étendus que des sous-arbres de profondeur 1 (que sont les règles de réécriture de GPSG) et permet donc davantage de restructurations que les métarègles des autres formalismes.

Ces métarègles pourraient également être utilisées pour relier des entrées morphologiquement apparentées, puisque le résultat est du même ordre (arbres élémentaires reliés par des liens; voir chapitre 7).

Formellement, il n'y aurait donc pas de difficulté à définir seulement certains arbres élémentaires (ceux correspondant aux principaux cadres de sous-catégorisation) et à déduire les autres par métarègles. Certains problèmes linguistiques se posent néanmoins.

La notion même de métarègle suppose qu'il s'agit de règles plus générales que les règles de réécriture indiquant par exemple les différents cadres de sous-catégorisation. Or, si les principales règles syntaxiques supposent qu'on regroupe les différents types de phrases simples auxquelles elles s'appliquent, elles nécessitent également des distinctions supplémentaires, en tant qu'elles peuvent être sensibles à la valeur des têtes lexicales en jeu. On peut donc se demander si les métarègles devraient être notées directement dans le lexique (c'est-à-dire "activées" seulement par les verbes concernés), ou bien rester à un niveau plus abstrait d'organisation des familles, les arbres élémentaires ainsi produits étant dans un second temps sélectionnés par les bons verbes. D'autre part, certains verbes (ou certaines phrases figées) n'existent qu'au passif, ou qu'à la forme interrogative. Il faudrait donc que les métarègles ne soient pas orientées, ce qui semble contradictoire, ou qu'il soit possible aux entrées lexicales considérées de sélectionner directement les arbres élémentaires (de base ou produits par métarègle) qui leur correspondent. Mais la notion même de métarègle perd alors beaucoup de son sens.

Enfin, les travaux du LADL ont montré l'importance et la productivité lexicale de certaines relations entre phrases simples qui se laissent malaisément représenter en

termes transformationnels. Par exemple, les relations qui définissent les classes 12 et 34LO respectivement:

(classe 12)

N_0 V que P := *Jean approuve que Marie ait fait cela*

N_0 V N_1 de Vinf W := *Jean approuve Marie d'avoir fait cela*

(classe 34LO)

N_0 V de N_1 := *Le jardin grouille d'abeilles*

N_1 V Loc N_0 := *Les abeilles grouillent dans le jardin*

Leur représentation par métarègle n'a pas de sens, puisque, par définition, un seul schéma de phrase simple est concerné.

Nous avons donc préféré considérer tous les arbres phrastiques élémentaires comme les arbres de base de la grammaire et nous les avons regroupés en classes d'équivalence. Cette méthode n'évite pas certaines redondances (ainsi nous avons plusieurs dizaines d'arbres de phrases passives) mais elle a la souplesse nécessaire en pratique.

3.3. Lien avec les classes du LADL

Nous nous appuyons sur les informations fournies par les tables du LADL pour le français.

Nous avons pour l'instant moins de familles qu'il n'y a de classes au LADL. Cela tient en partie au fait que nous avons codé moins d'informations, en partie au fait que nous codons certaines informations différemment.

Les tables du LADL sont globalement fondées sur les cadres de sous-catégorisation (avec en général des distinctions selon la valeur de la préposition) et éventuellement certaines sous-distinctions selon les restrictions de sélection (Cf les différentes tables 32).

De nombreuses tables correspondent aussi à une ou plusieurs propriétés syntaxiques régulières³⁵. En cas de conflit, c'est généralement le critère de la propriété syntaxique qui l'emporte. Ainsi, la table 5 regroupe les verbes à complétive sujet et complément prépositionnel: *Que P V Prep N₁*. La propriété associée est l'extraposition de la complétive (contrairement à la table 4 à complétive sujet et verbe transitif). Les quelques verbes transitifs à complétive sujet (par exemple *démanger* dans *Faire cela Luc* = *Il démange Luc de faire cela*) sont donc classés en table 5 avec Prep=0, et non en 4 (ce qui obligerait à une colonne pour la structure extraposée peu motivée pour cette table). Il est clair qu'une telle démarche relève surtout d'une position théorique. Dans un but de reconnaissance automatique, et avec le souci de ne pas multiplier les éléments vides qui rendent le traitement coûteux, c'est le principe de sous-catégorisation qui passe pour nous en premier.

Nous regroupons certaines de ces propriétés et nous en séparons d'autres. Certaines de ces informations font partie de l'entrée lexicale de l'expression, d'autres du nom de la famille, d'autres sont sous formes de traits.

³⁵ Pour certaines tables, il y a également homogénéité sémantique, mais ceci n'est pas un critère a priori de classification.

Ainsi, pour nous, les prépositions (sauf les locatives), avec leur valeur lexicale, font partie de l'entrée lexicale du verbe, tandis qu'elles peuvent donner lieu à des tables différentes (tables 7 et 8, par exemple, qui distinguent $N_0 V$ à ce que P et $N_0 V$ de ce que P), ou bien apparaître comme propriétés d'un verbe (en colonnes; tables 5,10..).

Les traits sémantiques des arguments (eg. objet humain ou concret) ne donnent pas lieu à des familles différentes mais à des traits différents associés à une même famille, ce qui est une simple variation de notation. Ainsi notre famille $n_0 V n_1$ regroupe les tables 32NM (N_1 = nom de mesure), 32C (N_1 = Nconcret), 32H (N_1 = Nhumain) etc...

L'optionalité d'un complément est notée dans les tables par une colonne (sous structure: $N_0 V$). Nous marquons cette propriété directement dans le nom de la famille. Ainsi *manger* a la famille $n_0 V(n_1)$ et *briser* la famille $n_0 V n_1$ (parmi d'autres).

Lorsque des propriétés syntaxiques sont régulièrement associées à ces tables, elles sont exprimées par un principe de cooccurrence de traits dans le lexique:

par exemple, la table 32NM regroupe des verbes dont les compléments sont des noms de mesure, et pour lesquels le passif est interdit. Nous aurons donc: $N_1: < \text{quant} > = + \Rightarrow \text{passif} = -$. Ce qui signifie que tous les verbes qui ont obligatoirement un complément de mesure n'ont pas le passif; on ne pourrait coder l'information au niveau du SN dans l'arbre du passif car la contrainte porte sur le verbe et non sur le seul complément:

La pendule coûte 10 francs
** 10 francs sont coûtés par la pendule*
Jean a volé 10 francs
10 francs ont été volés par Jean

Mais la table 32NM regroupe également tous les verbes transitifs sans passif (qui sont peu nombreux), par exemple *sentir* ou *regarder* dans *Cette pièce sent la rose* ou *Cette affaire regarde Jean*. Pour ces verbes, il faudra donc un trait passif explicite (voir section suivante).

De manière générale, on dispose de plusieurs moyens pour organiser les propriétés syntaxiques (i.e. obtenir les mêmes généralisations que la notion de classe), qui sont complémentaires:

Première possibilité: la composition même des familles indique quelles sont les propriétés possibles pour une structure donnée. Ainsi, les arbres correspondant à l'extraposition n'existeront que pour les familles de construction intransitive. Ce qui encode (implicitement) l'information que l'extraposition est interdite pour les verbes transitifs.

Deuxième possibilité : la restriction est suffisamment générale pour être codée directement (sous forme de traits) sur les noeuds feuilles des arbres élémentaires concernés. Par exemple, les traits sémantiques sur arguments pour le couplage obligatoire des arguments dans la structure symétrique: $N_0 V$ avec $N_1 = N_0$ et $N_1 V$.

Troisième possibilité: établir des relations d'implication entre traits (à la manière des relations de cooccurrence de traits de GPSG). Il peut s'agir de traits hétérogènes (règlant la combinaison des arbres élémentaires du verbe avec ceux de ses arguments, ou la composition de la famille d'arbres élémentaires associés au verbe), mais il faut qu'ils soient accessibles dans le lexique. Si la relation est générale, on peut poser un principe général, par exemple: $N_1: <quant> = + \Rightarrow passif = -$.

Sinon, on peut noter la relation au niveau des entrées lexicales concernées mais il n'est alors plus question de généralisation.

Quatrième possibilité: introduire des familles supplémentaires

On pourrait remplacer des rapports d'implication entre traits par des noms de sous-familles particulières. Cela peut revenir au même mais faciliterait en fait l'analyse automatique (qui peut être handicapée par l'usage de la disjonction ou de l'implication sur les structures de traits; Cf R. Kasper et W. Rounds 1986). Pour des informations que nous ne codons pas sous forme de traits, c'est sans doute la meilleure possibilité (sauf à introduire des traits supplémentaires). On pourrait ainsi introduire des sous-familles selon la valeur de certaines têtes lexicales, par exemple définir une sous-famille $Prep = \dot{a}$ dans la famille n_0Vpn_1 (puisque c'est la seule pour laquelle le trait datif est pertinent) ou une sous-famille $Clit = se$ dans les familles $n_0CIV...$

Il n'y aurait pas d'obstacle en fait à subdiviser certaines de nos familles pour se rapprocher des classes du LADL et ce serait sans doute plus économique si l'on voulait transférer la totalité des lexiques-grammaires dans notre grammaire.

Une autre différence avec la représentation du LADL est la suivante. Dans les lexiques-grammaires un couple verbe-table définit une entrée lexicale et une seule, alors qu'une même entrée lexicale en TAG peut se composer d'un verbe et de plusieurs familles. C'est le cas quand il n'y a pas de relations systématiques entre deux cadres de sous-catégorisation mais que sémantiquement il s'agit d'un seul et même verbe, par exemple *refuser* qui entre dans n_0Vn_1 et n_0Vs1 . L'on pourrait regrouper les familles qui sont ainsi associées, et considérer la construction $N_0 V N_1$ comme une sous-structure de $N_0 V P_1$ (ou vice versa) sélectionnée par les verbes adéquats par un trait *nominalisation*³⁶. Cependant, cela entraîne une grande complexification du système de traits puisque les N_0 admissibles en position sujet dans la construction à complément nominal ne seront pas forcément les mêmes, pour un verbe donné, que ceux admissibles comme sujets dans la construction à complément phrastique pour le même verbe. Nous avons donc préféré garder des familles différentes pour chaque cadre de sous-catégorisation, sauf dans le cas des verbes à sujets phrastiques auxquels sont régulièrement associés des sujets nominaux abstraits. Cette convention permet de ne pas créer de fausses ambiguïtés entre différents emplois d'un verbe.

Nous allons maintenant reprendre en détail la répartition des arbres élémentaires par familles.

36 C'est ce qui est fait dans les tables du LADL quand une construction est "mise en colonne" (ie considérée comme une propriété supplémentaire éventuelle des items de la table en question).

3.4. Familles des verbes simples

Nous ne considérons ici que la composition des familles correspondant à des phrases libres. Pour la composition des familles des phrases à verbe support et expressions figées, voir chapitre suivant. Il est à noter que, bien qu'il s'agisse de phrases libres, de nombreux arbres élémentaires ont des têtes lexicales multiples, ce qui montre, s'il en était besoin, que les phénomènes non compositionnels vont bien au delà de ce que l'on appelle communément "expressions figées".

Les arbres peuvent donc avoir des éléments lexicaux qui leur sont propres, par exemple *qui* pour les structures relatives à pronom sujet. Ils peuvent aussi porter des contraintes de traits à certains de leurs noeuds, indépendamment des contraintes de traits "apportées" par les verbes qui les sélectionnent. Les fichiers d'arbres forment donc un "réservoir" de contraintes structurales, auxquelles s'ajoutent les contraintes lexicales venant des verbes.

Voir la liste des familles avec des exemples en annexe 2. Les noms des familles sont normalement transparents et suivent le schéma: $\text{arg}_i\text{Varg}_j$ (pour une famille de tête verbale à arguments i et j), selon l'ordre canonique des arguments par rapport au verbe. Nous notons *ni* les arguments nominaux, *si* les arguments phrastiques, *pni* les compléments prépositionnels, *Cl* les clitiqes figés et *ppi* les compléments locatifs (qui commutent avec des adverbes). Sont entre parenthèses les arguments optionnels. Conformément aux conventions du LADL, nous notons le sujet avec l'indice 0 et les compléments de 1 à 3. Les familles de verbes intransitifs sont par exemple: n_0V ou n_0Vpn_1 , celles de verbes transitifs par exemple: n_0Vn_1 , $n_0Vn_1(pn_2)$. Certaines familles en incluent d'autres: par exemple, un verbe transitif neutre inclura les structures intransitives correspondantes.

Une famille se définit comme le nombre maximal de structures élémentaires correspondant à un cadre de sous-catégorisation donné. Ce qui ne veut pas dire qu'un verbe donné entrera forcément dans toutes ces structures à la fois. Chaque structure (c'est-à-dire chaque arbre) élémentaire est définie, pour une famille donnée, par un ensemble de traits: question, passif, extraposition ... Le verbe concerné peut bloquer certaines structures si son entrée lexicale contraint la valeur de certains de ces traits (ex: *passif*=-). L'on pourrait attacher ces traits au noeud V des arbres élémentaires mais cela semble un peu artificiel. Nous avons donc préféré les considérer comme des traits de type différent, associés directement à chaque arbre pris comme un tout d'une part, et à l'entrée du verbe dans le dictionnaire (c'est-à-dire au nom de sa famille) d'autre part. Pour les différencier des traits lexicaux, on les note sans crochets. La plupart de ces traits ne concernent que certaines familles (ex: *passif* ou *symétrie*).

La convention pour marquer ces traits est différente dans le dictionnaire de ce qu'elle est pour les arbres. Quand un trait syntaxique a la valeur - dans le dictionnaire, la ou les structures portant ce trait avec la valeur + sont impossibles pour l'item lexical considéré. Lorsqu'il porte la valeur +, cela ne signifie pas que la ou les structures correspondantes sont possibles, cela veut dire qu'elles sont obligatoires. Quand une

structure est possible, on ne marque rien, ce qui permet par exemple à un verbe comme N_0 *aimer* N_1 autorisant le passif de sélectionner aussi bien les arbres actifs que passifs. La justification de cette convention vient de l'intégration des expressions figées, qui peuvent n'exister qu'au passif, ou seulement sous forme extraposée (voir chapitre 4).

Nous présentons ci-dessous les arbres élémentaires non par famille mais par phénomène syntaxique. Pour les listes correspondantes par famille, voir en annexe. En général, nous présentons les arbres correspondant à une seule transformation à la fois. Les structures intégrant plusieurs transformations, qui suivent les mêmes principes, sont en annexe. Nous présentons ici les arbres à tête verbale correspondant aux phrases simples, puis les arbres correspondant à certains ajouts (prépositions ou adverbess). Pour les familles des phrases à verbe support ou expressions figées, voir chapitre 4.

3.4.1. Le passif

Le passif est représenté par un arbre à tête multiples, certains éléments sémantiquement vides (*par*, *être*) étant déjà dans l'arbre³⁷. Plusieurs arbres ont le trait *passif* = +, par exemple:

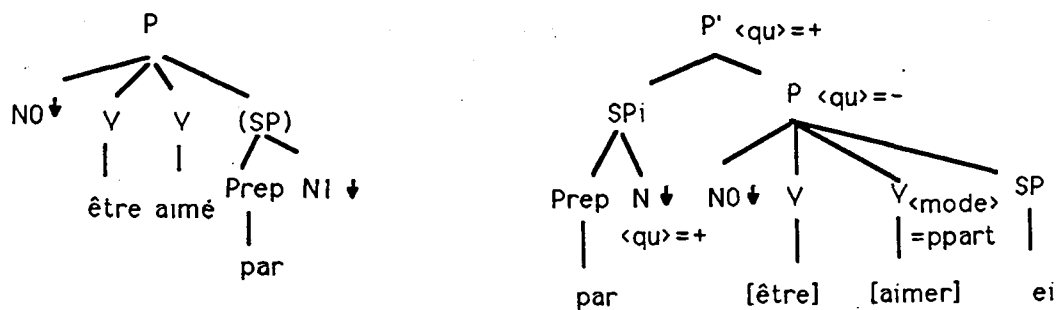


Figure 48

Le complément d'agent est optionnel. L'auxiliaire *être* pourrait l'être aussi, si l'on veut pouvoir obtenir à partir de l'arbre passif des phrases enchassées réduites telles que: *Paul croyait Marie aimée*.

Les familles suivantes sont concernées: n_0Vn_1 (*Jean aime Marie*), $n_0Vn_1pn_2$ (*Jean donne un livre à Marie*), n_0Vs_1 (*Jean a découvert que Jacques est un génie*) et $n_0Vs_1pn_2$ (*Jean affirme à Marie que Paul est stupide*).

Notons aussi le cas des passifs auxiliaires, qui apparaissent avec certains verbes à complément phrastique: N_0 *supposer que* $P = N_1$ *être supposé* V_{inf} , N_0 *dire que* $P = N_1$ *être dit* V_{inf} :

On suppose que Jean partira demain
= *Jean est supposé partir demain*

37 Nous ne traitons pas les cas de passif en *de*.

On dit que Jean a fait cela
= Jean est dit avoir fait cela

Le verbe *être censé* est un cas où le passif auxiliaire est la seule construction possible. Il s'agit d'un arbre particulier qui a le trait *pass-aux* et qui appartient à la famille n_0V_{s1} ³⁸:

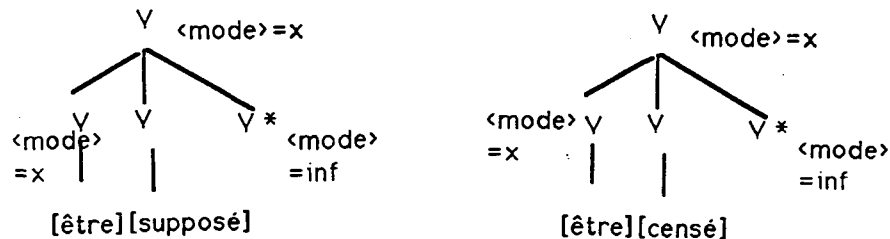


Figure 49

3.4.2. Les arbres questions

Les questions à élément interrogatif antéposé sont représentées par des arbres élémentaires particuliers, lorsqu'il y a "mouvement" d'un complément. Pour une structure donnée, on aura donc autant d'arbres questions qu'il y a de compléments de type SN (non dominés par un SP) ou SP³⁹.

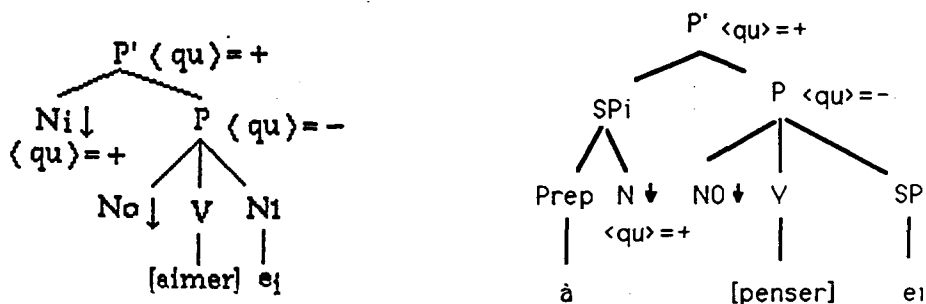


Figure 50

L'arbre correspondant à la question sur le sujet n'est pas différent de l'arbre déclaratif, car nous ne prenons pas en compte, du moins pour l'instant, les phénomènes

38 Les deux passifs peuvent exister pour un même verbe:

Il a été dit que Marie reviendrait demain. (extrap=+, passif=+, pass-aux=-)

Jean est dit être rentré hier. (extrap=-, passif=-, pass-aux=+)

39 La trace correspondant au syntagme déplacé ne sert pas, comme en grammaire générative, à interdire l'insertion d'un complément (puisque les compléments sont toujours déjà prévus dans l'arbre élémentaire). Elle ne pourrait servir qu'à autoriser des adjonctions (par exemple des relatives extraposées ou des déterminants complexes):

Quel homme as-tu vu qui est parti ?

De quel gâteau as-tu mangé la moitié ?

Dans l'état actuel de la grammaire, toute adjonction est interdite sur ces noeuds et ils pourraient donc être omis. Il en est de même pour les traces dans les structures relatives et clivées.

d'extraction qui sont particuliers au sujet et conduisent souvent à des phrases d'acceptabilité limitée⁴⁰: ?? *Qui croyez vous qui aime Marie ?* Nous n'avons donc pas besoin d'un arbre supplémentaire à noeud P interne comme pour les autres arguments (voir chapitre 5).

Les questions à pronom interrogatif (*qui, que, quoi*) sont générées à partir des mêmes structures que celles à SN interrogatif (*quel N, la fille de quel N*), puisqu'il y a propagation du trait <qu> dans le SN.

Il ne peut donc y avoir d'arbres correspondant à une double question avant: * *Qui quels livres a lu ?* ni * *Quels livres qui a lu ?*, puisque N₀ porte le trait <qu> =- dans la structure. Mais *Qui a lu quel livre ?* est normalement généré.

Nous utilisons les mêmes arbres élémentaires pour les questions directes et indirectes. Certaines questions en *est-ce que* sont également générées à partir des mêmes arbres, par adjonction de ce groupe figé au noeud phrastique intérieur. Les autres questions en *est-ce que* sont générées à partir des arbres déclaratifs par adjonction du groupe à la racine phrastique. Voir plus de détails au chapitre 5, en particulier concernant les contraintes d'inversion.

Les questions portant non sur des arguments mais sur des ajouts ne donnent pas lieu à des arbres spécifiques: elles sont obtenues par adjonction d'un ajout *qu-* sur l'arbre déclaratif (voir chapitre 5 pour plus de détails sur les extractions d'ajouts).

Les questions ne comportent pas de restrictions lexicales particulières, sauf pour certains verbes à complément infinitif et sujet humain, pour lesquels il faut noter une interdiction (par exemple avec N₀ *se dépêcher de* *Pinf* =: *Jean se dépêche de partir* = * *De quoi se dépêche Jean ?*; Cf supra, chapitre 2).

3.4.3. Les relatives

Les relatives sont représentées par des arbres auxiliaires à double tête: le pronom relatif et le verbe dont il est l'argument.

Il s'agit d'arbres auxiliaires de type nominal (racine N), même si le complément relativisé est de type prépositionnel, dans la mesure où l'antécédent est toujours nominal. C'est d'ailleurs là un argument supplémentaire pour nos SP étendus.

⁴⁰ L'acceptabilité de ces phrases varie selon les locuteurs: elle semble meilleure à des locuteurs très conservateurs ou au contraire en français populaire. Elle dépend aussi du verbe de la principale (Cf D. Godard 1980b).

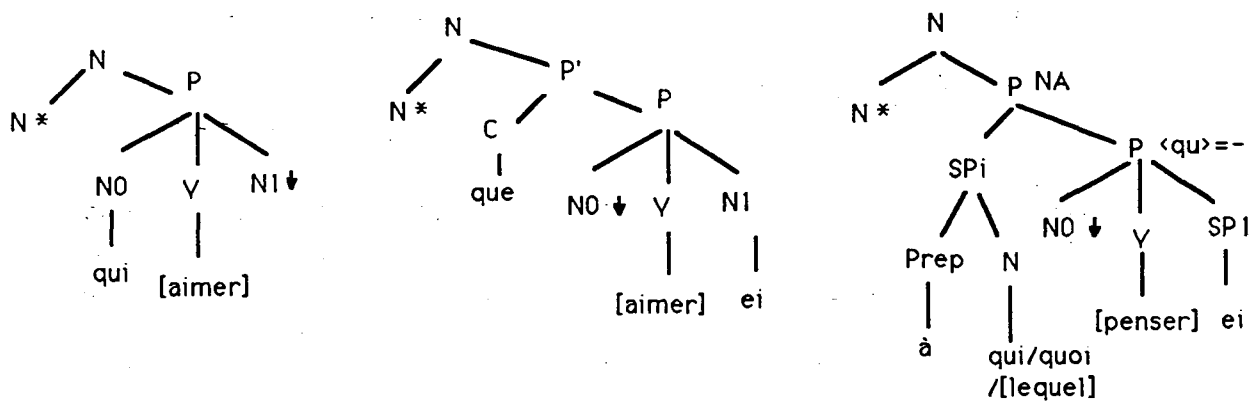
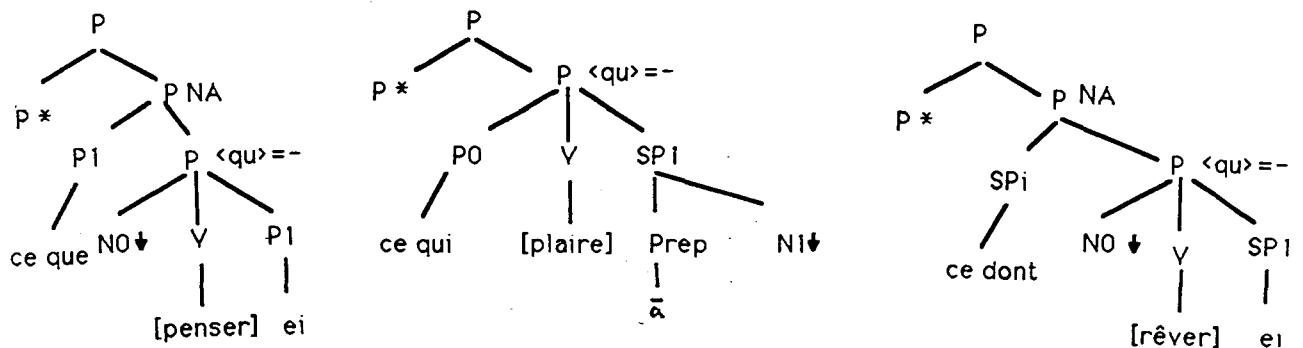


Figure 51

Nous ne traitons pas de l'enchâssement possible du pronom *lequel* dans les compléments prépositionnels, comme par exemple: *Je connais la personne à la fille de laquelle tu penses*. Pour ce faire, il faudrait (de la même façon que pour les questions) considérer que, dans ces arbres, l'élément relatif peut être substitué sous forme d'un pronom ou d'un SN relatif. Il faudrait donc ajouter un trait <rel> propagé à l'intérieur du SN et distinguer *qui* et *quoi* relatifs et interrogatifs. le même traitement pourrait être utilisé pour les relatives appositives sujets en *lequel N*.

Pour les verbes à argument phrastique, les relatives correspondantes sont introduites par *ce qui*, *ce que*, *ce dont*, *ce Prep quoi* ... que nous considérons comme des pronoms relatifs. L'antécédent de ces relatives étant une phrase, ce sont des arbres auxiliaires de racine phrastique:



Que P amuser N₁
 =: Que Paul soit parti amuse Pierre.
 Paul est parti, ce qui amuse Pierre.

N₀ rêver de ce que P
 Jean rêve de ce que Marie parte.
 Marie est partie, ce dont rêvait Jean.

Figure 52

Il y a partage de la plupart des traits (<hum>, <num>, <genre> ...) entre l'argument relativisé du verbe (c'est-à-dire le pronom relatif) et son antécédent nominal. En revanche, quand l'antécédent est une phrase, tous les traits ne sont pas partagés, le

mode par exemple peut être différent. Ainsi le complément de *craindre* doit être au subjonctif, alors que son antécédent n'est pas forcément contraint quant à son mode:

Je craignais que Paul soit venu
Jean est-venu, ce que je craignais

Les relatives à antécédent phrastique sont toujours appositives. Pour les relatives à antécédent nominal, nous ne faisons pas de différence structurale entre relatives restrictives et relatives appositives. Cette distinction devrait être déduite des traits du N modifié.

Nous ne traitons pas les "relatives sans antécédents" (*free relatives*) du type : *J'aime ce que tu as acheté*. Elles seraient représentées par des arbres initiaux de racine N à tête verbale, et ajoutées dans la famille du verbe correspondant (ici N_0 acheter N_1).

3.4.4 Les phrases clivées

Les phrases clivées sont représentées par des arbres élémentaires particuliers, à tête multiple: le verbe principal, *c'est* et *que, qui* ou *quoi* (ou *lequel*) selon que le foyer est sujet ou complément⁴¹. Pour une structure verbale donnée, on aura donc autant d'arbres de ce type que le verbe a d'arguments SN ou SP⁴².

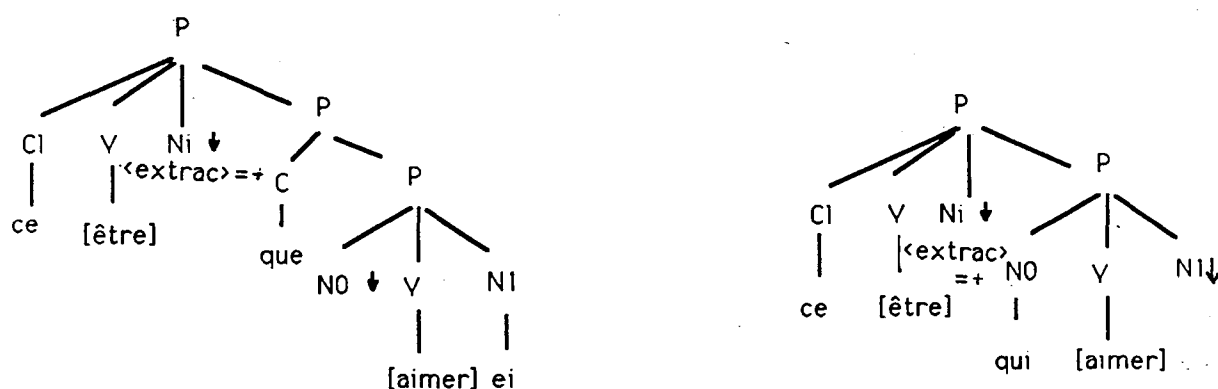


Figure 53

Le noeud N du foyer de clivée porte des restrictions sur son type de déterminant : $\langle \text{neg} \rangle = -$, $\langle \text{extrac} \rangle = +$ (Cf supra chapitre 2 et chapitre 3 section 1.4.2).

41 La co-présence de *être* et du verbe principal dans le même arbre élémentaire facilite par ailleurs l'expression de leurs corestrictions temporelles (Cf M-L. Moreau 1976).

42 L'on pourrait penser à dériver les arbres de clivées, par exemple à partir des structures relatives. Mais, comme l'a montré par exemple M-L. Moreau 1976, les contraintes sur les clivées et les relatives sont différentes, et l'on veut pouvoir analyser comme ambiguë une phrase comme:

C'est une fille qui est venue
 (présentatif+relative ou clivée)

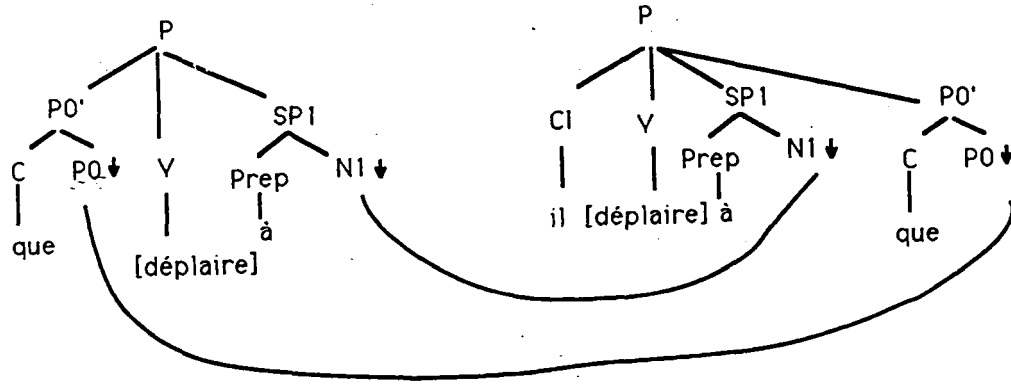


Figure 55

Le sujet SN doit être indéfini (Cf supra section 1.5). L'extraposition est normalement limitée aux verbes intransitifs, cependant on observe des cas avec un verbe transitif à objets figés ou semi-figés (voir chapitre suivant). M. Gross 1975 note également un emploi figuré (classé en table 5 avec Prep=0, de même que certains s0V tels que *survenir*) : *Il dérange Paul de faire cela*. Nous n'en avons pas tenu compte.

Elle est productive avec les verbes à complétive sujet, à l'actif ou au passif; de même pour les phrases copulatives:

Il déplaît à Max que Luc parte
Il a été rapporté que Luc partait
Il se raconte partout que Luc part
Il serait surprenant que Luc parte (M. Gross 1989b)

Elle est plus rare avec les verbes intrinsèquement intransitifs ou verbes neutres:

Trois hommes sont arrivés hier
Il est arrivé trois hommes hier

Trois branches ont cassé hier
 * *Il a cassé trois branches hier*

Il faut donc marquer la famille des verbes interdisant l'extraposition avec un trait *extrap* (= -) dans le dictionnaire⁴⁴.

L'extraposition est assez productive avec les verbes au passif. Certains verbes admettent *extrap* + passif mais pas le passif simple. Si leur structure admet le passif dans sa famille, on note un rapport d'implication *passif*=+ => *extrap*=+. C'est le cas de certains verbes à complétive objet pour lesquels le passif simple peut être difficile:

44 Pour les verbes impersonnels (*il faut que P*), nous ne les marquons pas *extrap*=+, mais les considérons directement comme phrases à sujet *il* figé (voir chapitre 4).

On a (statué + décrété) hier que tous les employés devraient porter cravate
*Que tous les employés devraient porter cravate a été (*statué + ?? décrété) hier*
Il a été (statué + décrété) hier que tous les employés devraient porter cravate

Dans les cas où la structure passive n'existe pas dans la famille des verbes considérés, il faut un trait spécial *extrap-passif* (Cf chapitre suivant)⁴⁵. C'est le cas de verbes à compléments prépositionnels comme *N₀ rire de N₁* ou *N₀ discuter de N₁ (avec N₂)*:

Il a été ri de cette histoire
Il a été discuté de ce problème (M. Gross 1975)

On note le trait *extrap* dans le dictionnaire des verbes pour les familles de verbes intransitifs *n₀V*, *n₀Vpn₁*, *n₀Vpp₁*. Pour les familles *n₀Vpn₁*, *n₀Vpn₁pn₂* et *n₀Vps₁*, on note *extrap* et *extrap-passif* (*passif* n'existe pas). L'extraposition est toujours interdite avec *n₀Vad* (ex: *N₀ aller Adv* =:Jean va bien) ou *n₀Va* (*N₀ devenir A* =:Jean devient rouge). Pour les familles *s₀V*, *s₀Vpn₁* et *s₀Va* (verbes non transitifs à sujet phrastique), on considère qu'elle est régulière avec les sujets phrastiques. On note en revanche l'extraposition des sujets nominaux correspondants, qui, elle, ne l'est pas:

(E + De) prendre cette décision appartient à Paul
Il appartient à Paul de prendre cette décision
De nombreuses décisions appartiennent à Paul
** Il appartient à Paul de nombreuses décisions* (M. Gross 1975)

Pour ces familles, le trait *extrap* ne concerne donc que les arbres à sujet nominal.

3.4.6. Les permutations

Nous représentons par des structures élémentaires différentes les permutations entre arguments⁴⁶. Ainsi les familles de verbes à deux compléments prépositionnels ont deux types d'arbres équivalents:

⁴⁵ Le problème d'indécidabilité qui naît de l'utilisation de la disjonction (ou de l'implication) sur les structures de traits (R. Kasper et W. Rounds 1986) ne se pose pas ici puisqu'il s'agit de contraintes simples, sans emboîtements.

⁴⁶ Une solution sans doute plus élégante mais plus complexe à mettre en oeuvre serait d'ajouter aux arbres élémentaires existants des règles de variation d'ordre des mots. Ce dispositif a été défini par A. Joshi 1987 pour les langues à ordre des mots "libres" (Cf chapitre 6). Pour une représentation du français comme langue à ordre des mots libre voir G. Bès, C. Gardent 1989.

Jean parle à Marie de son travail = Jean parle de son travail à Marie

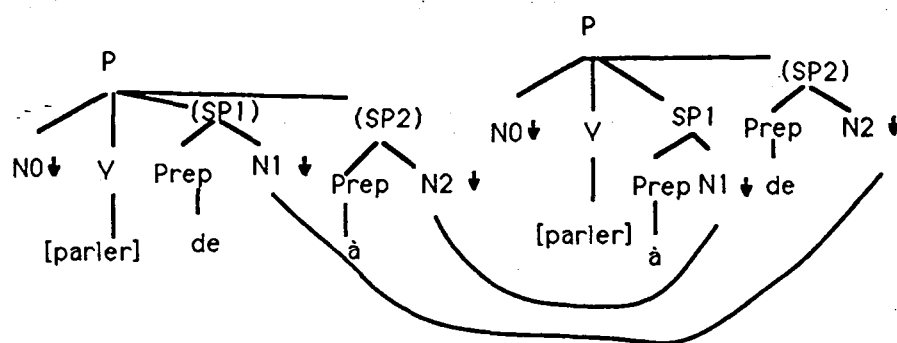


Figure 56

Nous n'avons pas représenté les permutations entre compléments directs et indirects. Elles pourraient être traitées de la même façon, en ajoutant une contrainte visant à "alourdir" le complément direct postposé (<modif> = + en top):

Jean donne des fleurs à Blanche
? Jean donne à Blanche des fleurs
Jean donne à Blanche des fleurs de son jardin

Les permutations entre sujets et compléments prépositionnels sont également représentées par des arbres spécifiques. Elles sont réservées à certains verbes, et les arbres correspondants doivent être marqués par un trait spécifique:

A cette question correspond cette réponse
** A Paul pense Marie*

Cette contrainte est représentée par un trait *permut*, pour les familles concernées (n0Vpn1, n0CIVpn1). Nous n'avons pas représenté de telles variations pour les passifs sans agents des verbes à deux compléments (*A ce salaire ont été ajoutées des primes*; M. Gross 1975).

3.4.7. La relation de neutralité

Comme le notent J-P. Boons, A. Guillet, C. Leclère 1976a, les verbes neutres (ou 0-moyens) sont assez rares en français: de l'ordre de 4 à 500 sur un total de plus de 10000. Ils n'en font pas une classe à part, étant donné la variété des cadres dans lesquels ils entrent mais une sous-structure associée à la plupart des classes de verbes transitifs. Deux classes sont particulièrement productives: 32RA (V = *rendre Adj*) et 32CV (V = *transformer en*)

Ils notent de nombreuses variations entre les deux structures à mettre en relation: $N_0 V N_1 W$ et $N_1 V W$. En particulier le complément W peut ne pas être le même, bien qu'il y ait intuition de neutralité⁴⁷:

Pierre gonfle le ballon (E + de gaz carbonique)
*Le ballon gonfle (E + * de gaz carbonique)*

Ce qui les conduit donc à la définition suivante: "il y a relation de neutralité entre $N_0 V N_1 W$ et $N_1 V W$ si et seulement si la factitive $N_0 faire V N_1 W$ est acceptable, entre en relation de synonymie relative avec $N_0 V N_1 W$ (au degré près de 'contrôle extérieur' exercé sur le procès) et représente l'application de *faire* à $N_1 V W$, ceci pour une même interprétation de la factitive" (p. 83).

Par souci de simplification, nous n'avons pas retenu tous les cas de neutralité mentionnés par J-P. Boons, A. Guillet, C. Leclère. En particulier, lorsque la complémentation (W) de la structure transitive et celle de la structure neutre sont différentes, nous avons considéré qu'il s'agissait de familles distinctes⁴⁸. De même lorsque les restrictions de sélection sont différentes pour l'objet de la construction transitive et le sujet de la construction neutre:

Pierre tourne (la manivelle + un film + ? Paul)*
*(La manivelle + * Un film) tourne*
Paul a tourné (E + à gauche)

L'humidité pourrit les fruits
Les fruits pourrissent
Pierre pourrit cet enfant
** Cet enfant pourrit*

On considèrera qu'il s'agit de deux constructions différentes, et en général de deux verbes différents. S'il n'y aucune raison sémantique de distinguer les deux emplois, il seront regroupés dans une même entrée lexicale. Nous ne créons donc pas d'ambiguïtés gênantes dans ces cas, mais nous ne notons pas la présence d'une relation de neutralité⁴⁹.

Les familles concernées par le trait *neutre* sont les suivantes: $n_0 V n_1$ ($=: N_0 cuire N_1, N_0 casser N_1$), $n_0 V n_1 p n_2$ ($=: N_0 augmenter N_1 de N_2, N_0 commencer N_1 par N_2, N_0 changer N_1 en N_2$, avec $sym = -$, ou $=: N_0 coller N_1 avec N_2, N_0 contraster N_1 avec N_2$, avec $sym = +$), $n_0 V n_1 p p_2$ (avec SP locatif $=: N_0 glisser N_1 dans N_2, N_0 tremper N_1 dans N_2$). Par défaut, *neutre* $= -$.

47 Ce qui rendrait difficile l'expression de la neutralité par métarègle.

48 Ce qui n'empêche pas de faire une seule entrée pour le verbe considéré.

49 On pourrait aussi ajouter des restrictions du type: *neutre* $= + \Rightarrow N_0: <hum> = -$ dans l'entrée lexicale de *pourrir*, et: *neutre* $= - \Rightarrow N_1: <hum> = -$ dans celui de *tourner* par exemple. Nous ne l'avons pas fait pour l'instant.

Les arbres utilisés sont les mêmes que ceux des familles intransitives correspondantes: n_0V , n_0Vpn_1 , n_0Vpp_1 .

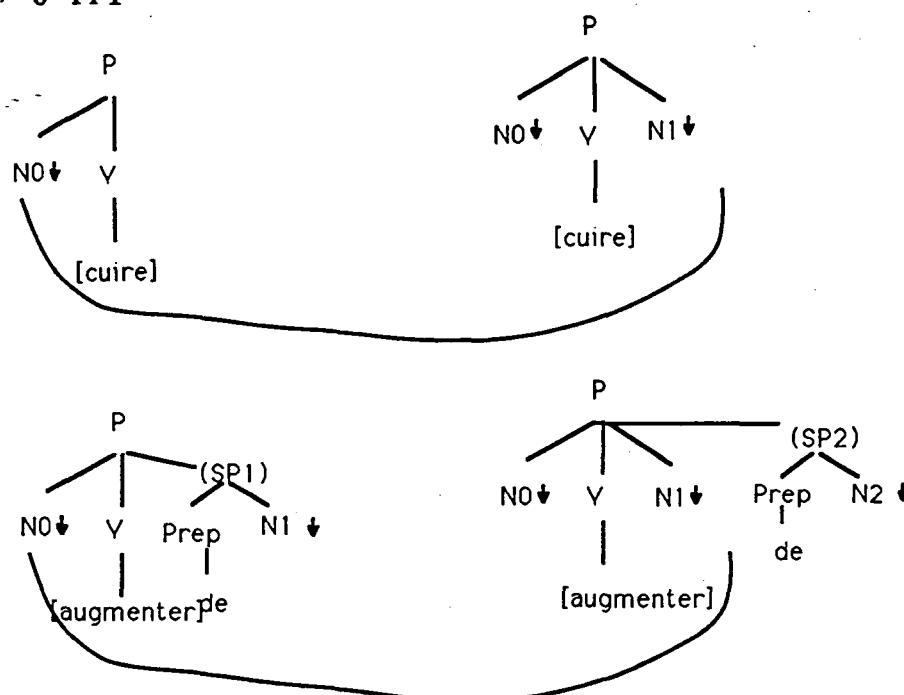


Figure 57

Nous n'avons pas représenté le *se*-neutre, ni le *se*-moyen, selon la distinction classique (Cf N. Ruwet 1982) entre:

se-neutre: *Le pont-levis s'est abaissé* (*E* + soudainement)

se-moyen: *Le grec se traduit* (**E* + facilement)

Le premier suppose d'ajouter un trait *se*-neutre aux familles concernées et correspondrait au même arbre que les verbes intrinsèquement pronominaux (*se* figé). Néanmoins, vu les nombreux problèmes d'ambiguïté que posent les constructions en *se*, nous avons préféré garder cette étude pour plus tard.

3.4.8 La relation de symétrie

La notion de symétrie a été étudiée pour les verbes par A. Borillo 1971 et J-P. Boons, A. Guillet, C. Leclère 1976.

- (1) *Jean flirte avec Marie*
- (2) *Marie flirte avec Jean*
- (3) *Jean et Marie flirtent*

Les phrases (1), (2) et (3) sont synonymes. C'est la relation entre (1) et (3) qui est considérée comme symétrique, au sens syntaxique du terme, la relation entre (1) et (2) appartenant au phénomène plus général de permutation⁵⁰.

Le prédicat peut avoir des cadres syntaxiques variés : $N_0 V Prep N_1$, $N_0 V N_1 Prep N_2$, $N_0 V Prep_1 N_1 Prep N_2$, avec $Prep := \text{à/avec}$ (et quelque fois *de/d'avec*). Il n'y a symétrie que lorsqu'il y a "couplage" d'arguments c'est-à-dire distributions identiques et dépendantes de N_i et N_j .

On associe donc aux familles concernées les arbres supplémentaires à N_i cordonnés, avec la contrainte particulière que la distribution des N_i doit être la même. Ce couplage d'arguments n'est pas forcément obligatoire dans la structure prépositionnelle, et il sera marqué lexicalement suivant les verbes. On aura donc:

Marie flirte avec (Jean + l'existentialisme)

Mais seulement:

*Marie et (Jean + *l'existentialisme) flirtent* (Boons, Guillet, Leclère 1976a)

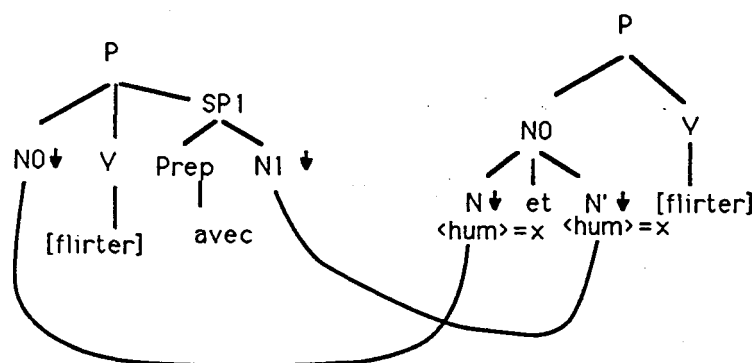


Figure 58

La relation de symétrie peut relier deux compléments:

Le maire a uni Jean à Marie
= *Le maire a uni Jean et Marie*

Mais elle peut également, à partir de la même structure mais avec un verbe différent, mettre en jeu sujet et complément:

Jean a négocié cette affaire avec Luc
= *Jean et Luc ont négocié cette affaire*

⁵⁰ Il peut y avoir permutation sans symétrie, par exemple avec N_0 être avec N_1 ou N_0 vit/habite avec N_1 :

Jean (est + vit) avec Marie
= *Marie (est + vit) avec Jean*
*= *Jean et Marie (sont + vivent)*

Six familles sont donc en jeu: $n_0Vn_1pn_2$ ($=: N_0 \text{ comparer } N_1 \text{ avec } N_2$), $n_0Vs_1pn_2$ ($=: N_0 \text{ fixer que } P \text{ avec } N_2$), n_0Vpn_1 ($=: N_0 \text{ rimer avec } N_1$), n_0ClVpn_1 ($=: N_0 \text{ se battre avec } N_1$), $n_0Vpn_1pn_2$ ($=: N_0 \text{ discuter de } N_1 \text{ avec } N_2$), et $s_0Vn_1pn_2$ ($=: \text{Que } P/N_0 \text{ séparer } N_1 \text{ de/d'avec } N_2$).

Pour les cinq dernières, la relation de symétrie est toujours la même: entre N_0 et N_1 pour n_0Vpn_1 et n_0ClVpn_1 , entre N_1 et N_2 pour $s_0Vn_1pn_2$ et entre N_0 et N_2 pour les autres⁵¹. Il suffit d'un trait *sym* à valeurs booléennes (et ce trait serait superflu si l'on faisait des sous-familles selon la préposition, puisqu'il faut et il suffit que la préposition ait *avec* ou *d'avec* parmi ses valeurs). En revanche, pour $n_0Vn_1pn_2$, il faut noter si la symétrie est entre N_0 et N_2 (type *négocié*) ou entre N_1 et N_2 (type *unir* ou *comparer*). Le deuxième type est beaucoup plus fréquent. D'où nos traits *sym* = 0 (lien entre N_0 et N_2) et *sym* = 1 (lien entre N_1 et N_2). Par défaut, *sym* = -.

Nous n'avons pas représenté les autres restructurations étudiées par J-P. Boons, A. Guillet, C. Leclère 1976 et surtout A. Guillet 1990, par exemple:

Les abeilles grouillent dans le jardin
Le jardin grouille d'abeilles

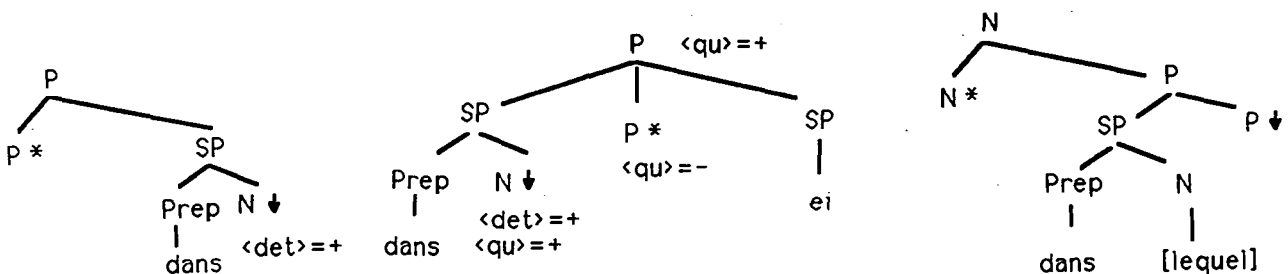
Jean a chargé les oranges dans le camion
Jean a chargé le camion d'oranges

Voir un tableau récapitulatif des différentes familles et de leurs arbres élémentaires en annexe 2.

3.5. Les familles d'ajouts

Les têtes d'adverbiaux modifiant phrase ou verbe aussi ont des familles, qui sont plus réduites mais suivent les mêmes principes que celles des arbres à tête verbale.

Ainsi la famille d'une préposition comme *dans* comprend un arbre pour la question, un pour la relative, deux pour la clivée.



⁵¹ On notera par convention le complément en *avec* comme N_2 lorsqu'il y a deux compléments prépositionnels, par exemple $N_0 \text{ parler de } N_1 \text{ avec } N_2$ (familles $n_0Vpn_1pn_2$ et $n_0ClVpn_1pn_2$).

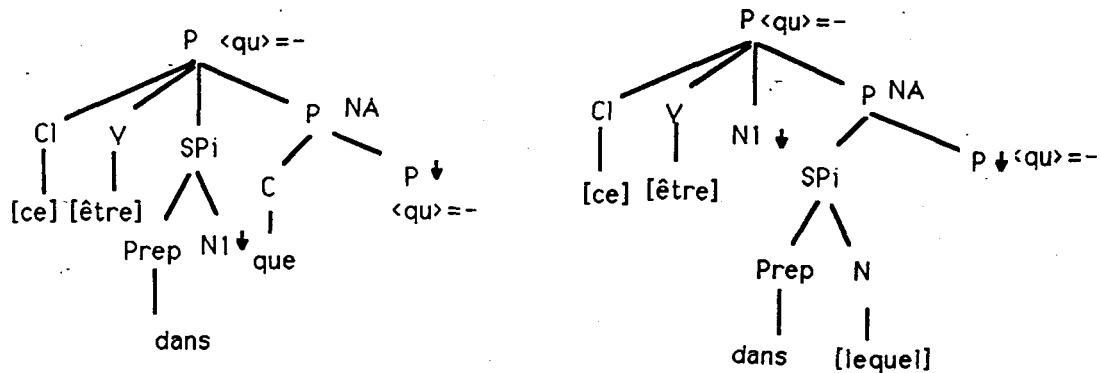


Figure 59

De même pour la famille d'un adverbe comme *conformément à N*. Les mêmes traits ($\langle qu \rangle$) que pour les structures phrastiques à tête verbale restreignent l'argument de *dans* ou de *conformément à* dans les arbres de la question ou de la relative.

Comme pour les verbes à compléments prépositionnels, deux clivées peuvent être autorisées :

C'est dans cette maison que Jean dort
C'est cette maison dans laquelle Jean dort

Certains adverbes ont des familles plus restreintes. Ainsi *où* est tête d'un arbre question et de deux arbres de relative, l'une à antécédent nominal et l'autre à antécédent adverbial:

le lit où Jean dort
là où Jean dort

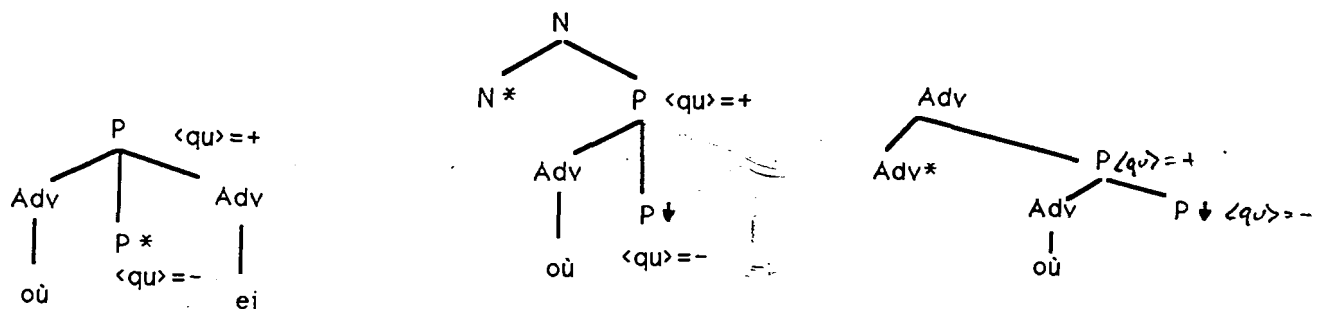


Figure 60

Les adverbes qui ne prennent pas d'argument ne peuvent autoriser que la clivée. L'on note par un trait *clivée* la différence entre les adverbes en *-ment* acceptant ou non l'extraction en *c'est que*, puisque cette distinction ne recoupe pas entièrement la distinction entre adverbes modifiant des verbes et adverbes modifiant des phrases⁵².

En conclusion, nous avons présenté les structures élémentaires correspondant aux différentes têtes lexicales simples, ainsi que le principe de leur regroupement en "familles" c'est-à-dire en classes d'équivalence. Nous avons essayé, autant que possible, de ne pas multiplier inutilement le nombre d'arbres élémentaires associés à une même entrée, chaque type d'arbre devant correspondre à un phénomène syntaxique bien identifié. Ainsi, les mêmes arbres sont utilisés pour les questions directes ou indirectes; les mêmes arbres également pour les "principales" et pour les complétives. On pourrait encore réduire le nombre d'arbres (par exemple, utiliser le même arbre pour un verbe à complément phrastique avec et sans complémenteur, ou pour les phrases permutées), mais au prix d'une grande complexification du système de traits. Le gain éventuel quant à l'"élégance" de la grammaire serait plus que compensé par une perte importante d'efficacité de traitement.

A la limite, deux ou trois types d'arbres, relativement abstraits, pourraient être le point de départ dont tous les autres seraient dynamiquement déduits. Mais une telle grammaire devient impossible à gérer.

⁵² La distinction ne recoupe pas entièrement la mobilité de l'adverbe. Ainsi les adverbes de temps, très mobiles, autorisent l'extraction (*dernièrement, récemment*). Elle ne recoupe pas non plus exactement la distinction entre adverbes de phrases et adverbes de manière (Cf C. Molinier 1985).

Chapitre IV.

Les phénomènes de semi- et non compositionnalité

Les grammaires d'arbres adjoints permettent une représentation originale des phénomènes non compositionnels, qui leur conserve une structure interne et des propriétés syntaxiques régulières. Nous examinons d'abord le cas des phrases à verbe support (qui sont des constructions semi-compositionnelles) puis celui des phrases figées.

1. Les phrases à verbe support

1.1. Discussion théorique

Les constructions à verbe support ont été étudiées en détail pour le français au LADL à partir de 1974, avec entre autres les travaux de J. Labelle 1974, J. Giry-Schneider 1978 et 1987, L. Danlos 1980. Il s'agissait au départ, reprenant l'idée de Z. Harris sur les verbes opérateurs et les mots appropriés, de représenter les nominalisations à partir d'une autre phrase simple que celle du verbe morphologiquement lié au nom. Ainsi, il n'y a pas de règle générale permettant de déduire la construction du nom *amour* (*l'amour de N_0 pour N_1*) de celle du verbe *aimer* (*N_0 aimer N_1*), ou celle du nom *agression* (*l'agression de N_0 contre N_1*) de celle du verbe *agresser* (*N_0 agresser N_1*). Il y a en revanche correspondance directe entre le groupe nominal et la phrase simple: *N_0 avoir de l'amour pour N_1 , N_0 commettre une agression contre N_1 .*

Ces constructions ont souvent été remarquées dans les études linguistiques traditionnelles¹, par leur proximité sémantique avec un verbe simple, et des études plus récentes sur la structure des groupes nominaux et l'assignation des rôles thématiques en grammaire générative leur ont donné un regain de faveur (Cf R. Cattell 1984), mais sans pour autant donner lieu à des études aussi détaillées pour d'autres langues.

Les constructions à verbe support posent plusieurs types de problèmes théoriques. Le choix du verbe support (Vsup), pour un nom prédicatif (Npred) donné, est relativement figé et en général non prévisible. Le verbe, en tant que verbe support, a des propriétés différentes (et un sens différent) de ses propriétés en emploi "plein". En revanche, le nom conserve beaucoup de ses propriétés et a en général le même sens que lorsqu'il est tête d'un groupe nominal. On a donc affaire à un phénomène de semi-compositionnalité sémantique.

¹ La *Grammaire Générale et Raisonnée* de Port Royal (1655) utilise l'existence d'une relation entre *Je suppose que P* et *Je fais une supposition (qui est) que P*, comme argument pour un traitement unifié de "que" comme relatif. Une liste étendue d'équivalents entre verbes simples et combinaisons verbe-nom a été élaborée dans le dictionnaire de B. Lafaye (1841). La tradition linguistique allemande les connaît sous le nom de *Funktionsverb*, l'anglais sous le nom de *light verbs* (depuis Jespersen).

Syntaxiquement, ces phrases ont plusieurs caractéristiques:

- le déterminant du nom prédicatif est plus restreint que lorsque le groupe nominal est dans une autre position. Comme en témoigne le contraste suivant (M. Gross 1976):

Max a rapporté (une + cette + ton + mon) agression contre Luc.

*Max a commis (une + ?cette + *ton + *mon) agression contre Luc.*

- le sujet du verbe support s'analyse comme sujet du nom prédicatif (restrictions de sélection et catégorie, interdiction du possessif et du complément de N_0 avec Npred)

(Le journal + Max) a rapporté (une agression + l'agression de Bob) contre Luc.
(phrase à verbe plein)

*(*Le journal + Max) a commis (une agression + *l'agression de Bob) contre Luc.*
(phrase à verbe support)

= *L'agression (de Max + *du journal) contre Luc*

- il y a formation régulière d'un SN complexe par incorporation du sujet de la phrase à support:

Max a commis une agression contre Luc

= *l'agression que Max a commise contre Luc*

= *l'agression de Max contre Luc*

Il y a proximité sémantique entre le SN complexe *l'agression de Max contre Luc* et la relative de la phrase à verbe support, mais non avec les relatives des phrases à verbe plein (e.g. *l'agression contre Luc que Max a rapportée*)

- on peut extraire le complément du nom prédicatif seul ou avec ce dernier (double extraction)

C'est une agression contre Luc qu'il a rapportée

* *C'est une agression qu'il a rapportée contre Luc*

* *C'est contre Luc qu'il a rapporté une agression²*

C'est une agression contre Luc qu'il a commise

C'est une véritable agression qu'il a commise contre Luc

C'est contre Luc qu'il a commis une agression

Sinon, la construction est syntaxiquement régulière, et le nom prédicatif peut avoir toutes les propriétés d'un complément direct ou indirect: être passivable, être questionnable, plus difficilement pronominalisable etc...

Ces phrases posent pour l'analyse automatique les mêmes problèmes que ceux des expressions figées en général: il faut reconnaître le groupe [Vsup Npred], car les

² Cette dernière phrase ne serait acceptable qu'avec l'interprétation : *C'est pour nuire à Luc qu'il a rapporté cette agression (contre Bob)*

conditions de bonne formation de la phrase ne sont pas les mêmes que lorsqu'il s'agit du même nom avec un verbe plein, et aussi pour préparer l'analyse sémantique, alors que, syntaxiquement, il ne s'agit pas d'un syntagme. De plus, il peut y avoir des discontinuités entre verbe support et nom prédicatif et leur ordre respectif peut être variable:

Max a commis hier soir je ne sais combien d'agressions contre les passants
Les innombrables agressions que Max a commises finissent par nous excéder.

Le fait de pouvoir manipuler des unités de base étendues, et pouvant comporter plusieurs syntagmes, est ici un avantage de notre formalisme.

Ces constructions posent aussi des problèmes théoriques. M. Gross 1976 a proposé une double analyse des phrases à verbe support pour rendre compte de la double possibilité d'extraction:

- a) N_0 Vsup [Det Npred Prep N₁]SN
- b) N_0 Vsup [Det Npred]SN [Prep N₁]SP

Selon le schéma a), c'est tout le groupe *une aggression contre Marie* qu'on extrait; selon le schéma b), c'est *contre Marie* et *une aggression* qu'on extrait séparément.

J. Bresnan 1982a propose pour l'anglais, dans le cadre de LFG, une double réanalyse pour expliquer des doubles passifs dans des cas analogues:

No notice was taken of Simone.
Simone was taken no notice of.

En réanalysant comme un verbe la séquence *take+of*, la règle du passif, normalement réservée aux verbes transitifs, peut s'appliquer, de même en réanalysant la suite *take+notice+of* comme un verbe. Un premier problème est qu'il peut y avoir interposition de modificateurs (ici *no*) entre les mots à réanalyser, ce qui entraîne à définir une réanalyse sur des séquences non bornées.

Un deuxième problème, qui se pose aussi pour la "double analyse", est que les phrases de départ ne sont pas ambiguës. C'est rompre la relation habituelle entre analyse syntaxique et analyse sémantique que d'assigner deux structures différentes à une phrase non ambiguë³.

Nous avons proposé une analyse simple, qui permet le même résultat sans supposer une double analyse, et sans reposer non plus sur deux niveaux de représentation linguistique (Abeillé 1988a).

La représentation des constructions à verbe support pose également d'autres problèmes, tous indépendants de notre système. D'une part, toutes les combinaisons Vsup-Npred

3 La critique a moins de portée dans le cadre de LFG, si l'interprétation sémantique ne s'appuie que sur la structure fonctionnelle (qui peut être la même pour les deux phrases) et non sur la structure en constituants. Mais il faut alors déterminer la structure fonctionnelle au niveau lexical et indépendamment de la structure en constituants, ce qui pose d'autres problèmes.

n'ont pas été décrites pour le français. Seules celles mettant en jeu des Vsup particuliers (*avoir, faire, être prep* et quelques autres) l'ont été (en particulier, *porter* et *commettre* n'ont pas été systématiquement étudiés). Nous avons donc dû les compléter pour les noms qui nous intéressent.

D'autre part, se pose une question théorique plus générale. La notion d'extension de verbe support (proposée par M. Gross 1981 et développée dans R. Vivès 1983 et G. Gross 1990) rend les limites de la notion incertaine. Les limites entre phrases à verbe support et expressions figées sont également floues (M. Gross 1988). Nous avons donc, suivant en cela L. Danlos (à paraître), retenu la formation du SN complexe par incorporation du sujet comme critère discriminant des constructions à support. De nombreuses constructions à nom prédicatif sont donc représentées dans notre grammaire non comme constructions à verbe support mais comme expressions figées (voir section 2).

1.2. Représentation en TAG

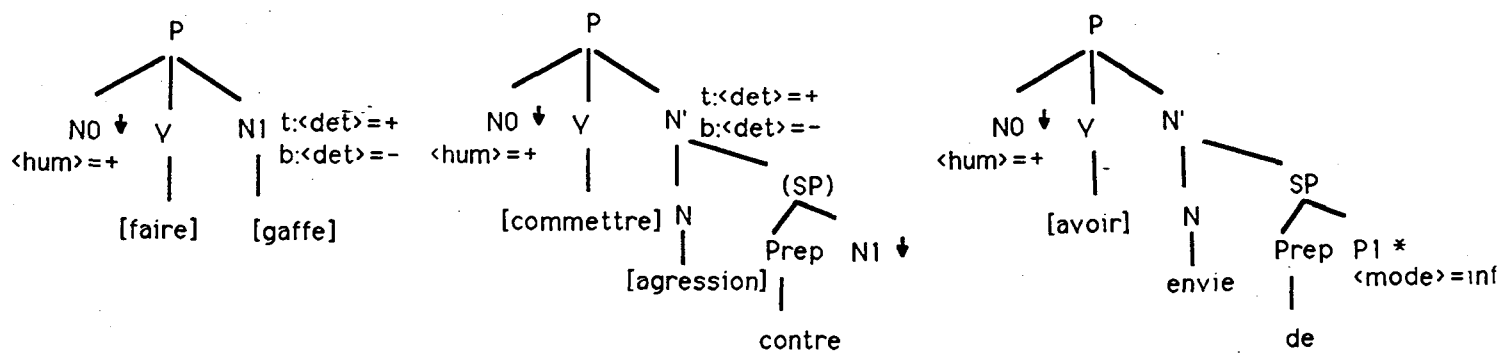
Les constructions à verbe support sont encore souvent ignorées en linguistique informatique, à l'exception du projet Eurotra (F. Namer et al. 1990, L. Danlos à paraître) et d'extensions en cours de la grammaire en chaîne (M. Salkoff 1990). La double extraction n'est intégrée dans aucune de ces implémentations.

Elles se représentent naturellement dans notre formalisme qui est particulièrement bien adapté aux phénomènes non compositionnels en général. En revanche, tous les problèmes théoriques restent entiers.

Nous avons défini des familles particulières, qui incorporent un ou plusieurs arbres de groupes nominaux, en plus des arbres phrastiques, et des relatives particulières (lorsque l'antécédent est le nom prédicatif).

Les phrases à verbe support sont représentées comme des arbres élémentaires à tête multiples, plus étendus que les arbres élémentaires des phrases simples. Leur tête lexicale comprend le verbe support, le nom prédicatif, éventuellement le déterminant de celui-ci et des prépositions⁴:

4 Nous notons donc chaque combinaison Npred-Vsup comme une entrée. D'après les études faites au LADL, il ne semble pas possible de prédire la valeur lexicale du Vsup à partir des propriétés morphologiques, syntaxiques ou sémantiques du nom prédicatif. Seules des corrélations très parcellaires ont été observées: ainsi tous les noms de maladie entrent dans la construction N_0 avoir Le Nmal (= : N_0 avoir la grippe, N_0 avoir le choléra... Cf J. Labelle 1974); un certain nombre de noms dérivés de verbes "psychologiques" (table 4 de M. Gross 1975) entrent dans la construction N_0 être en Npred devant N_1 (D. de Negroni-Peyre 1978): N_0 émerveiller N_1 = N_0 être en émerveillement devant N_1 . Ce n'est pas la situation générale. Pour d'autres tentatives de rationalisation sémantique, voir par exemple A. Wierzbicka 1982.



N₀ faire une gaffe N₀ commettre une agression contre N₁ N₀ avoir envie de/que P

Figure 1

Les phrases à support ont des cadres syntaxiques variés, mais ce sont les mêmes que ceux des phrases simples. Leur sujet peut être nominal ou phrastique, Npred peut avoir 0, 1 ou 2 compléments. Lorsque Npred a un complément phrastique, l'arbre élémentaire correspondant est un arbre auxiliaire (comme pour les verbes à complétive), et non un arbre initial.

Le verbe support peut avoir un autre complément que Npred. Nous notons comme complément de Vsup les syntagmes qui ne peuvent en aucun cas être considérés comme complément de Npred, c'est-à-dire que l'extraction avec Npred est interdite. Ainsi, nous représentons à Marie comme complément exclusif de faire dans l'expression *N₀ faire la cour à N₁*:

Jean fait (la cour + une cour assidue) à Marie.

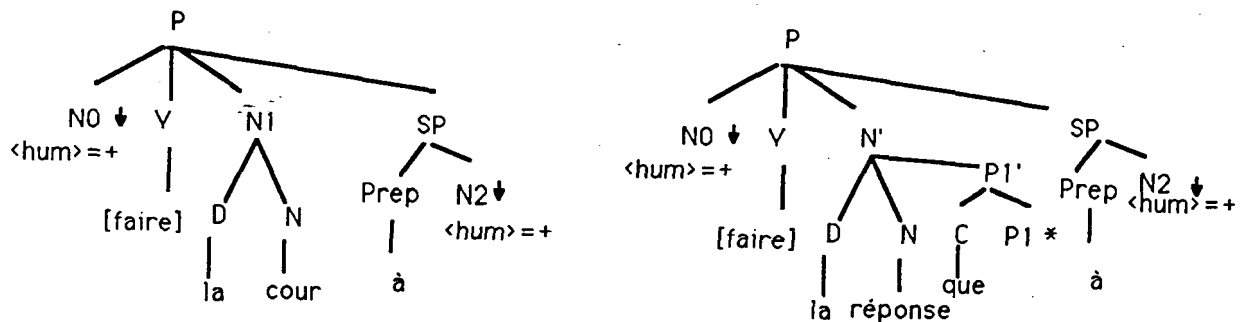
** C'est (la cour + une cour assidue) à Marie qu'il a faite.*

*C'est (*la cour + une cour assidue) qu'il a faite à Marie.*

C'est à Marie qu'il a fait (la cour + une cour assidue).

Cela n'empêche bien évidemment pas l'incorporation régulière de à Marie dans le SN complexe:

la cour de Jean à Marie



N₀ faire la cour à N₁

N₀ donner la réponse que P à N₂

Figure 2

Lorsque l'extraction avec Npred est possible, nous représentons le complément comme complément de Npred, mais cela n'empêche pas qu'il puisse également être extrait séparément (voir infra p. 184).

Le nom prédicatif peut a priori occuper n'importe quelle position syntaxique, même s'il est plus souvent objet direct. Il peut être complément prépositionnel (L. Danlos 1980):

Jean est dans l'ignorance de cet événement
 = *? L'ignorance de Jean de cet événement*

Le nom prédicatif peut aussi être sujet (le critère du SN complexe suppose alors l'incorporation des compléments libres éventuels):

Un accord avec Luc interviendra à trois heures (M. Gross 1977)
Une envie soudaine de s'en aller a pris Luc
 = *l'envie soudaine de s'en aller de Luc*

Les phrases à nom prédicatif sujet ont été les moins étudiées et nous ne les avons pas incorporées en tant que telles à notre grammaire. Il semble que dans la plupart des cas, une dérivation concurrente du SN complexe soit possible et préférable (voir infra).

1.2.1. Propriétés distributionnelles

a) Variations du déterminant de Npred

Le déterminant de Npred est par définition contraint mais en général assez variable. Lorsqu'il est variable, ce ne peut en général pas être un possessif, et cette contrainte sera marquée directement dans l'arbre. Dans les cas où l'on a affaire à un déterminant figé, celui-ci fait directement partie de la tête lexicale de l'expression et peut parfois être son: *N₀ donnerson aval à N₁*.

Le déterminant doit pouvoir être indéfini (ou partitif) pour qu'il puisse y avoir relativation (et donc formation du SN complexe), comme le note J. Giry-Schneider 1978 suivant l'analyse de S. Kuroda 1968 qui donne pour source à l'antécédent du pronom relatif (d'une relative restrictive) un substantif à déterminant indéfini.

Hormis ces contraintes, les variations du déterminant peuvent dépendre de plusieurs facteurs. Elles sont relativement libres dans des cas comme N_0 commettre Det agression contre N_1 :

Max a commis (une + un nombre vraiment effroyable de + je ne sais combien de) agression(s) contre Luc

Elles sont plus contraintes dans des cas comme N_0 avoir du courage ou N_0 faire la cour à N_1 :

*Jean a (du + *un + *le + *plusieurs) courage*
*Jean fait (*de la + *une + la + *plusieurs) cour à Marie*

On observe alors une alternance du déterminant figé avec Un-modif (M. Gross 1977):

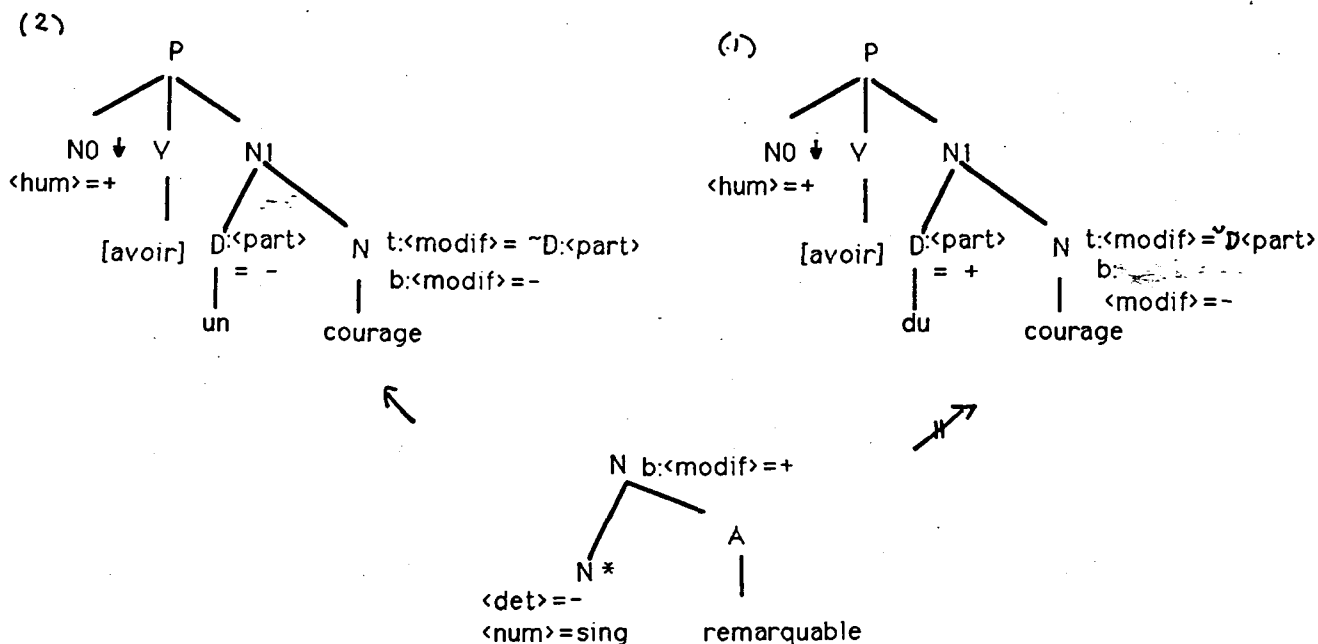
Jean a (un certain courage + un courage remarquable + un courage qui vous épate)
Jean fait une cour assidue à Marie

On note alors *un* comme tête lexicale de l'expression en disjonction avec le déterminant figé, et, associée à la famille de l'expression, la contrainte de modifieur obligatoire dans le cas où c'est *un* qui est réalisé. Ainsi pour N_0 faire la/une cour à N_1 , on note la contrainte de traits entre le déterminant (Det) et le nom prédicatif (N): Det: <def> = ~ N.t: <modif>⁵. Pour N_0 avoir du/un courage, on note Det: <part> = ~ N.t: <modif>. Dans le dictionnaire morphologique, les déterminants ont les traits suivants:

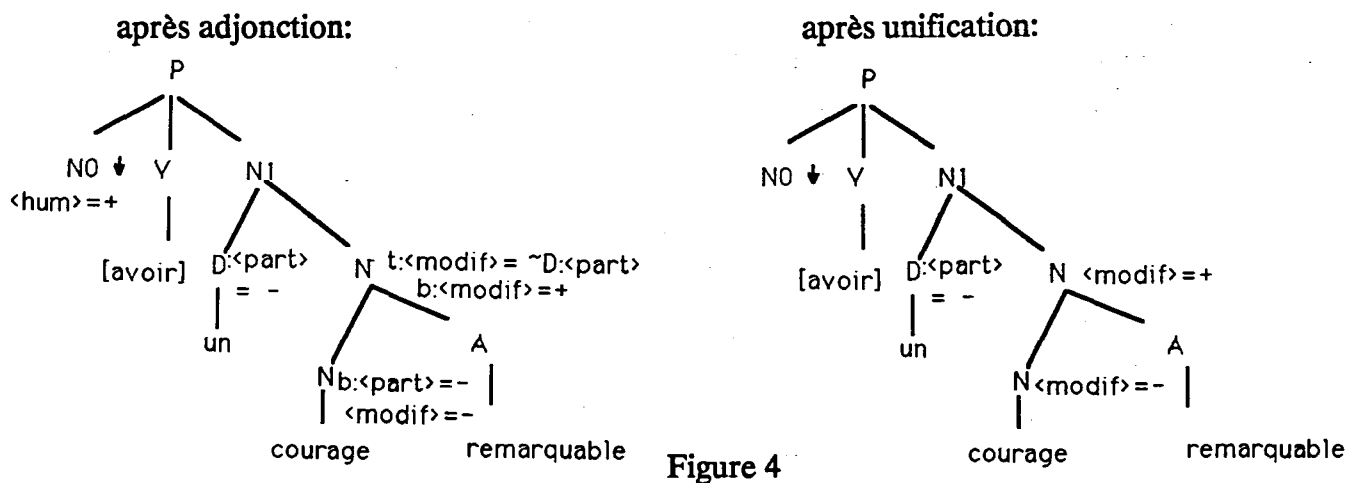
un: { <def> = -, <part> = - }
du: { <def> = -, <part> = + }
le: { <def> = + }

Ainsi, pour N_0 avoir du/un courage, si le déterminant est *du*, on a en top de N la contrainte <modif> = -. Aucune adjonction de modifieur (adjectif ou relative) n'est donc possible. En revanche, si l'on a *un*, le trait top de N devient <modif> = +. Comme par lui-même le nom prédicatif porte <modif> = - comme trait bottom, l'adjonction d'un modifieur est obligatoire.

⁵ Il s'agit d'une contrainte locale puisque les deux noeuds sont coprésents dans l'arbre élémentaire. Le signe ~ note la négation.



L'adjonction d'un modifieur sur le nom prédicatif est impossible en (1), car son noeud N porte en top le trait $\langle \text{modif} \rangle = -$ (puisque $\langle \text{part} \rangle = +$). En revanche, elle est obligatoire en (2) puisque N porte en top un trait $\langle \text{modif} \rangle = +$ (puisque $\langle \text{part} \rangle = -$) qui est incompatible avec son trait bottom $\langle \text{modif} \rangle = -$. Seule l'adjonction de *remarquable* (ou d'un autre modifieur) permet de résoudre cette incompatibilité (puisque l'adjonction sépare les traits top et bottom du noeud où elle a lieu en les transférant sur des noeuds différents):



De même, pour N_0 avoir peur (de N_1) ou N_0 faire son entrée (Loc N_1), c'est-à-dire pour tous les déterminants figés Det = *du, le, 0* ou *son*.

Nous n'avons pas représenté les variations avec Le-Modif où le modifieur est un modifieur d'unicité au sens de M. Gross 1977: *Jean a (? le plus grand courage de la terre + le courage qu'il faut)*.

Un autre cas d'alternance s'observe quand le complément du nom prédicatif est optionnel. Suivant la réalisation de ce complément, le déterminant (et le nombre) du nom prédicatif est figé ou relativement libre. Ceci est particulièrement vrai quand il s'agit d'un complément phrastique:

*Jean a pris (?*la + une + trois + cette + la plupart des) décision(s).*
*Jean a pris (la + *une + *cette + *trois + *la plupart des) décision de partir*

Npred doit être singulier et avoir Le pour déterminant. Il est à noter que la même contrainte s'applique quand Npred est tête d'un SN complexe:

(la + une + cette + trois + la plupart des) décision(s) de Jean
*(la + *une + *cette + *trois + *la plupart des) décision de Jean de partir*

Nous avons représenté cette contrainte directement dans les arbres à compléments phrastiques des familles concernées ($n_0VdN(ps_1)$ et $n_0VdN(ps_1)(pn_2)$):

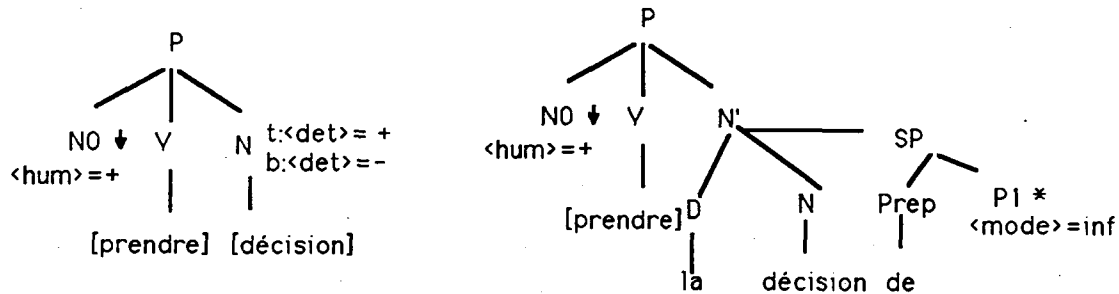


Figure 5

Le fait d'utiliser des arbres différents lorsque le complément est réalisé ou non permet aussi d'encoder une différence d'ordre préférentiel avec un deuxième complément: N_0 donner un/des conseil(s) à N_1 mais N_0 donner à N_1 le conseil de P_{inf} .

Dans le cas de compléments uniquement nominaux, il n'y a pas de contrainte régulière. Le complément optionnel peut être réalisé et Npred avoir des déterminants variés:

Jean a peur (E + de Marie)
Jean commet une agression (E + contre Luc)

Lorsque la présence du complément entraîne un figement du déterminant de Npred, le plus simple est alors de noter le déterminant figé comme tête lexicale optionnelle de l'expression et de décomposer la construction en deux familles (pour la même entrée lexicale): l'une où Npred est sans complément et son déterminant libre, l'autre à déterminant figé et complément réalisé :

[rencontre],N;([le],D);(de,P);[faire],V: $n_0VdN, n_0VDNpn_1\{N: <num> = sing\}$.

On aura donc bien⁶:

Jean fait (une + cette + plusieurs) rencontre(s)

*Jean fait (*une + *plusieurs + *cette + la) rencontre de Marie*

Nous avons ignoré (pour le moment) les phénomènes de dépendance entre déterminant de Npred et déterminant de son complément, qui sont particulièrement complexes.

c) Contraintes morphologiques sur Npred et Vsup

Nous avons vu les contraintes liées à la présence d'un complément. Les Npred sans compléments peuvent aussi être contraints. Ainsi, N_0 *avoir des besoins* est autonome par rapport à l'expression N_0 *avoir besoin de* N_1/P_1 . Il faut donc noter que, dans le premier cas, Npred est obligatoirement au pluriel. De même pour N_0 *faire des enfantillages* ou N_0 *avoir des difficultés*.

C'est pourquoi les têtes lexicales peuvent être des formes fléchies et non seulement des lemmes.

Le verbe support peut également avoir moins de formes fléchies qu'en emploi plein. Le participe passé du verbe *porter* est invariable dans l'expression *porter plainte*, puisqu'il n'y a ni passif, ni relative possible de *plainte*, alors qu'il est variable autrement : *les oranges que j'ai portées à Marie*.

Même sans restrictions syntaxiques particulières, la flexion du participe passé est toujours contrainte par la nature du Npred complément. Ainsi, dans l'expression N_0 *faire une gaffe*, *faire* peut être au participe passé masculin singulier (*Jean a fait une gaffe*), féminin singulier ou pluriel (*la gaffe que Marie a faite; les gaffes que Marie a faites*) mais pas masculin pluriel. Compte tenu de ce que les autres formes fléchies de Vsup sont les mêmes qu'en emploi "plein", nous disons qu'il s'agit morphologiquement du même verbe (même lemme) mais avec des restrictions particulières notées dans l'entrée syntaxique de la phrase à support.

d) Variations de Vsup

Lorsque plusieurs constructions à verbe support sont possibles, c'est en général que le nom prédicatif est ambigu, et il est non seulement légitime mais nécessaire de faire plusieurs entrées différentes (cette "désambiguation" est d'ailleurs une des justifications de la notion de construction à support):

Jean est en forme

Ce vase a une forme (parfaite)

⁶ L'interdiction du déterminant *Le* dans la construction à Npred sans complément semble être générale (J. Giry-Schneider 1978), elle est donc notée directement dans les structures arborescentes.

Jean fait de l'économie
Jean fait (une + des) économie(s)

Le nom *forme* a un sens différent (et des sujets différents) dans *N₀ être en forme* et dans *N₀ avoir une forme* (modif). On fait donc deux entrées différentes. De même, le nom *économie* au sens de science économique entre dans la construction *N₀ faire du N₁*, tandis qu'en tant que dérivé du verbe *économiser*, il entre dans la construction *N₀ faire un/des N₁*. On distinguera donc deux entrées pour ce nom.

Mais il y a parfois plusieurs constructions à Vsup possibles pour un même nom prédicatif:

Jean est en retard (E + sur N₁)
Jean a du retard (E + sur N₁)

Jean a (envie + une envie soudaine) de s'en aller
(L'envie + Une envie soudaine) de s'en aller prend Luc

Les verbes supports ne sont pas représentés par des traits mais directement comme tête lexicale de l'expression, il faudra donc faire plusieurs entrées. Le problème est que le SN complexe correspondant (*le retard de Jean*) va être analysé comme ambigu, puisque sa structure peut appartenir à la famille de *N₀ être en retard* et *N₀ avoir du retard*. De même pour *l'envie de Luc de s'en aller* qui peut être dérivé de *L'envie de s'en aller prend Luc* (*L'envie de Pinf prend N₁*) ou de *Luc a envie de s'en aller* (*N₀ a envie de Pinf*). Il faut donc choisir une dérivation. Nous choisissons de préférence celle à Vsup transitif qui est plus naturelle:

Jean a du retard
 = *le retard qu'a Jean*
 = *le retard de Jean*

Jean a envie de s'en aller
 = *l'envie de s'en aller qu'a Jean*
 = *l'envie de s'en aller de Jean*

Nous excluons ainsi la structure à SN complexe pour *N₀ être en retard* ou *L'envie de Pinf prend N₁*. Comme nous ne souhaitons pas introduire un trait spécifique pour cette contrainte, la solution est donc de considérer ces expressions comme phrases figées (dont les familles ne comportent pas d'arbres SN complexes par définition) et non comme phrases à verbe support.

Lorsqu'un nom prédicatif a deux constructions à support exactement parallèles, une possibilité est de représenter cette alternance par une disjonction sur la tête verbale de l'expression⁷:

⁷ La disjonction sur les têtes lexicales est un procédé plus général utilisé aussi pour les expressions figées: *N₀ casser la graine/croûte*.

N₀ faire/commettre une gaffe.

N₀ faire/donner (une réponse + la réponse que P) à N₂.

Mais cela suppose que les expressions soient sémantiquement équivalentes. Ce procédé ne convient donc pas quand il y a variation aspectuelle régulière:

Jean a le goût de se lever tôt

Jean perd le goût de se lever tôt

Jean prend le goût de se lever tôt (R. Vivès 1983)

Il ne convient pas non plus quand il y a variation des prépositions associées, puisque celles-ci sont des têtes lexicales de l'expression (et non représentées sous forme de traits). Dans ces deux cas, il faut faire des entrées lexicales distinctes, comme pour *N₀ porter de l'admiration à N₁* vs *N₀ avoir de l'admiration pour N₁* (sauf s'il y a relation régulière entre les deux phrases à support, voir infra p. 195 pour la relation de causalité)

e) Certaines combinaisons de verbes et de noms prédicatifs ne se comportent pas comme des phrases à support mais comme des auxiliaires. Ils sont suivis de l'infinitif uniquement et celui-ci n'a aucune propriété de complément:

Marc a tendance à exagérer

?* *A quoi Marc a-t-il tendance ?*

* *C'est à exagérer que Marc a tendance*

Paul a coutume de mentir

* *De quoi Paul a-t-il coutume ?*

* *C'est de mentir que Paul a coutume*

En général, ces auxiliaires composés sont "neutres" pour la sélection du sujet de la phrase, qui dépend du seul infinitif:

Il a tendance à pleuvoir beaucoup dans cette région.

? *La moutarde avait coutume de lui monter au nez chaque fois qu'il entendait parler d'elle.*

Nous les représentons par des arbres auxiliaires verbaux (comme les modaux et les auxiliaires simples) étendus au verbe et au nom prédicatif :

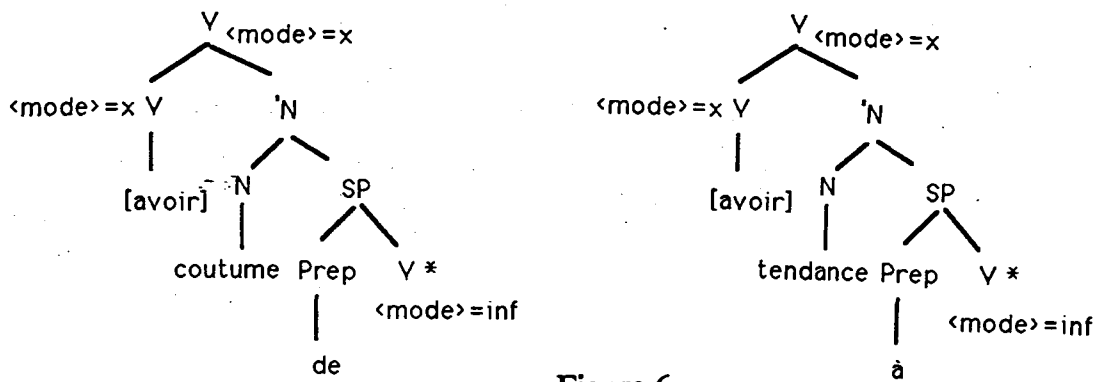


Figure 6

Nous ne considérons pas ces combinaisons comme des phrases à support, puisque la formation d'un SN complexe est en général interdite⁸:

* *La coutume de Marc de mentir*

1.2.2. Familles des constructions à verbe support

Ce sont des familles particulières puisque les arbres sont plus étendus que pour les phrases libres. Tous les arbres sont à têtes multiples : le verbe support, le nom prédicatif, souvent le déterminant de Npred, et parfois une préposition avant Npred ou après sont les éléments têtes. Alors que pour les familles des verbes simples seuls les arbres du passif, des relatives et des clivées sont à têtes multiples (Cf chapitre 3).

Sinon, les phrases à verbe support entrent dans les mêmes cadres syntaxiques: le Npred a 0, un ou deux compléments; ses compléments sont toujours prépositionnels (avec ou sans complétives); les sujets peuvent être nominaux ou phrastiques. On note N le nom prédicatif et P la préposition éventuelle qui suit le Vsup. La dénomination et la numérotation des arguments est autrement la même que pour les familles libres: n_0VN (ex N_0 faire un saut), $n_0VPNp_{n_1}$ ($=: N_0$ être en colère contre N_1) ...

Ces constructions suivent les mêmes règles syntaxiques que les familles libres, et on retrouve les mêmes types d'arbres élémentaires correspondant à chaque transformation ou à la combinaison de deux transformations: question, passif, symétrie etc...

Une particularité de ces familles est qu'en plus des arbres phrastiques, elles contiennent au moins un arbre élémentaire de racine N (pour le SN complexe).

⁸ Un syntagme comme *La tendance à exagérer de Marc* sera dérivée de *Marc a une certaine tendance à exagérer*, qui elle est une phrase à verbe support (à sujet humain Cf A. Meunier 1977).

a) Le SN complexe

C'est ainsi qu'on représente la relation systématique entre la construction du SN et la construction de la phrase à Vsup correspondante, relation définitoire de la notion de phrase à Vsup et essentielle pour faire la différence avec les phrases figées.

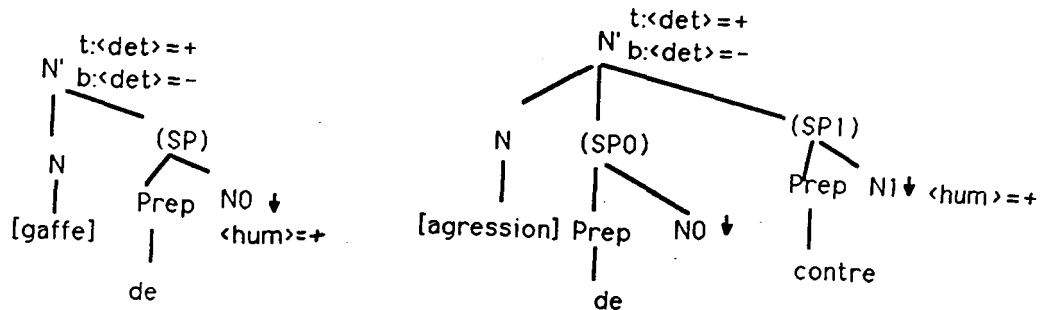


Figure 7

La préposition *de* pour l'incorporation du sujet N_0 est déjà dans l'arbre. La préposition du complément est la même que celle de la construction phrastique (et fait partie de l'entrée lexicale de l'expression). L'alternance de *de* N_0 (ou *de* N_1) avec un déterminant possessif est présentée au chapitre 6. Elle dépasse le cadre des phrases à support.

On a noté seulement les groupes nominaux actifs, pas les passifs (incorporation du sujet en *par* et complément introduits par *de* =: *L'agression de Marie par Luc*) ni les converses au sens de G. Gross 1990 (Cf aussi A. Daladier 1977 et L. Danlos 1991). On pourrait ajouter des arbres SN passifs correspondant aux arbres phrastiques passifs de la phrase à support (et sélectionnés avec eux par le trait *passif* puisqu'il semble y avoir une bonne corrélation entre les deux constructions) :

- Picasso a fait (le + un) portrait d'Alexandre*
 = *(Un + Le) portrait d'Alexandre que Picasso a fait*
 = *(Un + Le) portrait d'Alexandre de Picasso*
- (Un + Le) portrait d'Alexandre a été fait par Picasso*
 = *(Un + Le) portrait d'Alexandre qui a été fait par Picasso*
 = *(Un + Le) portrait d'Alexandre par Picasso*

Une telle construction ne s'observe pas avec les Npred sans complément (même si le passif s'applique) et il faudrait réserver ces structures de SN aux familles à Npred avec complément.

- Jean a fait une gaffe*
Une gaffe a été faite par Jean
*(La + Une) gaffe (de + *par) Jean*

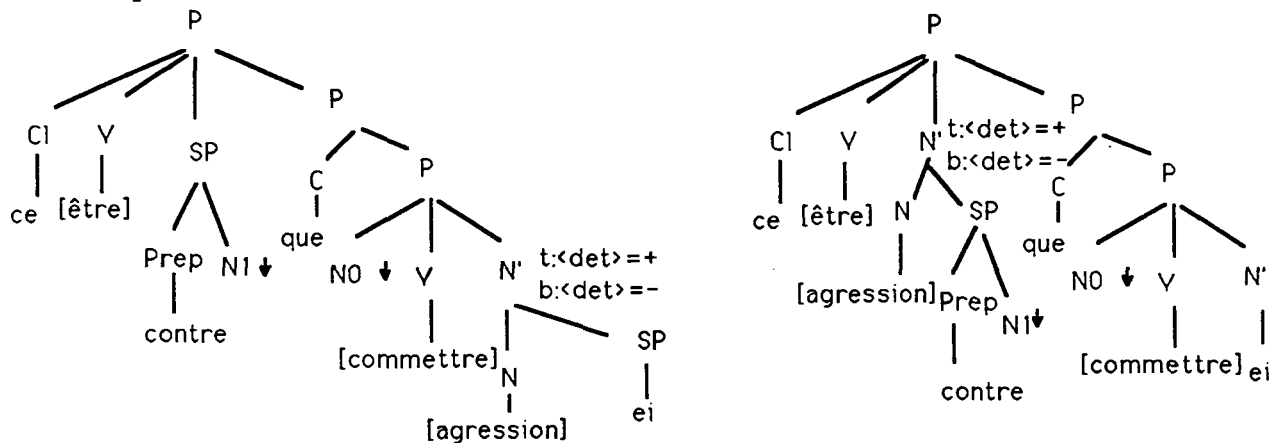
Il est à noter également que le complément de Npred, optionnel avec le sujet en *de*, est obligatoire avec le sujet en *par*.

Il faudrait enfin des traits supplémentaires pour exclure dans certains cas le SN actif (sujet en *de* N) lorsque le complément du nom prädicatif est réalisé (Cf A. Daladier 1978, J-C. Milner 1982) :

Jean a fait (cet achat + l'achat d'une mobylette)
*(L'achat d'une mobylette + *Cet achat) par Jean*
*(*L'achat d'une mobylette + Cet achat) de Jean*

b) Phrases clivées

Nous représentons les phrases clivées comme des arbres élémentaires à têtes multiples, comme dans le cas simple. Outre les clivées pour les arguments libres, il y a une structure pour l'extraction de Npred seul et une pour l'extraction éventuelle avec son complément:



*C'est contre Marie que Jean
a commis une agression*

*C'est une agression contre Marie
que Jean a commise*

Figure 8

Le fait d'avoir une structure étendue sur tous les éléments de laquelle peuvent s'appliquer les règles syntaxiques, permet donc d'autoriser naturellement la double extraction sans avoir besoin d'une double analyse (Cf A. Abeillé 1988a). Cette double extraction sera exclue dans le cas d'une phrase à verbe simple, puisque le complément étant réduit à un seul noeud (feuille), une seule clivée pourra lui correspondre:

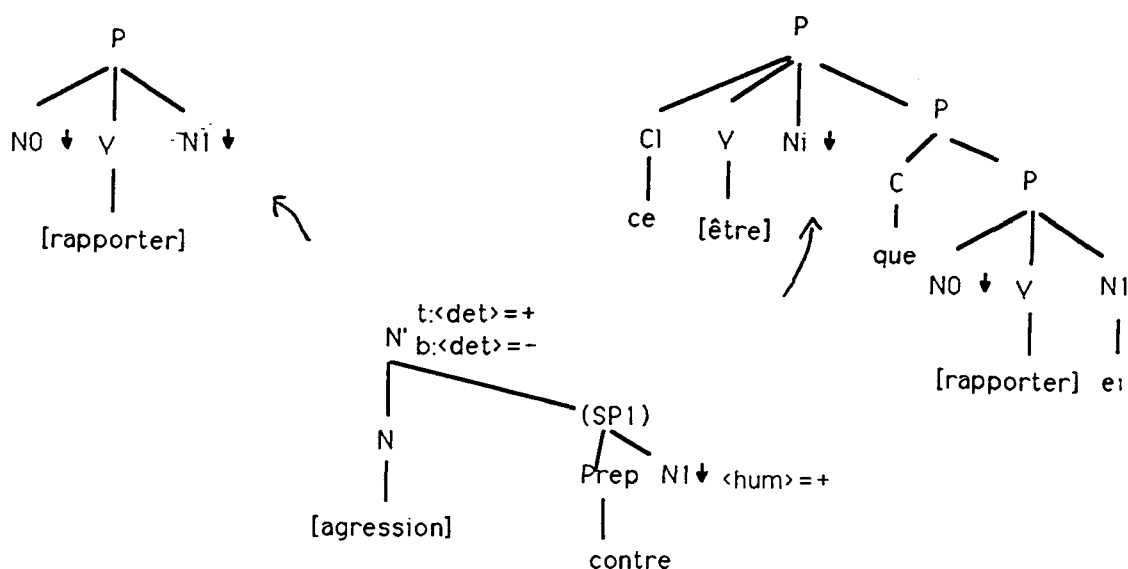


Figure 9

L'on autorisera bien la phrase 1 tout en interdisant les phrases 2 et 3:

- (1) *C'est une agression contre Luc que Max a rapportée*
- (2) * *C'est contre Luc que Max a rapporté une agression*
- (3) * *C'est une agression que Max a rapportée contre Luc*

* extraction de Npred

Le foyer de clivée a toujours le trait top <det> = +, ce qui exclut le déterminant zéro, et rend souvent l'adjonction d'un modifieur obligatoire (pour les Npred qui autorisent l'alternance 0/Un-modif):

Jean rend (hommage + un vibrant hommage) au professeur Minch
*C'est (*hommage + un vibrant hommage) qu'il a rendu au professeur Minch*

De plus, pour les Npred à déterminant figé, l'extraction n'est possible que si le déterminant figé est Un, ou alterne avec Un-modif (Cf chapitre 2). Nous plaçons donc comme contrainte <def> = - en top sur le Npred dans la structure clivée des familles à det de Npred figé (mais pas sur celles à det de Npred libre):

Jean a commis une agression (E + surprenante) contre Luc
C'est une agression (E + surprenante) que Jean a commise contre Luc

Jean fait (la cour + une cour assidue) à Marie
*C'est (*la cour + une cour assidue) que Jean fait à Marie*

Notons que cette contrainte ne concerne que la structure où Npred est extrait seul. Quand il est extrait avec son complément (libre par définition), la clivée est généralement possible, même si le déterminant de Npred est figé et différent de Un ou Un-modif:

*Jean fait (*une + *cette + la) rencontre de Marie*
C'est la rencontre de Marie que Jean a faite

Ces clivées, comme celles des familles où le déterminant du nom prédicatif est libre, portent les mêmes restrictions que les clivées des phrases libres.

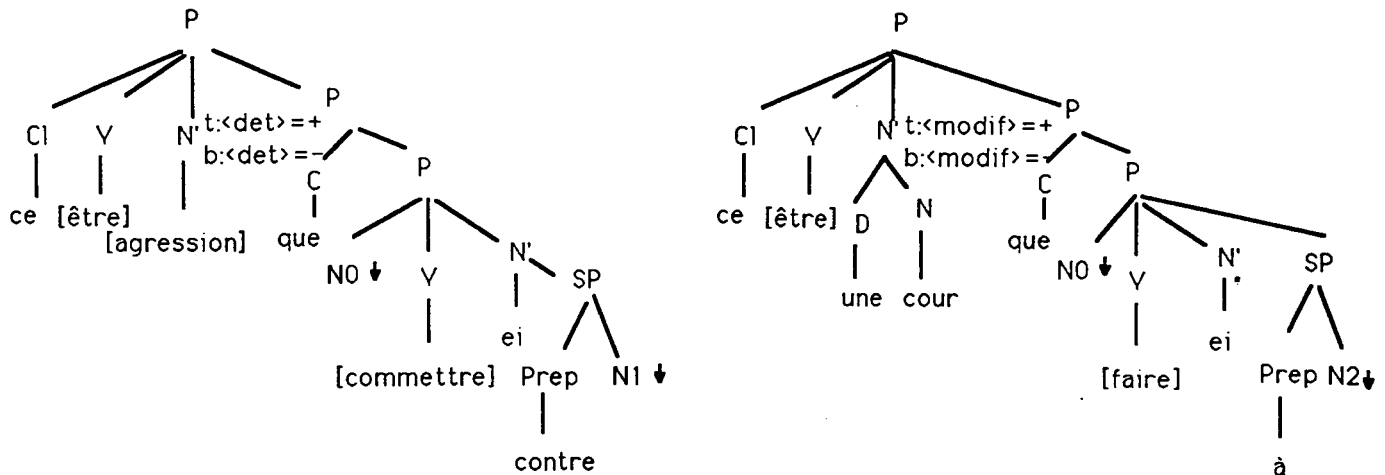


Figure 10

* extraction du complément de Npred

Il existe des restrictions à cette extraction. Comme le note L. Danlos 1980, l'extraction (et la question) est souvent interdite sur *de N₁* dans les phrases *N₀ être Prep Npred de N₁* :

- * *C'est de votre venue que Jean est dans l'attente*
- * *C'est de votre venue que Jean est dans l'ignorance*

Le seul autre exemple de Vsup de la même famille que nous ayons examiné est *procéder à*, avec lequel l'interdiction est la même. Il n'y a donc pas d'arbre correspondant à l'extraction de SP₁ dans la famille *n₀VPNpn₁*.

c) Les relatives

Comme pour les phrases simples, les relatives portant sur sujets et compléments donnent lieu à des arbres spécifiques. Une différence est qu'il y a trois relatives différentes correspondant à la fonction complément:

- (1) *Marie, contre qui Jean a commis une agression*
- (2) *l'agression contre Marie que Jean a commise*
- (3) *l'agression que Jean a commise contre Marie*

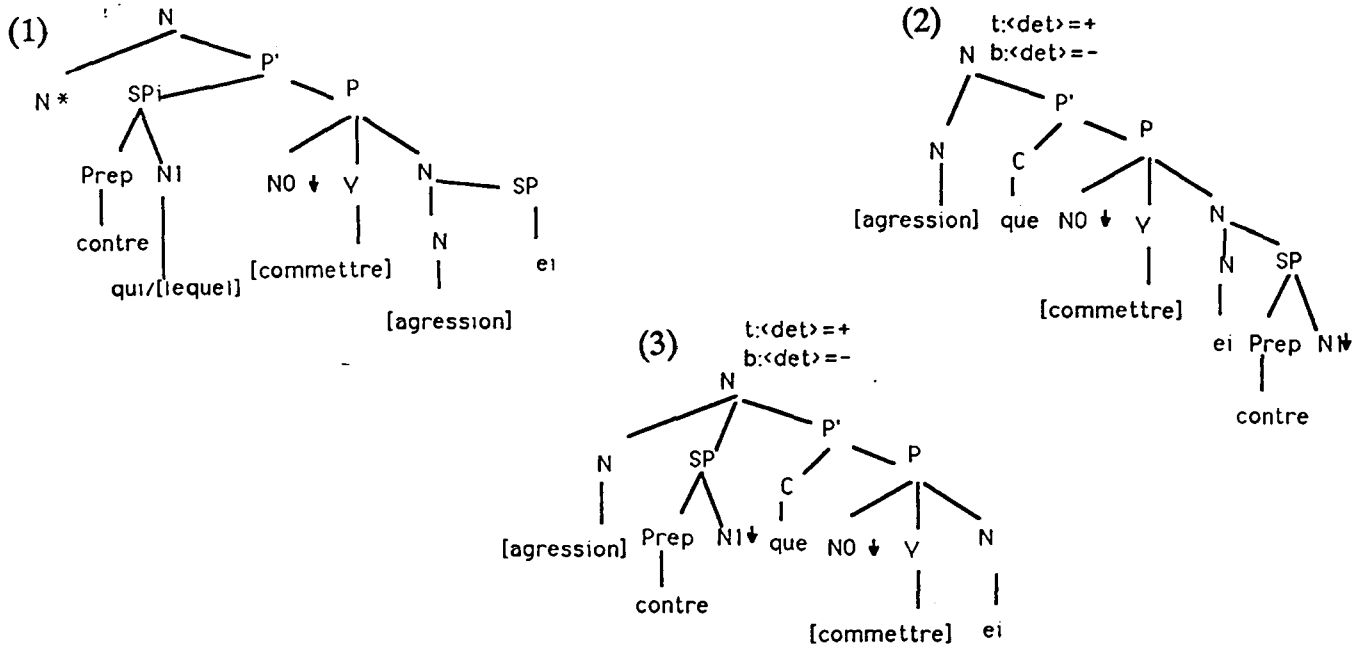
La phrase (1) est souvent considérée comme une violation de la contrainte de l'îlot nominal. Dans les constructions à verbe plein, elle est souvent exclue⁹:

** Marie, contre qui Jean a rapporté une agression*

La phrase (3) peut être dérivée de la phrase (2) par une règle d'extraposition du complément de nom. Néanmoins, cette règle est sujette à de nombreuses contraintes et s'applique plus difficilement dans le cas libre:

** L'agression que Jean a rapportée contre Marie*

Il est donc préférable d'en faire une structure explicite des constructions à Vsup. Nous avons donc les trois relatives suivantes pour n_0VNpn_1 :



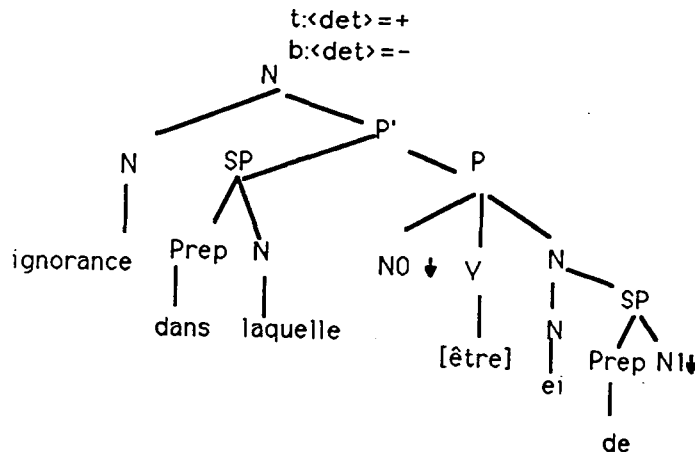
- (1) *Marie, contre qui Jean a commis une agression* (arbre auxiliaire)
- (2) *L'agression que Jean a commise contre Marie* (arbre initial)
- (3) *L'agression contre Marie que Jean a commise* (arbre initial)

Figure 11

⁹ Voir chapitre suivant pour plus de détails sur la représentation en TAG des contraintes d'extraction.

Lorsque le nom prédicatif appartient à l'antécédent de la relative, celui-ci fait partie de l'arbre élémentaire. Ces relatives ont donc trois têtes lexicales: Npred, Vsup et le pronom relatif (plus le déterminant de Npred s'il est figé). Il s'agit alors d'un arbre initial (correspondant à un SN qui sera lui-même argument d'une autre phrase) et non d'un arbre auxiliaire modifieur de N. (il y a donc bien parallélisme avec la structure du SN complexe).

Dans le cas d'une construction prépositionnelle à Vsup, l'antécédent est toujours nominal, même si le SP tout entier est antéposé (comme dans le cas libre):



N_0 être dans l'ignorance de N_1 = l'ignorance dans laquelle N_0 est de N_1

Figure 12

Les Npred à complément phrastique ont des relatives en *ce dont*:

Je vais aller en Espagne, ce dont je n'avais pas l'intention au départ
? Jean n' a pas été élu, ce dont Paul lui avait pourtant fait la promesse

Les sujets phrastiques donnent lieu à des relatives en *ce qui*. Il n'y a pas de relatives en *ce que* puisqu'il n'y a pas de complétives complément direct.

d) Les questions

Les questions sur les compléments sont représentées, comme dans le cas libre, par des arbres spécifiques, à N_i (ou SP_i) antéposés. Comme dans le cas libre, les mêmes arbres sont utilisés pour les questions directes et les questions indirectes (sauf *si*).

Il y a toutefois plusieurs différences. Rappelons que l'intérêt des arbres élémentaires étendus est de pouvoir appliquer les règles syntaxiques non seulement sur les compléments du verbe (au sens syntaxique) mais aussi sur tous les arguments de la structure verbale (au sens plus sémantique du terme). C'est-à-dire ici aussi bien sur le groupe (Npred + complément de nom) que sur le complément du Npred lui-même.

Nous ne faisons donc pas de différence syntaxique entre syntagme correspondant à un noeud intérieur et syntagme correspondant à une feuille. Tous les N_i (non dominés par un SP) et tous les SP_i sont des candidats à la question.

* question sur Npred

Une particularité tient au caractère figé du Npred, qui est tête lexicale de la structure. Il ne peut donc être remplacé par un pronom interrogatif:

* *Qu'as tu commis contre Marie ?*

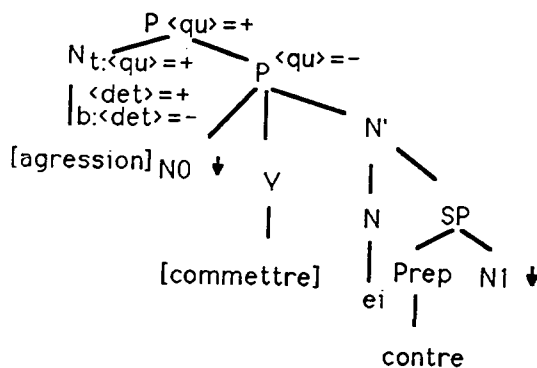
- *Une agression*

* *Qu'as-tu pris ?*

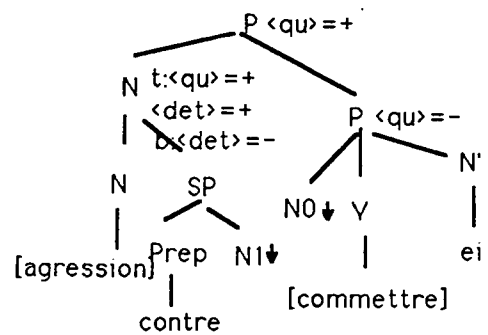
- *la décision de partir*

La seule question autorisée est celle où figure le nom prédictif accompagné d'un déterminant ou d'un complément de type *qu*-¹⁰. Le Ni interrogatif est donc étendu à Npred, et *quel* (ou tout autre déterminant interrogatif compatible: *combien de*, *quelle sorte de...*) devra s'y adjoindre (trait top <qu>=+), à moins que le complément de Npred (non extraposé) soit un interrogatif : *une agression contre (qui + quelle fille)*.

(1)



(2)



(1) *Combien d'agressions Jean a-t-il commises contre Marie ?*

(2) *Combien d'agressions contre Marie Jean a-t-il commises ?* ou *Une agression contre qui Jean a-t-il commise ?*

Figure 13

¹⁰ Ceci est une approximation de notre part, puisqu'il peut y avoir des cas où la question en *que* est plus heureuse, par exemple avec *N₀ avoir Le Nmaladie*:

Qu'est-ce que Jean a ?

- *la rubéole*.

Selon les expressions, cette commutation du déterminant de Npred avec *quel* (ou un autre interrogatif) est possible ou non :

Quelle décision as-tu prise ?

?* *Quelle décision d'aller à Bilbao as-tu prise ?*

Quand le déterminant zéro est obligatoire, la question n'est pas possible. Le trait top N.t.:<det> =- sur le nom prédicatif suffit à empêcher la sélection de l'arbre de la question, qui porte un trait <det> =+ incompatible. De même, quand le déterminant est entièrement figé: les familles $n_0\text{VDN}...$ n'ont pas d'arbres de questions. De même pour les arbres à complément phrastique réalisé des familles $n_0\text{VdN}(\text{ps1})$ et $n_0\text{Vdn}(\text{ps1})(\text{pn2})$ (Cf supra section 1.2.1 a).

Mais le déterminant peut n'être pas entièrement figé et la question être pourtant impossible:

Jean est dans (l'ignorance + une ignorance complète) des projets de Marie.

* *Dans quelle ignorance des projets de Marie Jean est-il ?*

Notons que *quel* est exclu même si le Npred n'est pas antéposé:

* *Jean est dans quelle ignorance des projets de Marie ?*

C'est une contrainte sur le déterminant de Npred en général qu'il faut noter dans le lexique et non une contrainte particulière à l'arbre de la question: N.t.:<qu> =-. Nous n'avons donc pas besoin d'un trait syntaxique spécifique.

* question sur *Prep N_I*

Elle donne lieu à un arbre spécifique par famille.

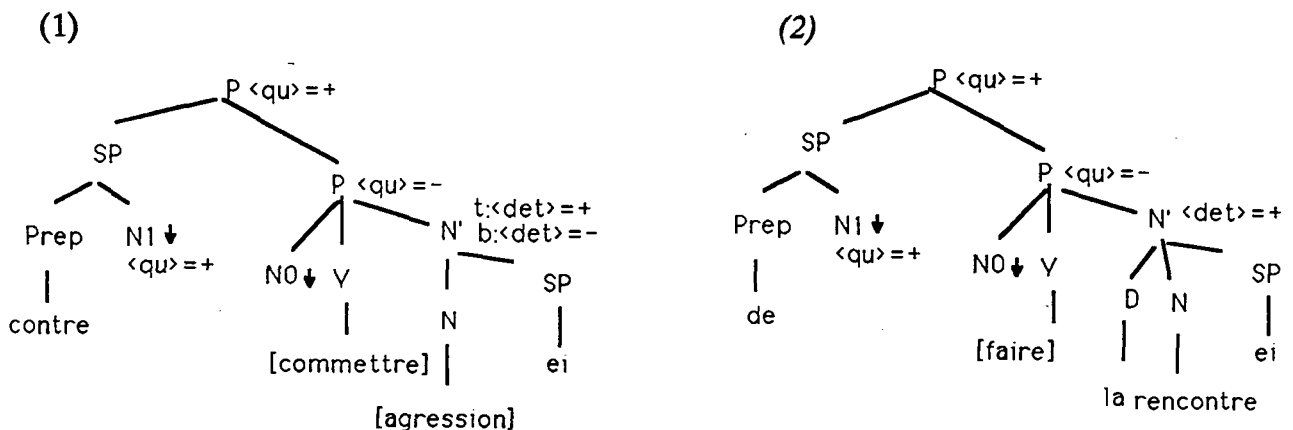


Figure 14

(1) *Contre quelle fille Jean a-t-il commis une agression ?*

(2) *De quelle fille Jean a-t-il fait la rencontre dernièrement ?*

Comme le note L. Danlos 1980, elle est souvent difficile lorsque Npred est dans un groupe prépositionnel:

- * *De quoi Jean est-il dans l'ignorance ?*
- * *De quoi Jean est-il dans l'attente ?*

En revanche, *quoi* est tout à fait possible en position complément dans la question "en écho" (structure déclarative):

- Jean est dans l'ignorance de quoi ?*
- Jean est dans l'attente de quoi ?*

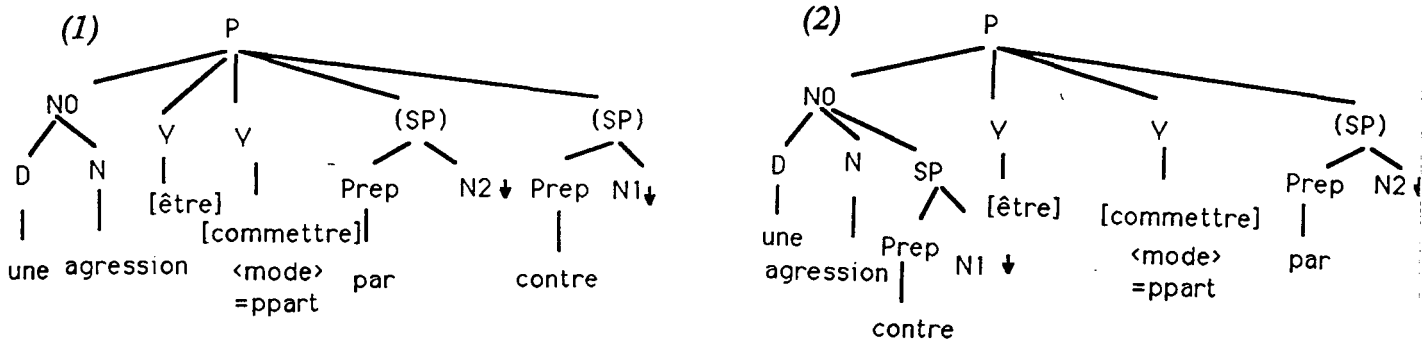
Il ne s'agit donc pas d'une contrainte sur le complément d'*ignorance* mais bien sur la structure interrogative elle-même. Il faut donc un trait *question* pour N₁.

e) Le passif

Les arbres du passif ont *être* et *par* parmi leurs têtes lexicales, comme dans le cas libre. Comme dans le cas des clivées et des relatives, deux constructions sont possibles:

- Une agression a été commise par Max contre Luc*
- Une agression contre Luc a été commise par Max.*

Elles sont représentées comme des structures différentes¹¹.



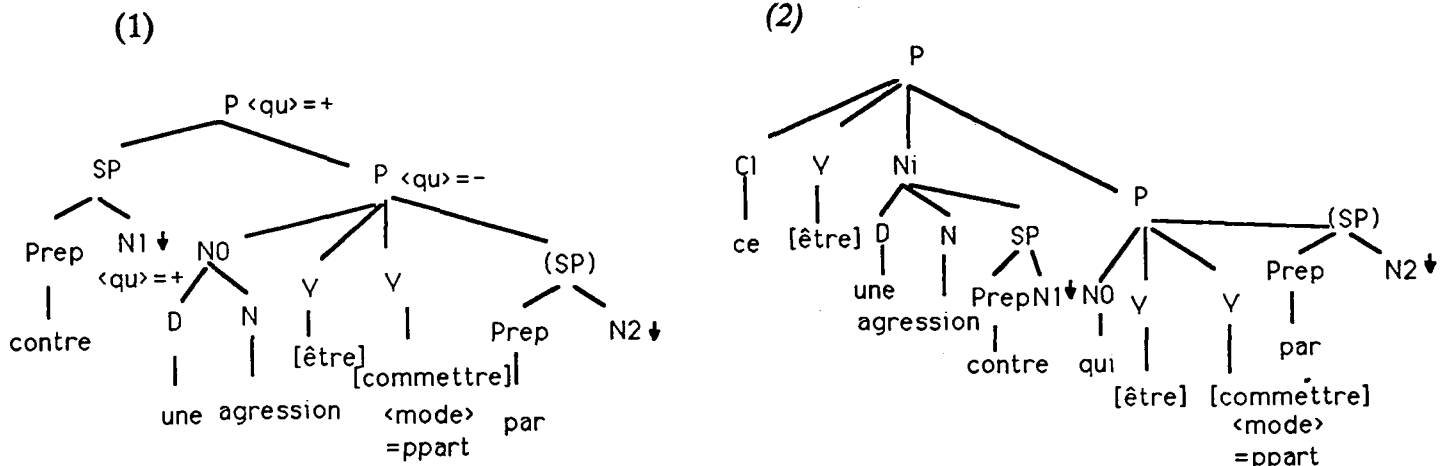
- (1) *Une agression a été commise par Jean contre Marie*
- (2) *Une agression contre Marie a été commise par Jean*

Figure 15

¹¹ Avec des règles de variation d'ordre des mots, ces deux arbres pourraient n'en faire qu'un seul. Ces règles portant sur les noeuds d'un même arbre élémentaire, la permutation serait impossible dans le cas d'un verbe "plein" puisque le complément de nom ne fait pas partie au départ de l'arbre phrastique de rapporter:

* *Une agression a été rapportée par Jean contre Luc.*

Des arbres passifs correspondent aussi à question+passif, relative+ passif ou clivée+ passif, comme pour les familles des verbes simples:



(1) *Contre quelle fille une agression a-t-elle été commise ?*

(2) *C'est une agression contre Marie qui a été commise par Jean.*

Figure 16

Comme l'a noté J. Giry-Schneider 1976, les contraintes sur le déterminant de Npred peuvent être différentes à l'actif et au passif.

Le déterminant zéro peut être possible à l'actif et non au passif:

Jean a pris (part + une part importante) à l'agitation

* *(Part + Une part importante) a été prise à l'agitation par Jean*

Mais c'est là une contrainte qui dépend de l'expression donnée, puisqu'il n'y a pas d'interdiction générale du déterminant zéro au passif:

On lui a rendu (hommage + un hommage éclatant)

(Hommage + Un hommage éclatant) lui a été rendu

On devra donc marquer dans l'entrée des expressions concernées une corestriction du type: *passif*=+ => N:<det>=+.

La situation inverse existe également où le déterminant zéro, impossible à l'actif, est possible au passif:

*Jean a pris (*E + la) décision de partir.*
(E + la) décision a été prise de partir (par Jean).
Max a fait la promesse que Marie viendrait.
(E + la) promesse a été faite (par Jean) que Marie viendrait.

Il semble que cette alternance soit régulière pour tous les Npred à complétive, et lorsque la complétive est extraposée. Nous avons donc noté cette possibilité directement dans les arbres passifs correspondants (familles $n_0\text{VDNps}_1$ et $n_0\text{VDNps}_1\text{pn}_2$).

Dans certains cas, également notés par J. Giry 1976, la phrase à verbe support n'existe qu'au passif:

* *On nous fait (E + l') interdiction de fumer en classe*
Interdiction nous a été faite de fumer en classe

* *La police nous a fait (une + E + l') invitation de faire cesser ce tapage nocturne*
Invitation nous a été faite par la police de faire cesser ce tapage nocturne

Nous utilisons le trait *passif* = + dans l'entrée lexicale correspondant à N_0 *faire interdiction de Pinf* (à N_2). Cette expression a donc la même famille que n'importe quel Vsup-Npred à complément phrastique + à N_2 mais y sélectionne seulement les arbres passifs. (C'est l'absence de trait *passif* qui dans le cas général permet la sélection des arbres actifs et passifs et signifie donc que le passif est possible).

f) L'extraposition

Comme pour le cas libre, l'extraposition est régulière avec les complétives sujets:

Il est du goût de Max de partir
Il est à l'avantage de Luc de savoir parler chinois (L. Danlos 1980)

Pour les sujets nominaux, il faut noter quand elle est interdite (*extrap* = -).
 L'extraposition + passif peut être meilleure que le passif seul:

On a porté plainte contre Luc
 ?* *Plainte a été portée contre Luc*
Il a été porté plainte contre Luc

Les traits *passif* et *extrap* étant indépendants, il faudra donc noter dans l'entrée lexicale des expressions concernées: *passif* = + => *extrap* = +.

g) La relation de symétrie

La relation de symétrie a été étudiée pour les noms par J. Giry-Schneider 1978 et 1987, D. Negroni-Peyre 1978 et L. Danlos 1980.

Elle pose un problème de localité à la plupart des théories linguistiques puisqu'ici le couplage d'arguments, leur permutation et leur coordination a lieu entre un sujet et un complément de nom:

Jean a eu une discussion avec Luc
Jean et Luc ont eu une discussion

Jean est en rivalité avec Luc
Jean et Luc sont en rivalité

Marc fait un pacte avec Bob
Marc et Bob font un pacte

Cette propriété montre bien que le nom prédicatif est syntaxiquement déterminant: les verbes *avoir*, *être en* ou *faire* ne sont pas par eux-mêmes symétriques.

Nous n'avons pas représenté les cas où, le nom prédicatif ayant deux compléments, la relation de symétrie joue entre ces deux compléments (J. Giry-Schneider 1978, table F7):

Jean fait le raccordement de la route avec la ville
Jean fait le raccordement de la route et de la ville

Cette construction permet en général également la préposition *entre*, que nous n'avons pas représentée:

Jean fait une raccordement entre la route et la ville

h) Relation de neutralité et causalité

Il n'y a pas de combinaisons Vsup-Npred neutres comme il y a des verbes neutres:

$N_0 V N_1 W =:$ *Jean blanchit les tissus*
 $N_1 V =:$ *Les tissus blanchissent*

$N_0 V_{sup} Det N_{pred} Prep N_1 =:$ *Jean fait le blanchiment des tissus* (Giry-Schneider 1978 table F2A)

$N_1 V_{sup} Det N_{pred} =:$ *Les tissus ont fait ((*le + *leur) blanchiment + ?* un blanchiment rapide)*

Sauf peut-être avec quelques variantes aspectuelles de Vsup:

Cela a plongé Max dans le désespoir
Max a plongé dans le désespoir

Une relation dont le format est proche mais qui est beaucoup plus générale met en jeu un changement de verbe :

N_0 Vsup1 Det Npred Prep N_1 =: *Les événements ont donné du poids à cet argument*

N_1 Vsup2 Det Npred =: *Cet argument a du poids*

Le premier verbe (Vsup1) est en fait un opérateur causatif de verbe support (M. Gross 1981) et non un véritable verbe support. Ainsi, il n'y a pas de SN complexe correspondant : * *Le poids de ces événements à cet argument.*

Si l'on reprend le critère de neutralité de J-P. Boons, A. Guillet, C. Leclère 1976a (Cf chapitre 3 section 3.4.6), il doit y avoir une factitive N_0 faire Vsup2 Det Npred Prep N_1 W parallèle à N_0 Vsup1 Det Npred Prep N_1 W, et de même interprétation (au contrôle extérieur près). Le critère de la factitive permet d'exclure des cas comme:

*Jean fait peur (E + *du loup) à Marie*
Marie a peur (E + du loup)
*Jean fait avoir peur (E + du loup) à Marie*¹²

Une relation similaire est étudiée par L. Danlos à propos des phrases en être prep:

N_0 Vsup1 N_1 Prep Npred =: *Jean a mis Marc au désespoir*
 N_1 être Prep Npred =: *Marc est au désespoir*

L'opérateur est "approprié" au verbe support, ou plutôt à la combinaison Vsup-Npred. Avec avoir, le causatif est en général donner, avec être Prep, c'est en général mettre (Cf L. Danlos 1980).

Nous considérons dans tous les cas la deuxième construction comme phrase de base, puisque c'est elle qui a les propriétés de phrase à support (formation du SN complexe). Nous notons donc un trait caus=donner ou caus=mettre qui ajoute les arbres phrastiques désirés (mais pas d'arbres de SN) dans les familles correspondantes:

¹² La relation en cause ici est différente (et plus proche du passif):

N_0 faire peur à N_1
 = N_1 avoir peur de N_0

(1) *extrap* = -
caus = -
passif = -

(2) *extrap* = -
caus = donner
passif = -

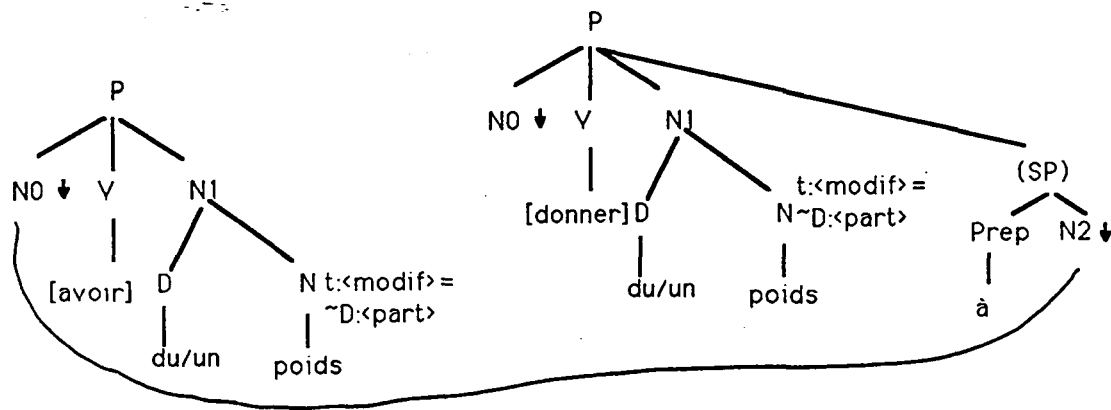


Figure 17

Par défaut *caus* = -

Nous n'avons pas incorporé dans notre système les nombreuses autres variations de constructions à Vsup notées par G. Gross 1990 et R. Vivès 1983, par exemple:

Jean a de l'admiration pour Luc
Luc a l'admiration de Jean

1.2.3. Relations sémantiques

Comme nous l'avons noté pour commencer, les phrases à verbe support ne sont pas par elles-mêmes ambiguës et reçoivent donc une seule analyse. Il pourrait y avoir ambiguïté avec des phrases simples puisque les Vsup peuvent généralement être également employés comme verbe plein (sauf *commettre*, *avoir lieu*...) mais leur construction (ou leurs restrictions de sélection) sont rarement les mêmes. Ainsi, le verbe *prendre* dans ses emplois pleins (*N0hum prendre N1 à N2hum* ou *N-hum prendre* =: *la crème prend*) ne peut avoir de complément abstrait, il n'y aura donc pas d'ambiguïté possible avec une construction à support comme:

N0hum prendre une décision (pas de cplt à *N2*)
N0hum prendre plaisir à (*N2abs* + *Pinf* + *ce que P*)

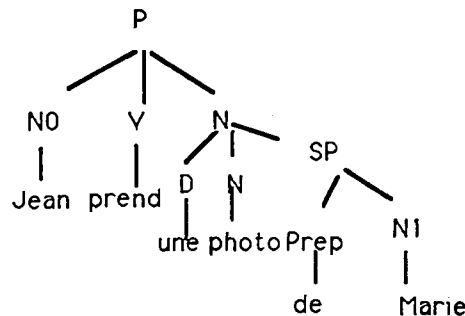
Le deuxième complément étant optionnel dans *N0hum prendre N1 à N2hum*, et le sujet de *N0* étant optionnel lorsque le Npred est tête d'un SN complexe, une phrase ambiguë serait par exemple,

Jean prend une photo de Marie

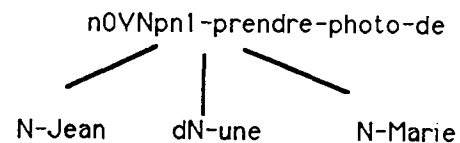
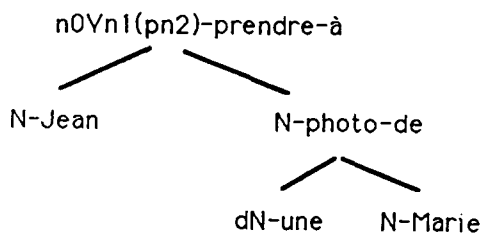
où soit le deuxième complément peut avoir été omis (= *Jean a pris une photo de Marie (de Nadar) sur la table*; verbe plein) soit il s'agit de la phrase simple complète (= *Jean photographie Marie*; verbe support).

L'arbre dérivé sera le même, c'est au niveau de l'arbre de dérivation que se fera la différence. En effet, l'agencement syntagmatique est le même mais les unités lexicales en jeu sont différentes :

arbre dérivé



arbres de dérivations



Phrase à verbe "plein"

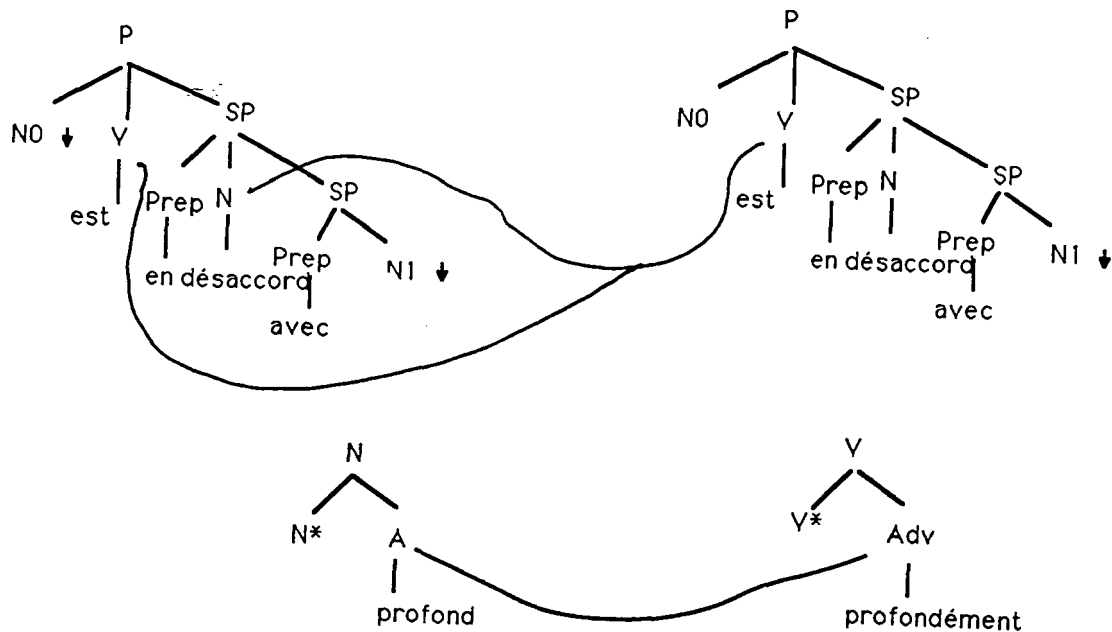
Phrase à verbe support

Le fait que l'arbre de dérivation soit la base de l'interprétation sémantique se retrouve également si l'on considère une autre particularité sémantique des phrases à verbe support: l'alternance entre adverbe et adjectif.

- Jean fait une confiance aveugle à Marie*
 = *Jean fait aveuglément confiance à Marie*
- Jean est en désaccord profond avec Marie*
 = *Jean est profondément en désaccord avec Marie*

Cette alternance est représentée par le fait que les deux phrases auront des dérivations synchrones. Les adjectifs et les adverbes en *-ment* sont liés, de manière générale, mais c'est une particularité des phrases à support que le verbe y soit lié au nom prédicatif et réciproquement. La relation de synonymie potentielle entre adjectifs et adverbes ne sera donc "activée" que dans ce cas:

arbres élémentaires



arbres de dérivation

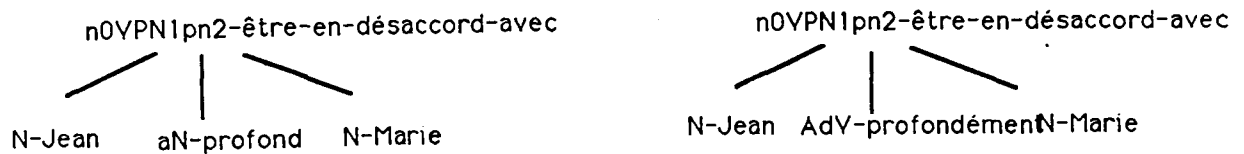


Figure 18

Lorsque le nom prédicatif est tête d'un SN complexe dans une phrase à verbe plein, cette alternance ne s'observera pas, puisqu'il n'y aura pas de liens entre le Npred et le verbe principal (qui n'appartiennent pas au même arbre élémentaire)¹³. Les phrases suivantes ne seront donc pas équivalentes, en dépit du lien existant entre *profond* et *profondément*:

- ≠ *Je désapprouve ce désaccord profond de Jean avec Marie*
 ≠ *Je désapprouve profondément ce désaccord de Jean avec Marie*

¹³ En fait, il faudrait aussi prévoir de telles alternances pour certaines phrases à verbe plein:

Jean a bu une rapide tasse de thé
 = *Jean a bu rapidement une tasse de thé*

C'est un trait des constructions non compositionnelles à têtes multiples d'avoir des liens entre certains de leurs noeuds de catégories différentes. De manière générale, la plupart des éléments têtes d'un arbre sont liés entre eux (sauf pour verbe et Prep).

L'on note également par de tels liens la relation éventuelle avec une phrase à verbe simple si Npred est dérivé d'un verbe.

Max prend la décision de partir
= *Max décide de partir*

Une telle relation n'est pas systématique et doit être notée à chaque fois (elle n'est pas déductible d'une simple analyse morphologique). De même pour les relations entre arguments:

Jean a de l'attachement pour Marie
Jean est attaché à Marie

et les restrictions spécifiques à chaque type de phrase:

(Jean + Cette histoire + Que Marie boude) offense Luc
*(Jean + *Cette histoire + *Que Marie boude) fait une offense à Luc* (J. Giry-Schneider 1978)

Dans d'autres cas, la restriction sur la construction à Vsup est en fait le reflet d'une véritable ambiguïté du verbe simple:

(Jean + l'immeuble) a sauté
*(Jean + *l'immeuble) a fait un saut*

Ce peut être un argument supplémentaire pour distinguer deux verbes *sauter*, et seul celui à sujet animé sera lié à la construction en Vsup.

L'équivalence éventuelle entre adjectif sur le Npred et adverbe sur le verbe plein s'exprime de la même façon que dans le cas où elle est interne à la phrase à support.

Voir en annexe 3 des exemples d'entrées de phrases à Vsup dans le dictionnaire TAG.



2. Les expressions figées

2.1. Discussion théorique

Les expressions figées ont été longtemps un domaine marginal et souvent négligé des études linguistiques, alors qu'elles sont plus nombreuses (dans la langue) et tout aussi fréquentes (dans le discours) que les schémas de phrases 'libres' (correspondant aux différents emplois des verbes simples), comme l'ont montré les études en cours au LADL depuis 1982, sur lesquelles nous nous appuyons. Un autre résultat essentiel de ces travaux est que ces expressions, figées lexicalement et sémantiquement, ont en général une syntaxe régulière et sont donc susceptibles d'une 'flexibilité' qui rend difficile leur représentation formelle comme leur reconnaissance automatique (E. Laporte 1990).

Nous nous sommes surtout intéressée aux phrases figées, qui présentent le plus de variations syntaxiques (A. Abeillé et Y. Schabes 1989, 1990). Suivant en cela M. Gross 1982, nous appelons 'figées' des phrases soit totalement idiomatiques, telles *N₀ casser sa pipe*, soit dont la combinatoire (par substitution synonymique) se bloque de façon imprévisible: *N₀ manger ses mots/ # N₀ manger ses paroles*¹⁴, ce qui inclut dans notre définition bon nombre de métaphores.

2.1.1. Aspects linguistiques

Les expressions idiomatiques ont longtemps servi d'argument à l'appui de telle ou telle transformation (passif, montée du sujet) mais sans faire l'objet d'études spécifiques (à l'exception de B. Frazer 1970 ou F. Newmeyer 1974 pour l'anglais).

La notion de structure argumentale a suscité un regain d'intérêt puisqu'il s'agit, pour reprendre l'expression de R. Cattell 1984, de "prédicats complexes". Diverses contraintes sur la formation de tels prédicats ont été proposées. Il a été suggéré que le figement ne concernerait que les groupes "frères".

A. Marantz 1984 avance ainsi que le figement se produit de façon privilégiée à l'intérieur du "groupe verbal". Ceci est vrai si l'on ne considère que des combinaisons du type "mettre fin à", "prendre corps", c'est-à-dire des combinaisons verbe + nom sans déterminant, où le nom est en position d'objet. Mais dès que l'on considère les phénomènes de figement dans toute leur étendue, cette idée d'une configuration privilégiée est à nuancer. L'inventaire systématique des phrases figées du français présenté par M. Gross 1982 montre en effet que sur un total d'environ 20 000 phrases, 3500 comportent sujet et verbe figés, contre environ 9000 avec verbe plus objet figé et 4300 à verbe et complément prépositionnel figé¹⁵.

N. Ruwet 1984 critique également l'idée que les sujets figés seraient (conformément à l'hypothèse ergative de L. Burzio 1981 sur les verbes intransitifs à auxiliaire *être*) des objets profonds. L'hypothèse peut s'appliquer à une expression comme *La moutarde monte au nez de N₁*, mais pas à de nombreuses autres expressions comme *La chance sourit à N₁* ou *Dieu a rappelé N₁ à lui*.

14 Nous employons le signe # pour marquer que l'interprétation idiomatique est impossible.

15 Il faut également ajouter 350 phrases à sujet, verbe et complément prépositionnel figés et 2000 à verbe, complément d'objet et complément prépositionnel figés.

L'on voit donc que la proportion de phrases à sujet figé est loin d'être négligeable et très comparable à celle des phrases à complément prépositionnel figé. S'il y a une configuration de figement privilégiée, c'est celle qui met en jeu verbe et objet, et elle ne recoupe donc pas la notion de groupe verbal. Il est également à noter que les trois types de phrases figées autorisent des discontinuités (et donc des défigements) comparables:

La moutarde risque de monter au nez de Max
Paul cassera bientôt sa pipe
Ce projet tombera complètement à l'eau

Nous traitons donc toutes les configurations de figement à égalité.

2.1.2. Aspects informatiques

Les expressions figées sont souvent mentionnées dans les expériences de traitement informatique les plus importantes, mais font généralement l'objet d'une représentation ad hoc et parcellaire.

a) Réanalyse et prétraitement

Un des premières méthodes a consisté à introduire une étape de prétraitement qui regroupe les mots considérés sous la catégorie syntaxique du mot libre qui leur correspondrait. Cela correspond à une représentation par réanalyse en linguistique théorique.

Comme le note E. Laporte 1990, il faut distinguer en analyse automatique le cas des mots composés qui sont en général très rigides de celui des phrases figées qui sont en général susceptibles d'un grand nombre de variations. Parmi les mots composés, les noms et adjectifs composés semblent les plus rigides; pour les déterminants, prépositions et adverbes composés, on note souvent des variations possibles de déterminants et des insertions de modificateurs.

Il peut y avoir réanalyse ou prétraitement pour les mots composés et cela est souhaitable afin de soulager l'analyse syntaxique proprement dite (voir par exemple, M. Silberztein 1990 pour les noms composés, D. Maurel 1989 pour certains adverbes figés)¹⁶. Pour les phrases figées, cela n'est pas le cas. Une réanalyse (ou un prétraitement) suppose en effet :

- qu'une catégorie syntaxique "naturelle" regroupe les différents éléments figés de l'expression (pas de figement sujet-verbe par exemple).
- que les éléments de l'idiome sont toujours continus (ou séparés par des éléments lexicaux facilement repérables sans analyse)¹⁷.

16 Nous avons néanmoins incorporé à notre grammaire certains déterminants figés (Cf chapitre 3) et certains adverbes figés (Cf chapitre 7) en leur assignant une structure syntaxique interne régulière.

17 Des systèmes à prétraitement comme celui d'A. Bérard-Dugourd 1990 peuvent autoriser des discontinuités à l'aide d'un joker dans l'entrée lexicale. Le problème est qu'il n'y a aucun contrôle sur la nature ou la longueur de l'élément inséré. Il y aura donc surgénération d'interprétations figées, et risque de blocage ultérieur de l'analyse.

Ce traitement est donc contredit par:

- l'existence de figement sujet-verbe comme *La grâce touche N₁* et d'expressions intrinsèquement discontinues comme *N₀ prendre N₁ en compte*.
- l'insertion régulière d'adverbes et autres modifieurs entre les éléments figés de l'expression, comme dans: *Jean a pris hier une sacrée veste*.
- l'application régulière des transformations syntaxiques, par exemple du passif ou de la relative, qui oblige à considérer qu'un SN figé, loin de "faire partie" du verbe, est bien un complément de celui-ci:

La glace a finalement été brisée par Luc

La veste que Jean semble avoir prise hier était bien méritée

Comme le note également I. Simatos 1986, une réanalyse supposerait que le groupe V-SN figé puisse se comporter comme un verbe intransitif eu égard à l'extraposition du sujet ou à la construction factitive. Or, il n'en est rien, comme le montre la comparaison de *casser sa pipe* et *mourir*:

Beaucoup de gens meurent ces jours-ci

Il meurt beaucoup de gens ces jours-ci

*L'alcool a fait mourir (Jean + * à Jean)*

Beaucoup de gens ont cassé leur pipe ces jours-ci

** Il a cassé sa pipe beaucoup de gens ces jours-ci*

** L'alcool a fait casser sa pipe (* Jean + ? à Jean)*

La reconnaissance des phrases figées suppose donc une analyse syntaxique complète de la phrase. L'inverse est également vrai, puisqu'il faudra reconnaître la séquence *faire mine de* pour accepter que le nom *mine* soit sans déterminant et l'expression *aller son train* pour accepter une phrase où *aller* est construit transitivement.

Des représentations ultérieures ont un peu raffiné cette première approche.

b) M. Salkoff 1979

M. Salkoff 1979 considère dans sa grammaire en chaîne ce qu'il appelle les "idiotismes" (p.79), c'est-à-dire des combinaisons verbe-objet figées du type : *rendre hommage à, avoir intérêt à, faire grand cas de*, où l'objet est sans déterminant, et qui mêlent en fait aussi des phrases à verbe support.

Il distingue les expressions passivables qu'il considère comme analysables, et les autres. Les premières sont reconnues comme n'importe quelle chaîne verbe-objet, avec éventuellement des contraintes particulières sur leur déterminant. Elles n'apparaissent comme figées à aucune étape de l'analyse et ne figurent pas dans le dictionnaire.

Les autres sont considérées comme un verbe, et entrées comme telles dans le dictionnaire. Pour celles où l'insertion d'un adverbe ou d'une incise est possible, l'entrée lexicale le prévoit et forme en fait une sous-chaîne: "avoir adverbe intérêt". Il ne semble

pas que dans ce cas, l'ajout d'un adjectif ou d'un déterminant sur le nom soit prévu. Il est également curieux de faire du passif le seul critère de la fixité (ou non fixité) d'une expression, puisque d'autres transformations peuvent très bien s'appliquer à des expressions non passivables : *quel intérêt aurait-t-il à faire ceci ?*

D'une façon générale, une bipartition simple entre phrases idiomatiques complètement figées et phrases idiomatiques à variation syntaxique ne se justifie pas. Peu d'expressions appartiennent au premier groupe. Pour le second, les règles syntaxiques s'appliquant différemment aux différentes expressions, il faudrait distinguer pratiquement autant de degrés de flexibilité qu'il peut y avoir de règles¹⁸.

c-) Les grammaires d'unification

Les grammaires d'unification semblent apporter une souplesse tout à fait utile pour la représentation des expressions figées. Et l'on devrait s'étonner qu'elles ne s'en soient pas plus préoccupé. C'est qu'en fait la représentation des phrases figées par traits lexicaux pose autant de problèmes qu'elle en résout.

De manière générale, si l'on veut introduire en même temps (dans une même unité) tous les éléments figés d'un idiome avec des règles de réécriture de type hors contexte, il faudrait "mettre à plat" les expressions figées et donc perdre leur structuration en constituants internes. Ainsi pour introduire en même temps les trois items de l'expression *briser la glace*, il faudrait avoir une règle de réécriture : $P \rightarrow NP V Det N$, c'est-à-dire perdre le constituant objet, pourtant utile pour expliquer l'existence d'un passif par exemple: *La glace a finalement été brisée*.

L'autre solution, qu'utilisent les formalismes GPSG et LFG¹⁹, est de garder les mêmes règles de réécriture que pour les structures libres, mais d'utiliser un trait <Form> pour contraindre la valeur lexicale des différents éléments de l'expression.

Il s'agit alors de produire (ou reconnaître) les expressions figées par les règles générales de la grammaire contraintes par des traits vérifiant la valeur lexicale de tel ou tel élément.

Ainsi, en LFG, le verbe *briser*, dans son entrée idiomatique, sélectionnera bien une structure <sujet, objet> (comme le *briser* littéral) mais avec un trait (Obj Form) =_c 'glace', et (Obj Spec) =_c la. Notons qu'il s'agit de contraindre les bonnes valeurs lexicales mais aussi de noter les propriétés syntaxiques et sémantiques particulières de l'expression (c'est-à-dire de distinguer la combinaison obtenue de l'expression littérale éventuellement identique). C'est pourquoi J. Bresnan 1982 réserve le trait <Form> aux éléments lexicaux en tant qu'ils appartiennent à une expression figée. Ce trait a donc un usage particulier, comme le marque l'indice _c supra. Sa valeur ne va pas être contrainte par unification, car tous les mots qui ne sont pas marqués pour ce trait seraient des

18 Ce qui ne veut pas dire que certaines règles ne sont pas corrélées, par exemple, la question en *quel* et la relative (voir infra).

19 Ces formalismes distinguent en fait deux types d'idiomes: ceux qu'ils considèrent comme complètement figés, du type *NO casser sa pipe* ou *NO perdre les pédales* et ceux qu'ils considèrent comme plus flexibles, du type *NO briser la glace*. Pour les premiers, ils opèrent une réanalyse, c'est-à-dire considèrent la séquence comme un verbe morphologique. Avec les mêmes problèmes que ceux mentionnés ci-dessus. Nous ne nous intéressons ici qu'à leur traitement des expressions 'flexibles'.

candidats potentiels, l'unification leur ajoutant ce trait Form, et le système ne contraindrait plus rien. Il s'agit en fait d'une simple vérification d'identité: le mot candidat doit porter ce trait (Form), et la contrainte ne sera pas satisfaite si, sans porter de trait à valeur incompatible, il ne le porte pas déjà explicitement.

Ce système conduit donc à dupliquer tous les mots qui peuvent appartenir à une expression libre ou à une expression figée. Les phrases pourront donc être éventuellement analysées comme ambiguës puisque l'interprétation littérale sera syntaxiquement différente de l'interprétation idiomatique, et les propriétés syntaxiques spécifiques à l'interprétation idiomatique pourront être marquées. Cette duplication empêche de rendre compte du fait que les expressions figées sont globalement formées du même matériel lexical que les expressions libres.

Un deuxième problème est que les éléments d'une même expression peuvent être séparés par une longueur non bornée. Il faudrait donc, pour propager ce trait, ajouter un mécanisme du type Incertitude fonctionnelle ou trait Slash. Le problème est que ce type de mécanisme est prévu pour des valeurs catégorielles et non lexicales (Cf chapitre 5, p. 249).

Le problème le plus sérieux est d'ordre sémantique: quel est le sens de *glace* et de *la* à combiner avec le sens de *briser* idiomatique ? La réponse de J. Bresnan 1982 est que les éléments marqués du trait Form ont forcément un sémantisme vide²⁰. Il y a donc (au moins) deux entrées pour *glace* (et deux pour *la*), l'une pour le sens littéral et sans trait (Form), l'autre avec un trait (Form) = 'glace' et un sémantisme vide. Tout le sens de l'idiome est donc contenu dans le verbe en tant qu'il sélectionne les autres éléments figés. La sémantique de LFG ne peut donc être strictement compositionnelle c'est-à-dire monotone.

Une autre possibilité, toujours pour les expressions variables syntaxiquement, suggérée par G. Gazdar et al. 1985 dans le cadre GPSG (mais non développée) est de considérer chaque élément figé comme porteur d'une partie du sens de l'expression, et celle-ci comme sémantiquement compositionnelle. Il s'agit, reprenant une hypothèse de T. Wasow et al. 1982, de relier propriétés syntaxiques et propriétés sémantiques des idiomes. Les arguments avancés ne semblent convaincants que pour très peu d'exemples (en particulier quand un élément de l'expression peut être repris par un pronom dans une phrase suivante). Dans le cas général (et contrairement à F. Newmeyer 1974), les propriétés syntaxiques des idiomes ne se laissent pas déduire de leurs propriétés sémantiques, comme l'indiquent les nombreux contrastes existant entre un idiome et sa paraphrase libre. Par exemple *N₀ louter le coche* se passive difficilement, alors que le passif de son équivalent non figé est parfaitement naturel:

Jean a loupé le coche
 * *Le coche a été loupé (par Jean)*
Jean a loupé cette superbe occasion
Cette superbe occasion a été loupée (par Jean)

²⁰ Cela permet aussi de ne pas avoir autant d'entrées pour un mot qu'il y a d'expressions figées différentes dans lesquelles il peut entrer.

Un deuxième problème est qu'il faudrait distinguer autant de degrés de compositionnalité qu'il y a de combinaisons de règles syntaxiques qui peuvent s'appliquer. En pratique, un tel "détour" sémantique est une complication inutile et nous préférons un marquage direct des règles syntaxiques.

Enfin, le système de G. Gazdar et al. aboutit à multiplier considérablement le nombre d'entrées lexicales, de nombreux items pouvant appartenir à de nombreuses expressions différentes (ex *manger, pain* ...) et à un système de co-sélection (par des traits ou par les "fonctions partielles" proposées par T. Wasow et al. 1982) particulièrement ingérable.

Le domaine de localité étendu du formalisme TAG permet d'introduire en même temps (dans une même unité) tous les éléments figés d'un idiome, tout en respectant par ailleurs la structure syntaxique interne de l'expression qui peut être fort complexe.

Si l'objectif d'un formalisme grammatical est bien de représenter les phénomènes linguistiques par la combinaison strictement compositionnelle, c'est-à-dire monotone, de structures élémentaires, les grammaires TAG sont donc doublement monotones, de ce point de vue:

- toute combinaison entre unités syntaxiques est une combinaison entre items lexicaux (la grammaire est lexicalisée et il n'y a pas de tête vide)
- toute combinaison entre unités syntaxiques est une combinaison entre unités sémantiques (il n'y a pas d'arbres élémentaires "vides" sémantiquement).

Les éléments qui seraient dans d'autres cadres considérés comme vides sémantiquement peuvent être traités de deux façons:

- ils figurent dans le dictionnaire syntaxique mais uniquement dans des entrées à tête multiples (comme *escampette* ou *haro* existant uniquement dans un (ou plusieurs) idiomes, verbes employés uniquement comme verbe support Cf *commettre*)
- ils figurent directement dans les arbres élémentaires des familles. Ce sont des mots outils, ils ne figurent pas dans le dictionnaire syntaxique (ni morphologique). Par exemple, les pronoms relatifs ou *être* auxiliaire passif.

2.2. Représentation en TAG

2.2.1. Arbres étendus et têtes multiples

Nous représentons les expressions figées par des arbres élémentaires étendus, c'est-à-dire en général plus "profonds" que les arbres élémentaires vus jusqu'ici, et dont la tête lexicale est constituée de tous les items figés de l'expression (Cf A. Abeillé 1988a et A. Abeillé, Y. Schabes 1989). Les arguments libres d'un verbe sont contraints par la catégorie du noeud qui leur correspond (et par des traits morphologiques et sémantiques). Les arguments figés ne sont pas introduits selon les mêmes principes: ils sont directement spécifiés dans l'entrée lexicale de l'expression (et appartiennent à la tête lexicale de celle-ci).

Voici par exemple les arbres élémentaires associés aux expressions N_0 *tomber à l'eau*, N_0 *perdre les pédales* et *la moutarde monte au nez de N_1* :

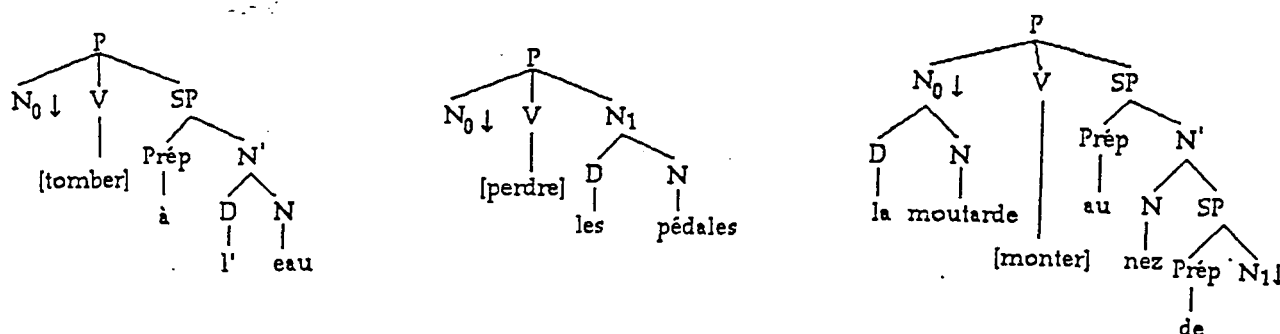


Figure 19

Les éléments fixes de l'idiome sont considérés comme définissant la tête de l'arbre élémentaire correspondant, qui a donc une tête multiple. Nous avons donc une entrée lexicale par expression idiomatique dans le dictionnaire, et nous ne créons pas d'entrées séparées pour chacun de ses éléments (comme en GPSG). Nous aurons par exemple (avec les notations vues au chapitre 3)²¹:

[tomber],V; à,P1; l',D1; eau,N1 : n0VPDN1.

[perdre],V; les,D1; pédales,N1 : n0VDN1.

La structure de ces arbres est syntaxiquement régulière, c'est-à-dire que c'est la même que celle qu'on obtiendrait en combinant les arbres élémentaires associés à chaque élément lexical indépendamment (chacun étant alors pris dans son sens littéral). Il s'agit donc ici de considérer comme des arbres élémentaires des structures qui jusqu'à présent (pour les expressions "libres") ne pouvaient être que des arbres dérivés.

C'est une première définition du figement dans notre système: un arbre dérivé devient un arbre élémentaire. Et l'on pourrait considérer un système qui, selon une vue dynamique du processus de figement ainsi conçu, engendrerait les arbres figés à partir des structures élémentaires²².

Une deuxième définition, qui est en fait plutôt une définition de la non compositionnalité, est que les têtes lexicales des structures élémentaires sont généralement liées les unes aux autres dans l'optique des TAG autosynchrones: un ajout

21 L'ordre dans lequel sont indiquées les co-têtes lexicales importe peu. Nous donnons ici des entrées simplifiées. En particulier, nous utilisons en fait les lemmes pour les noms et non les formes fléchies; voir *infra* p. 209.

22 Un tel système ne serait pas sans rappeler les grammaires W, définies dans le cadre des travaux sur le langage ALGOL: l'idée était de considérer que les phrases produites par une grammaire sont les règles d'une autre grammaire (Van Wijngaarden 1969, Greibach 1974). Bien sûr, il resterait à rendre compte des propriétés syntaxiques spécifiques des phrases figées par rapport à leurs équivalents littéraux, et il ne semble pas y avoir de procédure automatique pour cela.

à l'une équivaut à un ajout à l'autre (par exemple un adverbe sur le verbe et un adjectif sur le complément figé, Cf supra p. 197 et infra p. 224.).

Le formalisme TAG exprime donc bien le fait que figement et non compositionnalité sémantique sont liés.

Dans certains cas, l'arbre n'est pas plus étendu que l'arbre sélectionné par un verbe libre. La seule différence est que plus d'une feuille porte un terminal tête. Par exemple, pour les verbes à sujet impersonnel figé, comme *il faut que P* ou *il s'agit de/que P* :

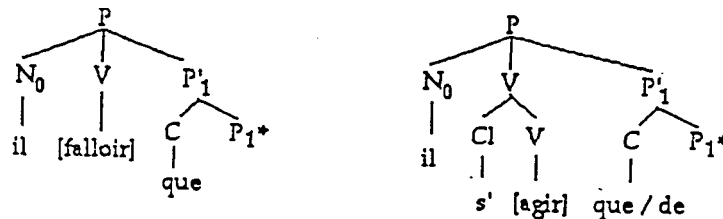


Figure 20

Il y a plusieurs avantages à considérer tous les éléments d'une expression figée comme appartenant à un même arbre élémentaire. Ces éléments formant une seule entrée lexicale dans le dictionnaire syntaxique, on n'a pas besoin de créer d'entrées lexicales artificielles pour les items n'existant pas en dehors d'une expression donnée, comme *escampette* ou *haro* qui n'apparaissent respectivement que dans *N0 prendre la poudre d'escampette* et *N0 crier haro sur la baudet*.

Un autre intérêt est de pouvoir exprimer directement des phénomènes de dépendance entre des éléments figés situés à des niveaux syntagmatiques différents (Cf infra 2.2.2). Il est aussi de pouvoir encoder directement dans l'entrée lexicale complexe les propriétés syntaxiques particulières de l'expression (Cf infra 2.2.4).

Les variations sur les têtes lexicales de l'expression sont représentées par la disjonction: *N0 perdre la tête/boule*, *N0 marcher/aller sur les brisées de N1*. Ces disjonctions peuvent associer des mots dont le sens est habituellement éloigné ce qui confirme bien le caractère non compositionnel de l'expression.

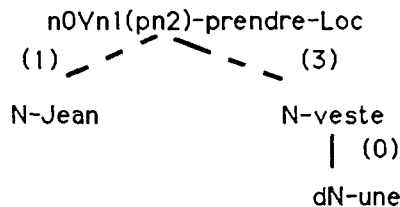
Des variations de déterminants sont également observées. Le déterminant d'un nom figé peut être lui aussi figé (*N0 demander la lune*) ou complètement libre (*N0 brûler un/trois/ce... feu(x) (rouge(s))*). Lorsque la variation du déterminant est entraînée par l'adjonction d'un modifieur, sa représentation est la même que dans le cas des phrases à verbe support.

N0 donner (le ton + un ton nouveau) à N1

Mais l'intérêt des arbres étendus est aussi sémantique. Un arbre élémentaire étant par définition une unité sémantique, la non compositionnalité des expressions figées est

donc définitoire en TAG. Ceci est manifeste si l'on considère l'arbre de dérivation associé à une phrase comprenant ce type d'expression. Il sera beaucoup plus réduit que l'arbre de dérivation d'une structure libre similaire. Ainsi, la phrase *Jean prend une veste* sera représentée par un seul arbre dérivé, qu'il s'agisse de l'interprétation idiomatique ou littérale, mais par deux arbres de dérivation distincts, selon l'interprétation:

sens littéral



sens figé

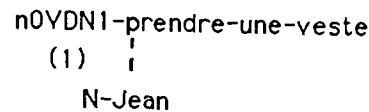


Figure 21

Il n'y aura bien sûr pas d'ambiguïté, et donc un seul arbre de dérivation, si l'un des mots de l'expression n'existe pas par ailleurs, ou n'existe qu'avec une construction différente. Ainsi *Jean prend la poudre d'escampette* ou *Jean va son train* ne seront pas analysées comme ambiguës.

2.2.2. Phénomènes d'accord

Les phrases figées sont en général composées du même matériel lexical que les phrases libres, et les mots y ont en général même catégorie. Nous représentons cette communauté de forme en utilisant les mêmes ensembles de formes fléchies dans les deux cas. Mais la morphologie des éléments figés de l'expression peut ne réaliser qu'un sous-ensemble des formes disponibles dans le cas libre. Par exemple, le nombre des noms figés devient souvent invariable:

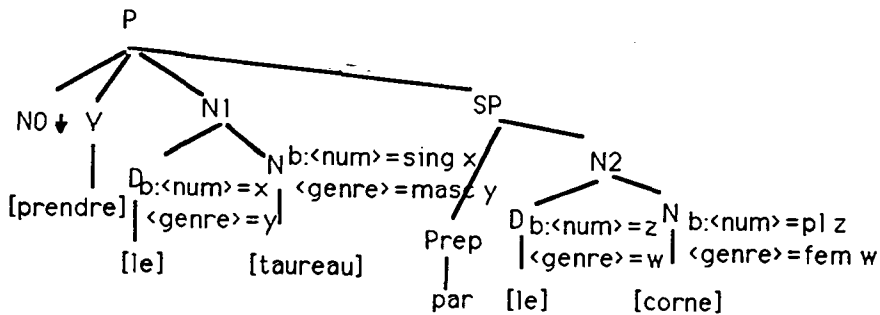
Jean a pris (le taureau + #les taureaux) par (#la corne + les cornes).
*(La moutarde + #Les moutardes (monte + #montent) (au + *aux) nez de Paul.*

Ces formes seront donc contraintes dans l'entrée syntaxique de l'expression. Nous pourrions contraindre de la même façon les déterminants ou les verbes qui leur correspondent. Nous préférons laisser ce soin à la contrainte générale d'accord inscrite directement dans la structure, manifestant ainsi que les expressions figées obéissent aux mêmes règles que les expressions libres²³.

[prendre], V; [le], D; [taureau], N₁ {N: <num> = sing}; par, P; [corne], N₂ {N: <num> = pl} :
 n₀VDN₁RN₂{N₀: <hum> = +}.

²³ Nous aurions bien sûr pu contraindre lexicalement le déterminant (ou le verbe) et laisser à la règle d'accord le soin de contraindre la forme du nom.

avant unification



après unification

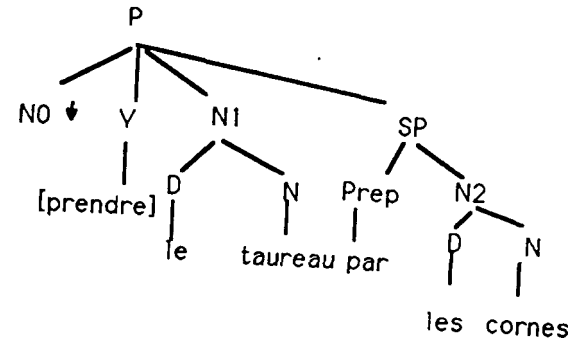


Figure 22

Les expressions figées présentent également de nombreux phénomènes de dépendance qui amènent à élargir la notion d'accord. Il en est ainsi des cas où un déterminant possessif attaché à un élément figé est obligatoirement "coréférent" avec un argument libre de l'expression²⁴:

N₀ casser Poss₀ pipe
Jean casse sa pipe
Tu casseras ta pipe un jour toi aussi

N₀ remettre N₁ à Poss₁ place
Tu as remis Marie à sa place
Jean a remis les étudiants à leur place

Le déterminant possessif peut même être enchâssé à l'intérieur d'un SN argument:

N₀ garder à N₁ un chien de Poss₀ chienne
N₀ rendre à N₁ la monnaie de Poss₁ pièce

L'on ne peut prédire à partir de la seule forme de l'expression avec quel argument se fera l'accord, comme en témoigne le contraste existant entre les structures parallèles:

N₀ prendre Poss₀ courage à deux mains
N₀ dire Poss₂ quatre vérités à N₂

²⁴ Les indices sont mis dans les exemples pour faciliter la lecture. Nous ne les utilisons pas dans notre représentation.

Il faut donc marquer ces phénomènes d'accord dans l'entrée lexicale des expressions (c'est-à-dire dans leurs arbres élémentaires). Nous utilisons deux traits <Poss-num>²⁵ et <Poss-pers> spécifiques :

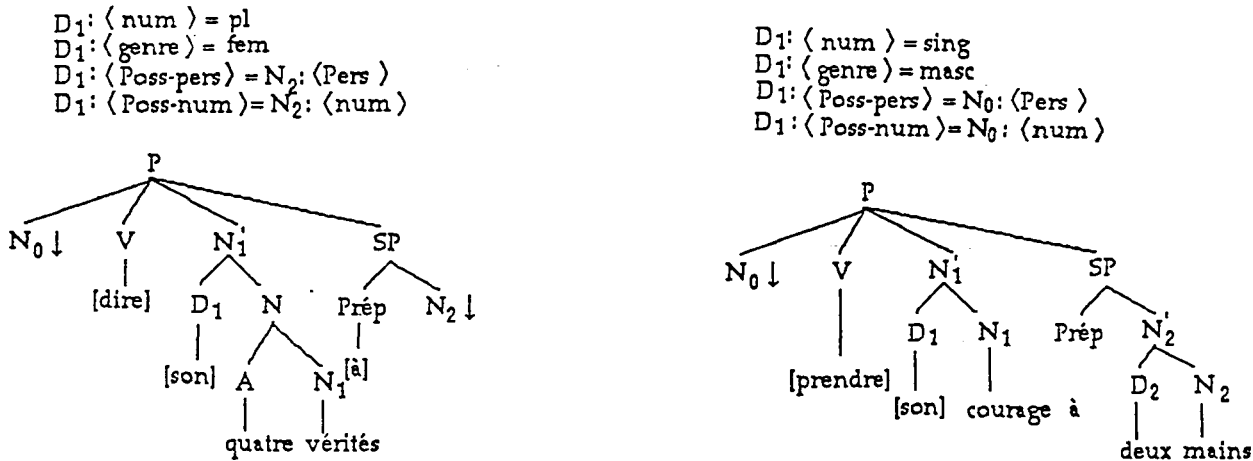


Figure 23

Dans une grammaire basée sur des règles hors contexte, il faudrait recopier les traits d'accord autant de fois qu'il y a de noeuds intermédiaires, car les éléments dépendants n'appartiennent pas à la même règle de réécriture.

Dans d'autres cas, le possessif n'est pas coréférent à un argument de l'expression, et il alterne régulièrement avec un complément en *de N* "libre" :

*Luc marche sur les brisées (*E + de Jean).*

*Luc marche sur ses brisées (E + *de Jean).*

On utilise un trait <Poss> qui est associé au noeud N concerné, et qui est un trait top. Le nom figé étant marqué <Poss> = - en bottom, l'adjonction d'un élément possessif est obligatoire, et celui-ci sera soit un déterminant possessif soit un groupe en *de N* (mais pas les deux, puisque chaque arbre auxiliaire porte <Poss> = - à son noeud pied²⁶):

²⁵ Ce trait est différent du trait <num> qui est le nombre de l'article possessif lui-même. Par exemple *ses* : <num> = pl, <Poss-num> = sing; *leur* : <num> = sing, <Poss-num> = pl.

²⁶ L'on peut noter que cette contrainte n'est pas spécifique aux phrases figées. De fait, cette alternance entre déterminant et modifieur possessifs s'observe de la même façon pour les phrases libres (Cf chap. 6).

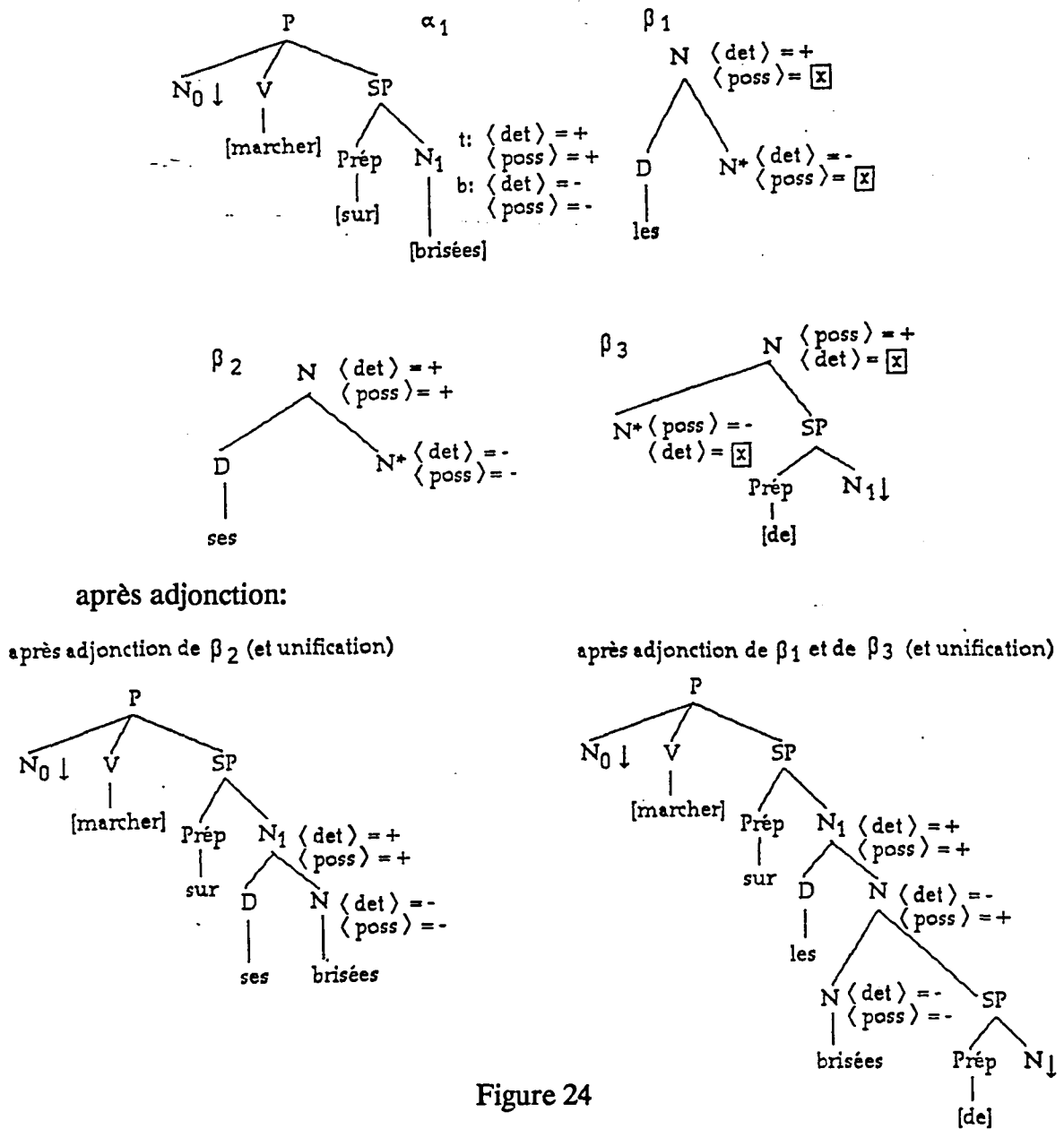


Figure 24

Dans certains cas, notés par L. Danlos 1980 et M. Gross 1989b, le complément *de N* est mieux accepté que le déterminant possessif. Il peut s'agir d'expressions où le nom est en fait plutôt à analyser comme un déterminant complexe (et le possessif est interdit comme dans les phrases libres; Cf D. Godard 1988):

Jean convoque le ban et l'arrière-ban (E+ des collègues).

* *Jean convoque leur ban et (E+ leur) arrière-ban.*

Dans ce cas, le déterminant défini et le complément *de N* sont figés et font directement partie de l'expression.

Souvent, il s'agit de noms de parties du corps, et l'on observe alors une alternance entre *de N* et un clitique datif (comme d'ailleurs pour les phrases libres; Cf R. Kayne 1976):

Jean casse les pieds de Marie
**Jean casse ses pieds*
Jean lui casse les pieds

Dans ce cas, le déterminant défini fait également partie de l'expression, qui comporte en plus une disjonction de traits, ici: $N_1.t:\langle \text{Poss} \rangle = +$ / $V.t:\langle \text{datif} \rangle = +$, les clitiques étant représentés par des arbres auxiliaires s'adjoignant sur le verbe, dont ils changent le trait $\langle \text{datif} \rangle$. Seul un groupe en *de N* peut s'adjoindre en N_1 , puisqu'il y a déjà un déterminant.

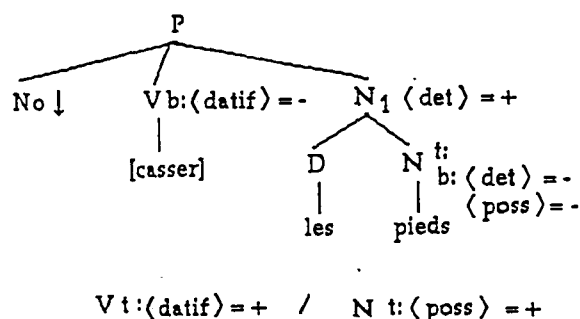


Figure 25

Enfin, les expressions qui autorisent à la fois une triple alternance entre clitique datif, *de N* et déterminant possessif présentent la même équation de traits mais sans le déterminant défini déjà dans l'arbre:

*Jean casse du sucre sur le dos (*E + de Luc).*
*Jean casse du sucre sur son dos (E + *de Luc).*
Jean lui casse du sucre sur le dos.

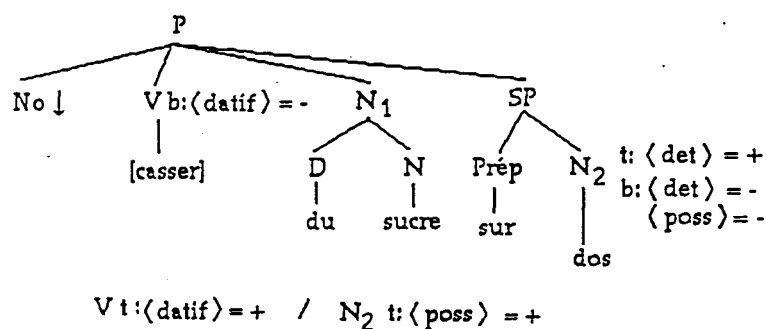


Figure 26

Les phénomènes d'alternance entre complément verbal en *à N* et clitique datif, ou entre complément nominal en *de N* et déterminant possessif, se règlent de la même façon pour les phrases libres. (voir chapitre 6)

2.2.3. Phénomènes de discontinuités et variations d'ordre des mots

L'association de structures syntaxiques régulières aux expressions figées permet une représentation naturelle de leur flexibilité syntaxique. Celle-ci se manifeste notamment par des phénomènes de discontinuités.

Comme nous l'avons montré dans A. Abeillé et Y. Schabes 1990, c'est la topologie même de l'arbre qui permet les discontinuités. Celles-ci peuvent faire partie de la définition même de l'expression, qui est "intrinsèquement" discontinue comme dans *N₀ prendre N₁ en compte* ou *N₀ passer N₁ par les armes* :

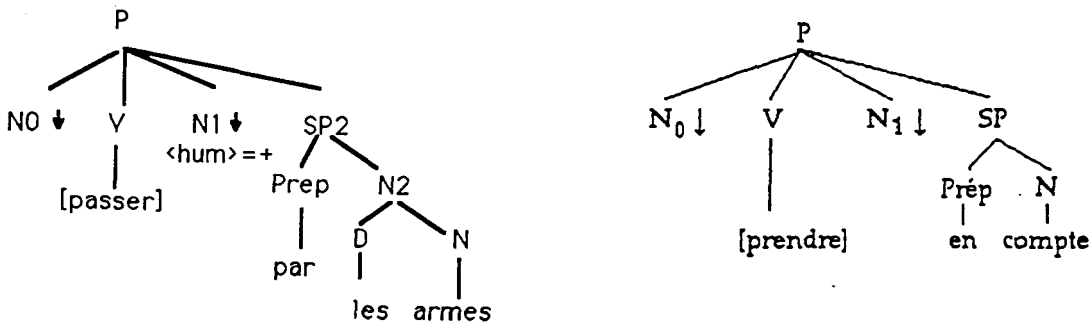


Figure 27

Les discontinuités peuvent venir également de l'adjonction régulière d'arbres auxiliaires, c'est-à-dire de modifieurs, aux noeuds intérieurs de l'expression. Cette adjonction peut être obligatoire, comme dans les expressions à déterminant libre *N₀ attendre Det enfant(s)* ou *N₀ griller/brûler Det feu(x)*, qui portent en N₁ un trait top <det> = + incompatible avec le trait bottom <det> = - de ce même noeud, rendant obligatoire l'adjonction d'un déterminant:

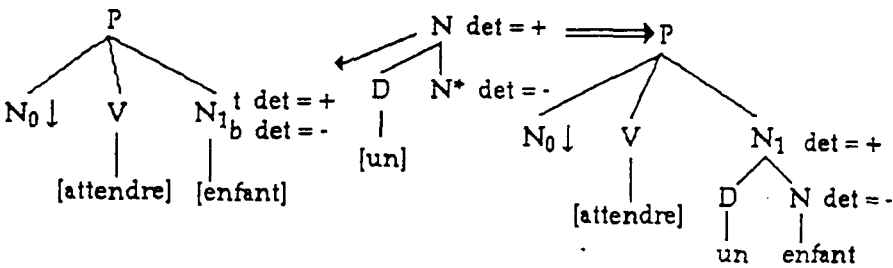


Figure 28

L'adjonction est souvent optionnelle, pour les adverbes, les auxiliaires ou les autres modifieurs, comme par exemple: *Jean a pris une sacrée veste*. Elles peuvent être de longueur non bornée, comme par exemple pour: *La moutarde semble vraiment monter au nez de Luc*.

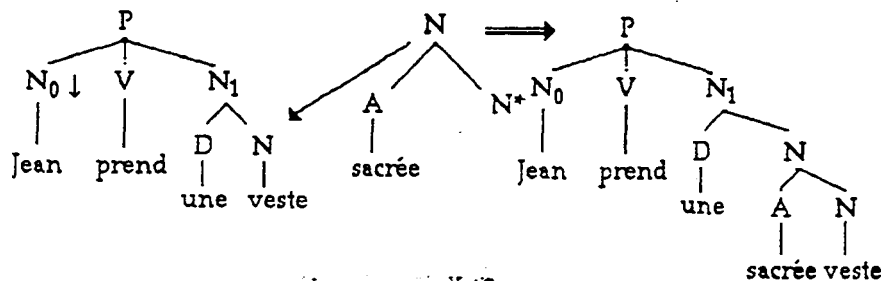


Figure 29

Quand toute adjonction est interdite à un noeud, cette contrainte doit être explicitement marquée, mais ceci est rare.

Ces discontinuités ne handicapent ni la représentation des entrées lexicales, qui indique simplement pour chaque item le noeud auquel il s'attache dans l'arbre élémentaire, ni le calcul sémantique, puisqu'une des différences entre l'arbre dérivé et l'arbre de dérivation est que le second ne tient pas compte de l'ordre des mots. Ainsi les arbres de dérivation de *Jean prend Marie en compte* ou *Jean prend une sacrée veste* sont respectivement:

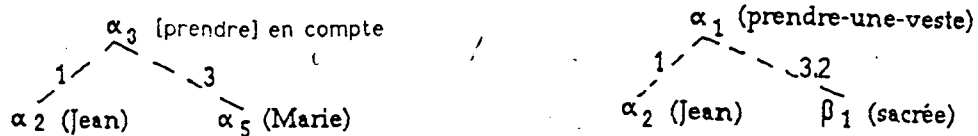


Figure 30

Les discontinuités peuvent venir également de l'application régulière des règles syntaxiques. Par exemple le passif crée une discontinuité par l'adjonction de l'auxiliaire, et d'éventuels autres modifieurs:

La glace ne risque pas d'être bientôt brisée.

Les questions et autres phénomènes d'extractions nous ramènent au cas plus classique des dépendances non bornées:

*Sur le dos de qui pensez-vous que Luc a cassé du sucre ?
C'est sur le dos de Marie que tout le monde s'accorde à dire que Jean a cassé du sucre.*

Les arbres élémentaires correspondants (au passif ou à la question ou à la phrase clivée) sont sélectionnés par l'idiome, qui a une famille d'arbres syntaxiquement régulière, sauf indication contraire, et les adjonctions se font de la même façon que dans le cas libre.

Notre représentation ne suppose donc pas un ordre canonique des expressions figées, ce qui est conforme aux phénomènes observés. La flexibilité des expressions figées devient le cas général, leur figement absolu l'exception.

Le parseur ne sélectionne un arbre idiomatique que si tous les éléments fixes de l'expression sont présents dans la phrase (leur ordre pouvant être variable), voir chapitre 7.

2.2.4. Familles figées

Les expressions figées sélectionnent en fait un ensemble d'arbres élémentaires regroupés en familles d'arbres selon les mêmes principes que pour les verbes simples. Les familles d'arbres encodent la structure argumentale complète de l'expression. Dans le nom de la famille, les arguments figés sont notés en majuscule, les arguments libres en minuscule. Certains arguments figés, comme certains arguments libres, peuvent être optionnels, et cela est noté dans le nom de l'expression. Ils sont numérotés comme dans le cas libre, de 0 à *i* en commençant par le sujet. Ainsi *N₀ prendre une veste* appartient à la famille *n₀VDN₁*, *N₀ prendre des vessies pour des lanternes* à *n₀VDN₁PDN₂*.

Nous avons représenté la plupart des familles figées à arguments figés ou libres, nominaux, prépositionnels ou phrastiques, ces différents arguments pouvant comporter des syntagmes figés enchassés. Nous n'avons pas implémenté les expressions à adverbes figés, ni celles comportant plusieurs phrases figées enchassées ou coordonnées, du type:

*N₀ partir sans laisser d'adresse.
Ne réveillez pas le chat qui dort
J'en passe et des meilleures.*

Leur représentation suivrait les mêmes principes que pour les autres expressions.

Les différentes règles syntaxiques s'appliquent aux arguments libres comme aux arguments figés, aux syntagmes correspondants à des "feuilles" des structures arborescentes comme à ceux correspondant à des noeuds intérieurs.

Toutes n'ont pas encore été systématiquement étudiées, et les tables de phrases figées comportent peu de colonnes de propriétés par rapport à celles des phrases à verbes simples ou à verbe support. De plus, les jugements d'acceptabilité sont souvent délicats

et souvent plus variables que dans le cas libre. Nous sommes donc loin d'avoir représenté toutes les propriétés des phrases figées.

Comme pour les phrases libres, et avec les mêmes réserves que pour celles-ci (voir supra), on pourrait générer par métarègles les arbres correspondant aux différents types de structures figées élémentaires. Il devrait bien entendu s'agir des mêmes métarègles, ce qui supposerait des métavariabes de taille non bornée²⁷.

Les structures élémentaires correspondant à l'application des différentes règles syntaxiques disponibles pour une famille donnée portent en elles-mêmes des contraintes associées à leurs noeuds. Leur sélection peut être également contrainte pour une expression donnée à l'aide des mêmes traits que pour les verbes simples. Ces traits sont simplement plus souvent utilisés que dans le cas libre, qu'il s'agisse d'interdire une règle (valeur -) ou de la rendre obligatoire (valeur +).

a) Passif

Les arbres passifs ont *être* et *par* pour têtes lexicales (en plus du participe passé), comme dans les familles libres et à verbe support. Le passif s'applique aux compléments libres comme aux compléments figés. Comme précédemment, on note *passif*= - dans le dictionnaire si l'expression ne sélectionne aucun des arbres passifs existant dans sa famille.

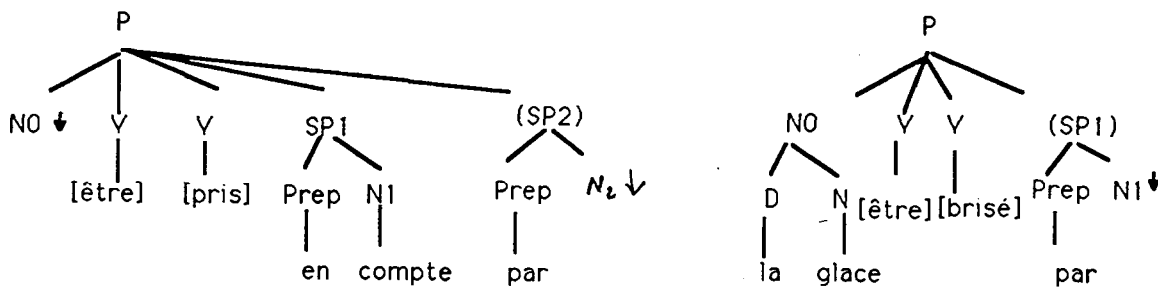


Figure 31

27 Si les structures figées élémentaires devaient être elles-mêmes dérivées de celles des phrases libres, ces métarègles seraient en fait des transformations puisqu'opérant sur n'importe quelle structure dérivée (il n'y a pas de limite a priori à la profondeur d'un arbre idiomatique) et non seulement sur des arbres élémentaires (finis) définis a priori. Le système serait alors formellement trop puissant pour permettre des traitements informatiques raisonnables.

On note quelques cas où l'expression n'existe qu'au passif:

Cette solution est tirée par les cheveux
Jean est pris par le temps

Deux formes actives sont a priori possibles pour ces phrases (puisque'on ne sait pas si le complément en *par* est agentif ou non), mais aucune n'est attestée (avec le sens voulu):

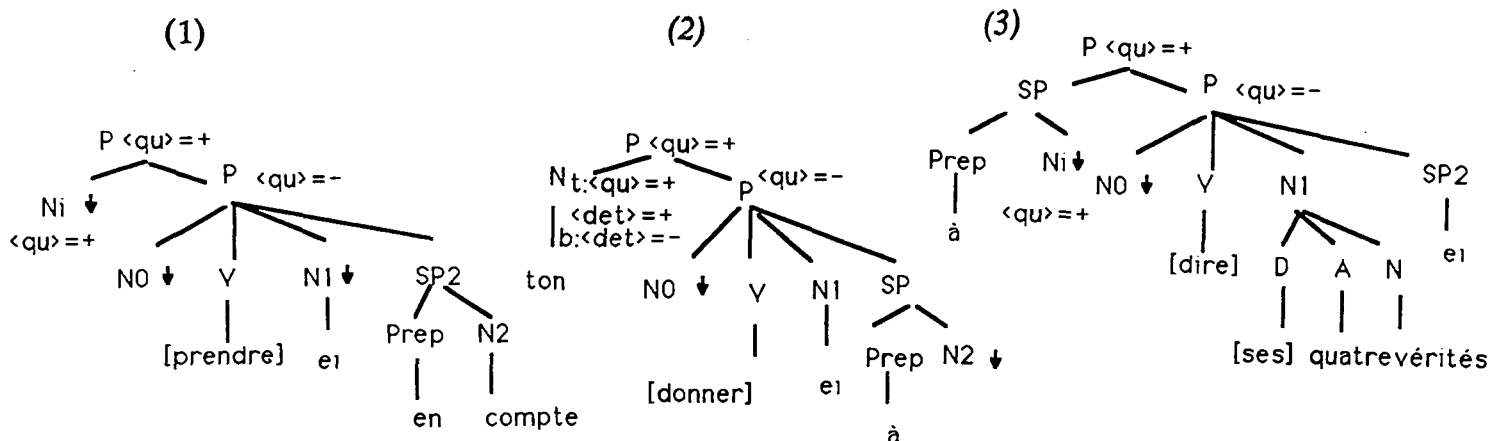
- * *Les cheveux tirent cette solution*
- * *N₀ tire cette solution par les cheveux*
- * *Le temps prend Jean*
- * *N₀ prend Jean par le temps*

Ces expressions ont les mêmes familles que celles existant à l'actif et au passif, mais elles sont marquées *passif*=+, ce qui empêche la sélection d'autres arbres que ceux au passif.

b) Questions

Elles peuvent s'appliquer aux compléments figés comme aux compléments libres. Lorsqu'il s'agit d'un complément figé, son déterminant doit pouvoir être variable car c'est la question en *quel* qui s'applique (Cf supra).

La question est productive lorsque le déterminant figé est *un*, il est alors marqué comme optionnel dans l'entrée lexicale, ce qui permet à l'arbre de la question d'être sélectionné, mais il sera bien figé dans la plupart des autres structures de sa famille (M. Gross 1989b explique la possibilité du questionnement par la commutation de l'argument figé avec *quelque chose*, ce qui explique le rôle de l'article indéfini).



- (1) *Quel argument Jean prend-il en compte ?*
 (2) *Quel ton Jean donne-t-il à leurs soirées ?*
 (3) *A quelle fille Jean a-t-il dit ses quatre vérités ?*

Figure 32

Comme dans le cas libre, les mêmes structures sont utilisées pour les questions directes et indirectes.

Comme le note M. Gross 1989a, certaines expressions sont réservées à la forme interrogative:

Quelle mouche pique Jean ?

Comme dans le cas libre, cette question est générée à partir de l'arbre déclaratif (où le SN sujet <qu> = + est figé). Le passif s'applique et deux variantes sont alors possibles:

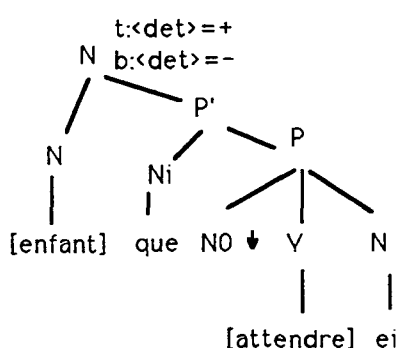
? Par quelle mouche N a-t-il été piqué ?

N est piqué par quelle mouche ?

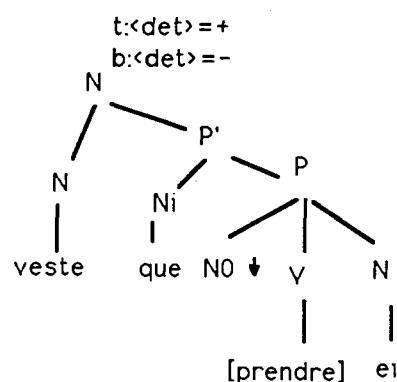
Ce qui montre bien que c'est un trait lexical sur le groupe *quelle mouche* et non une contrainte structurale sur la phrase qui est à l'oeuvre. Nous n'avons donc pas besoin d'un trait *question* = +.

c) Les relatives

Les relatives correspondent à des sujets ou à des compléments, libres ou figés. Dans le cas d'un argument figé, il figure comme antécédent de la relative déjà spécifié (comme pour les Npred des phrases à support). L'arbre de la relative est alors un arbre initial (correspondant à un SN qui sera lui-même argument d'une autre phrase) et non un arbre auxiliaire. Son déterminant doit donc pouvoir être variable.



L'enfant que Marie attend



La veste que Jean a prise

Figure 33

Comme pour la question, la relative est productive quand le déterminant figé est *un*, mais aussi chaque fois que le déterminant figé alterne avec un-modif:

Jean a pris une veste

La veste que Jean a prise ...

Jean donne (le ton + un ton nouveau) à nos soirées

Le ton (E + nouveau) que Jean donne à nos soirées ...

Le déterminant figé (et éventuellement *un* avec lequel il alterne) est alors marqué comme optionnel dans l'entrée lexicale (ce qui permet à l'arbre de la relative d'être sélectionné, mais il sera bien figé dans la plupart des autres structures de sa famille).

Nous n'avons donc pas besoin d'un trait *relative* particulier.

d) Extraposition

Les verbes à sujet *il* figé (*il faut que P*, *il s'agit de/que P*) sont rangés dans la famille des expressions à sujet figé, et ne sont donc pas analysés comme des extrapositions (même si cela serait justifié diachroniquement).

Comme dans le cas libre, l'extraposition concerne les verbes à sujet phratique ou à sujet nominal indéfini:

Que P pend au nez de Luc

Il pend au nez de Luc : que P

Beaucoup de projets sont déjà tombés à l'eau

Il est déjà tombé à l'eau beaucoup de projets

Des ailes poussent à Max

Il pousse des ailes à Max

Les sujets nominaux extraposés doivent être indéfinis, ils peuvent être libres ou figés.

La complétive peut être enchassée à l'intérieur d'un sujet figé, et dans ce cas N_0 n'a pas besoin d'être indéfini:

Le bruit que P court

Le bruit court que P

Il court le bruit que P

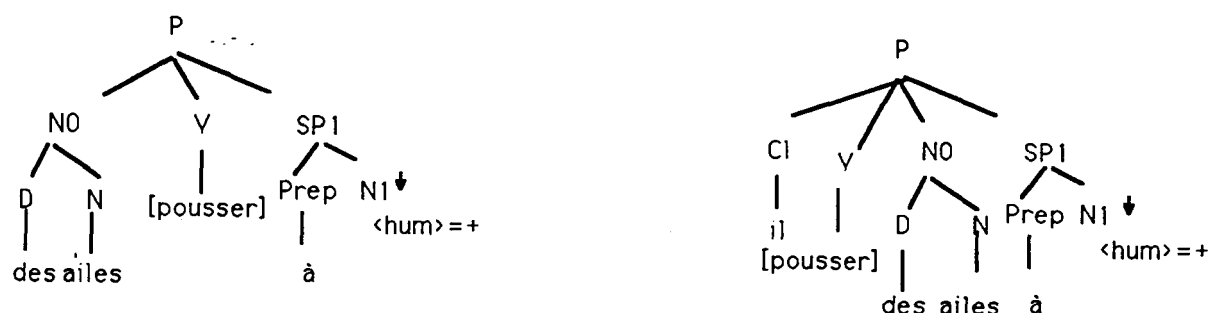


Figure 34

Nous ne représentons pas le fait que ce sont en général des verbes admettant aussi l'extraposition dans le cas libre (M. Gross 89b).

Les familles concernées sont donc s_0VPN_1 , s_0VPDN_1 , n_0VPN_1 , n_0VPDN_1 , DN_0V , DN_0Vp_{n1} , N_0V ...

Comme l'ont noté R. Kayne 1976 et M. Gross 1975, l'extraposition est parfois possible avec des verbes à objet figé et déterminant zéro:

si:
Il prend corps dans le pays une grande espérance (Kayne)
Il a (traversé + effleuré) l'esprit de Paul que Marie soit partie (Gross)

Il faut donc noter le trait *extrap* dans les entrées lexicales correspondant aux familles concernées: n_0VN_1 et s_0VN_1 . Par défaut *extrap* = -, puisque ce phénomène est plutôt rare:

?? *Il a pris froid beaucoup de monde cet hiver*

Comme dans le cas libre, l'extraposition est régulière au passif. Comme dans le cas libre également, l'extraposition + passif peut être autorisé pour des structures interdisant le passif personnel :

On a paré au plus pressé
 * *Au plus pressé a été paré*
Il a été paré au plus pressé

Cela est noté dans l'entrée lexicale de l'expression par l'implication: passif = + => extrap = +.

e) Clivées

L'extraction s'applique aux arguments libres comme aux arguments figés, selon les conditions étudiées au chapitre 2.

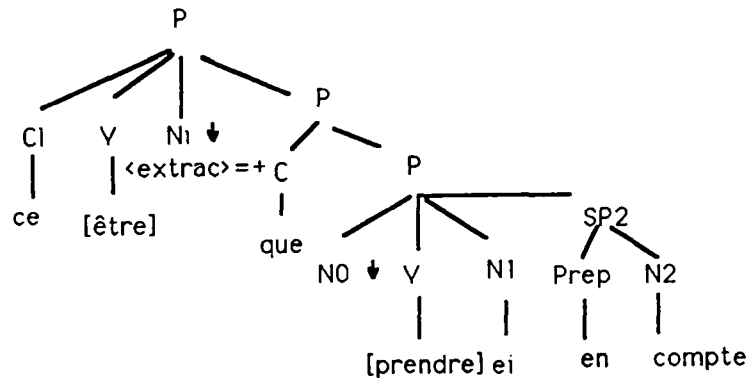


Figure 35

Lorsque l'argument figé a lui-même un complément libre, l'extraction est en général possible:

Jean marche sur les brisées de Luc.

C'est sur les brisées de Luc que Jean marche.

C'est une caractéristique de cette famille d'arbres.

Dans les autres cas, il faut prendre en compte le caractère figé ou non du déterminant de l'argument (Cf chapitre 2). S'il est libre, l'extraction est possible (avec les mêmes restrictions que dans le cas libre). S'il est figé, l'argument figé doit en général être modifié (ce qui entraîne souvent un changement de déterminant).

Jean a pris une (E + sacrée) veste

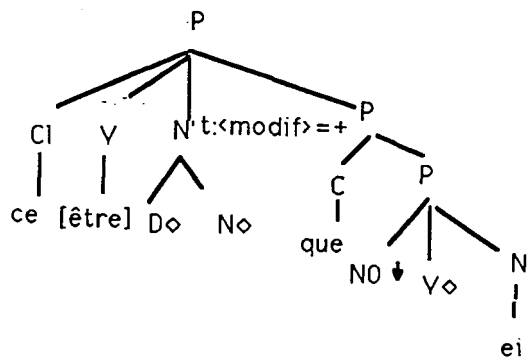
*C'est une (*E + sacrée) veste que Jean a prise*

Marc donne (le ton + un ton nouveau) à nos soirées

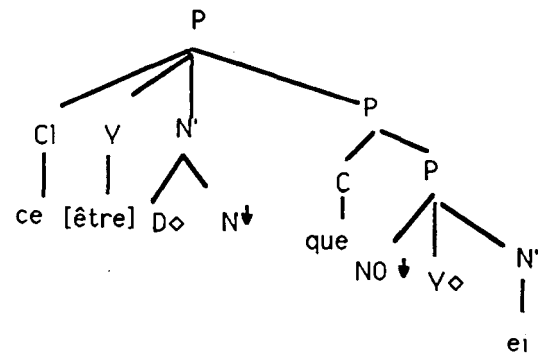
*C'est (*le ton + un ton nouveau) que Marc donne à nos soirées*

Nous notons donc, dans ce cas, une contrainte $\text{top} \langle \text{modif} \rangle = +$ sur le foyer de clivée, ce qui crée une contrainte d'adjonction obligatoire²⁸ :

²⁸ Les têtes lexicales sont notées ♠.



famille n_0VDN_1 ($=: N_0$ prendre une veste)
chasser LE Ngibier)



famille n_0VDn_1 ($=: N_0$

Figure 36

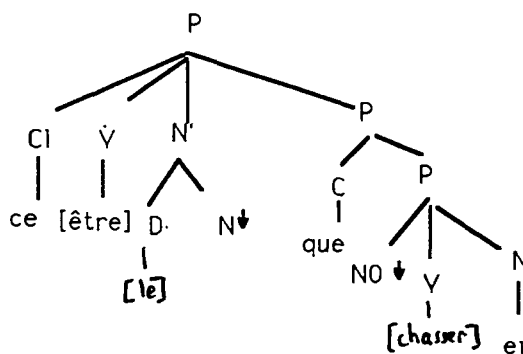
La structure ne pourra donc être sélectionnée que si l'expression autorise la modification du nom figé, et avec les conséquences lexicalement spécifiées sur la variation du déterminant. Dans le dictionnaire, N_0 mettre le paquet ou N_0 casser sa pipe sont marqués $N_1.t: <modif> = -$, on n'aura donc pas:

- # C'est sa pipe qu'il a cassée
- # C'est le paquet qu'il a mis

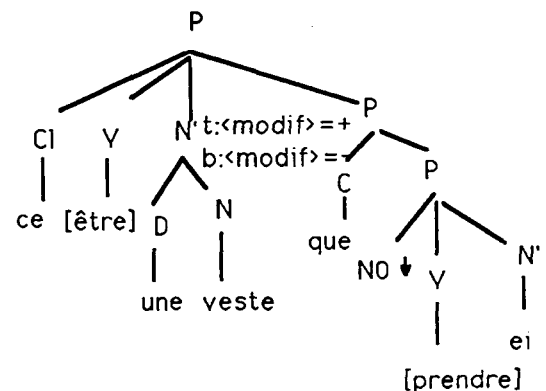
Dans le cas des arguments à déterminant figé (mais nom libre), la clivée est possible (Cf chapitre 2):

- Il chasse la (palombe + grive + poule d'eau ...)
- C'est la (palombe + grive + poule d'eau ...) qu'il chasse

Il est donc justifié de distinguer les familles selon le caractère figé ou non de leurs noms et de leurs déterminants, puisque leurs arbres élémentaires (et leurs contraintes) ne sont pas les mêmes.



C'est la palombe qu'il chasse



C'est une (*E + sacrée) veste que Jean a prise

Figure 37

Lorsqu'un argument figé comprend un syntagme enchassé, des phénomènes de double extraction sont possibles, comme dans le cas des phrases à verbe support, puisque le principe de l'arbre étendu est le même. De fait, de nombreuses expressions, que nous ne considérons pas comme phrases à support car elles ne permettent pas la formation d'un SN complexe par incorporation du sujet, permettent la double extraction. Ainsi pour N_0 faire le jeu de N_1 (* le jeu de N_1 de/par N_0):

Marc a fait le jeu de ses ennemis
C'est le jeu de ses ennemis qu'il a fait
C'est de ses ennemis qu'il a fait le jeu

Quand, pour de telles structures, la double extraction est interdite, il faut le noter par un trait *clivée_i* = - (i pour l'indice du syntagme enchassé). Un marquage par défaut est bien entendu possible.

Comme dans le cas des phrases à verbe support, nous considérons que l'extraction du complément de nom n'est pas possible quand le syntagme est enchassé à l'intérieur d'un complément prépositionnel:

Jean marche sur les brisées de Luc.
C'est sur les brisées de Luc que Jean marche.
 * *C'est de Jean qu'il marche sur les brisées.*

Il y a également des expressions dont la forme de base est la phrase clivée (N. Ruwet 1984):

C'est au pied du mur qu'on voit le maçon.
C'est ici/là que les Athéniens s'atteignirent.

Pour celles-ci, on ne crée pas une famille particulière, on les marque avec le trait *clivée_i* = + (i étant l'indice du syntagme obligatoirement clivé), ce qui contraint la sélection des arbres désirés.

f) Neutralité

M. Gross 1989a mentionne des cas de neutralité tels que:

Cette loi a sonné le glas de nos projets
Le glas de nos projets a sonné

Nous les représentons par des arbres associés aux familles concernées, de la même façon que pour le cas libre (mais avec des arbres plus étendus), et un trait *neutre* qui marque si cette possibilité est réalisée dans l'entrée lexicale de l'expression. Nous n'avons pas de moyen de noter le fait que les verbes des expressions concernées sont également neutres dans leur construction libre.

Les phrases figées symétriques, peu étudiées, seraient représentées de la même façon que dans le cas libre (toujours avec des arbres plus étendus).

2.2.5 Analyse sémantique

a) Ambiguïté des phrases figées

Même si le phénomène est relativement rare dans les textes, nous prévoyons une représentation ambiguë entre sens littéral et sens idiomatique de la plupart des phrases figées. Ces ambiguïtés permettent de nombreux jeux de mots:

La pollution nous pompe l'air (publicité Fina, été 1989)

Les mots de l'idiome ayant en général également des entrées 'libres' dans le dictionnaire, leur combinaison aboutira au même arbre dérivé que pour l'idiome, mais l'arbre de dérivation (sur lequel se base la représentation sémantique) sera différent. Si un idiome comporte des mots qui n'ont jamais d'occurrence dans des phrases libres, ou dont la sous catégorisation est différente dans les phrases libres (comme *aller* dans : *N₀ aller son train*), il ne sera bien sûr pas analysé comme ambigu.

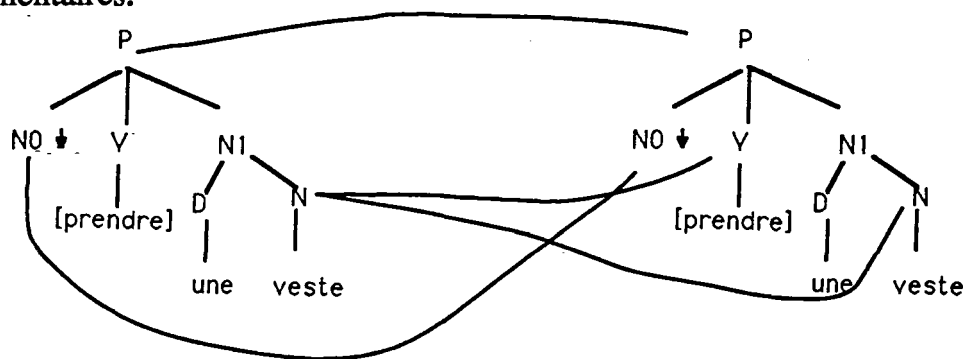
b) Alternance adverbes/adjectifs/déterminants

L'on retrouve un phénomène déjà vu pour les phrases à verbe support:

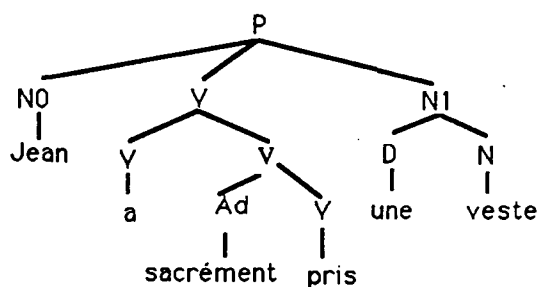
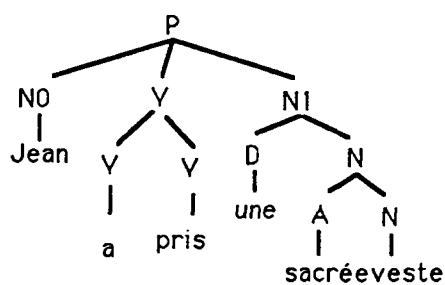
Jean a pris une (vraie + sacrée) veste
Jean a (vraiment + sacrément) pris une veste

La représentation est la même: le noeud N correspondant à *veste* est lié au noeud V correspondant à *prendre*, et qu'il y ait adjonction d'un modifieur en V ou d'un modifieur (lié au premier) en N, les phrases résultantes sont dites équivalentes, suivant le principe des dérivations synchrones:

arbres élémentaires:



arbres dérivés:



arbres de dérivation:

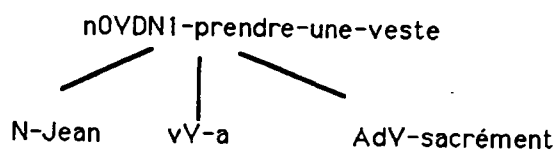
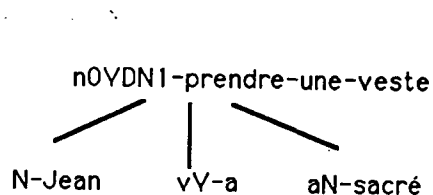


Figure 38

Les déterminants également peuvent jouer le même rôle que des adverbes, mais nous n'avons pas encore implémenté cette alternance:

Jean a brûlé trois feux (rouges)
 = *Trois fois, Jean a brûlé un feu (rouge).*

c) Alternance adjectifs/ de N

D'autres exemples mettent en jeu l'alternance de l'adjectif avec un complément de nom en *de N*.

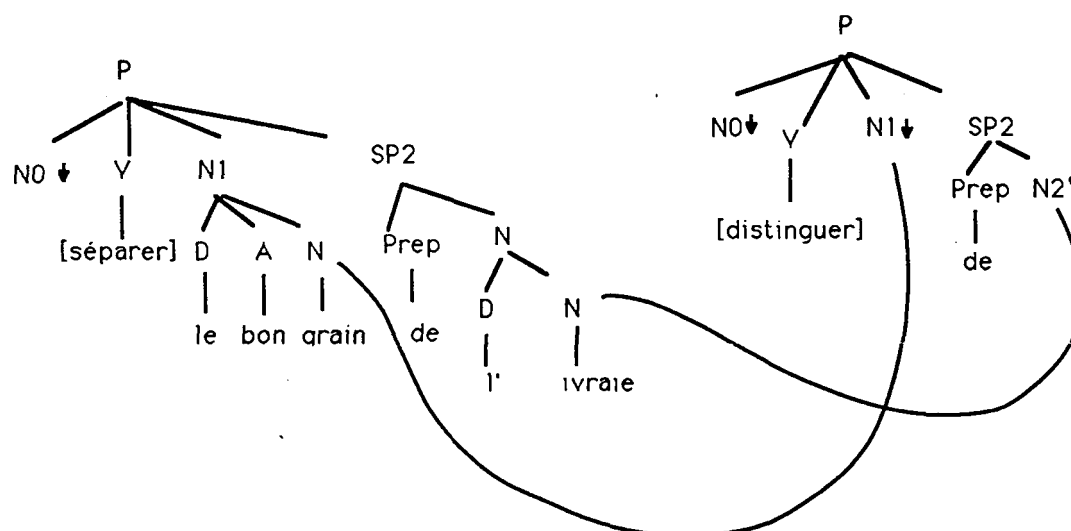
- (1) *Ils apportent de l'eau au moulin indépendantiste* (Le Monde, juin 1989)
 (2) *Cet examen prétend séparer le bon grain individualiste de l'ivraie collectiviste.* (Le Monde, Octobre 1989)

La phrase (1) est équivalente à *Ils apportent de l'eau au moulin de l'indépendantisme* et la phrase (2) à *Cet examen prétend séparer le bon grain de l'individualisme de l'ivraie du collectivisme*. L'adjonction d'un adjectif n'est donc pas équivalente ici à un adverbe mais à un complément de nom en *de N*. Syntactiquement, les adjectifs *individualiste* et *collectiviste* s'attachent aux noms *grain* et *ivraie* respectivement. Mais sémantiquement, ils sont arguments de toute l'expression qui se comporte comme un verbe composé à deux compléments libres.

Cette alternance dépend de l'adjectif, selon qu'il est lié à un adverbe ou à un nom (certains adjectifs sont à cet égard ambigus). Elle dépend aussi de l'expression idiomatique, selon que le noeud N où s'adjoint l'adjectif est lié à un noeud V (ou P) où peut s'adjoindre un adverbe ou non.

Lorsque l'adjectif est équivalent à un complément de nom, il s'agit en général d'un complément libre (et optionnel) de l'expression idiomatique, comme l'indique sa paraphrase libre:

- N₀ séparer le bon grain (X) de l'ivraie (Y)*
 = *N₀ distinguer X et Y*
- N₀ prendre des vessies (X) pour des lanternes (Y)*
 = *N₀ confondre X et Y*



- N₀ séparer le bon grain individualiste de l'ivraie collectiviste*
 = *N₀ distinguer individualisme et collectivisme*

Figure 39

N₀ séparer le bon grain individualiste de l'ivraie collectiviste
 = *N₀ distinguer individualisme et collectivisme*

Les adjectifs (*individualiste*) et les noms (*individualisme*) sont reliés entre eux par ailleurs.

L'on peut donc représenter par le biais de telles paraphrases, même pour les expressions qui ne comportent pas de compléments libres, une structure argumentale sémantique à arguments libres différente de la structure syntaxique:

idiomes	nb de compléments	nb d'arguments sémantiques ²⁹
prendre une veste	1	0
nettoyer les écuries d'Augias(X)	1	1
mettre de l'eau dans son vin	2	0
prendre le train(X) en marche	2	1
jeter un pavé dans la mare(X)	2	1
prendre des vessies(X)		
pour des lanternes(Y)	2	2
séparer le bon grain(X) de l'ivraie(Y)	2	2

La représentation d'un argument libre pour *N₀ jeter un pavé dans la mare* et *N₀ prendre le train en marche* se fonde sur les exemples suivants:

Ce vigoureux pamphlet est un pavé jeté dans la mare de cette rentrée sereine (Le Monde, septembre 1989)

La Grande-Bretagne allait une nouvelle fois prendre le train européen en marche (Le Monde, juin 1988)

Une option alternative serait de prévoir au niveau sémantique une interprétation appositive de ces compléments *de N* ; cependant, le risque certain de surgénération nous fait préférer pour l'instant le marquage systématique des éléments fixes de l'idiome dont les ajouts en *de N* peuvent être interprétés comme des arguments.

²⁹ sans compter les sujets.

Chapitre V

Phénomènes d'extraction et dépendances non bornées

Ces phénomènes ont été les premiers à être étudiés dans une optique TAG pour l'anglais (A. Kroch et Joshi 1985, A. Kroch 1987). Ces études se situaient dans la perspective chomskyenne du Gouvernement et Liage (*Government and Binding* ou GB; N. Chomsky 1981). Dans cette optique, le formalisme permet une formulation plus simple de certaines contraintes d'extraction. Ce qui, en grammaire générative, doit s'exprimer comme contrainte sur des mouvements, et donc sur des structures non bornées, peut être exprimé en TAG comme condition de bonne formation sur les arbres élémentaires, c'est-à-dire comme contrainte locale.

Nous avons élargi le domaine d'investigation à des contraintes non étudiées par ces auteurs (contrainte du sujet phrastique) ou dont la formulation ne rendait pas compte des faits observés en français (contrainte du SN complexe).

Nous allons présenter d'abord les contraintes portant sur l'extraction (question, relative, clivée) des compléments et des ajouts, c'est-à-dire le statut des différents îlots isolés par la grammaire générative. Puis nous montrerons comment le formalisme TAG permet une représentation aisée des dépendances non bornées (c'est-à-dire l'extraction hors d'un domaine phrastique enchâssé). Au lieu de représenter ces phénomènes en termes de mouvement (comme en grammaire générative) ou par la propagation de traits sur des distances non bornées (comme en GPSG), nous pouvons en donner un traitement purement "local".

1. Contraintes d'extraction

1.1 Discussion théorique

1.1.1. Asymétries sujet-compléments

L'on note une alternance entre *qui* et *que* pour les clivées selon que l'élément extrait est sujet ou non. Une alternance entre *qui* et *que* est aussi observable pour les relatives, mais *que* ne concerne alors que les compléments nominaux directs :

C'est Max qui a vu Luc
C'est Luc que Max a vu
C'est hier que Max a vu Luc

Il n'y a pas de difficultés à rendre compte de ces alternances dans notre grammaire puisque des extractions de syntagmes de fonctions différentes correspondent à des structures élémentaires différentes et que l'élément *qu-* fait partie de la tête lexicale de

ces arbres.

Une autre alternance est marginalement observée lorsque la phrase à extraction est une complétive. Lorsqu'il y a extraction du sujet de la complétive, le complémenteur devient *qui*, et non plus *que*:

Quel homme crois-tu que Marie aime ?
*? Quel homme crois-tu qui aime Marie ?*¹

Lorsque la phrase enchâssée est une relative ou une phrase clivée, l'ordre entre pronom *qu-* et complémenteur s'inverse: il devient *que... qui* et non plus élément *qu-* ..*que*:

C'est l'homme que tu crois que Marie aime.
C'est l'homme à qui tu crois que Marie pense.
C'est l'homme que tu crois qui aime Marie.

Ces alternances sont plus difficiles à représenter dans notre cadre, puisque le complémenteur appartient à la structure élémentaire du verbe principal et non à la structure enchâssée (Cf chapitre 3). Il est à noter en outre que l'enchâssement d'une phrase à extraction du sujet donne souvent un résultat beaucoup moins acceptable qu'avec l'extraction d'un élément d'une autre fonction.

1.1.2. Contraintes d'îlots

Différents types d'îlots sont distingués en grammaire générative:

- les contraintes de J. Ross 1967: SN complexe, sujet phrastique, syntagmes coordonnés.
- le principe A/A (N. Chomsky 1973)
- la sous-jacence (N. Chomsky 1977)

a) Contrainte du SN complexe

Elle exprime l'impossibilité d'extraire un élément dominé par un noeud P lui-même dominé par un noeud SN.

Cette contrainte a été abondamment discutée. Elle a été notamment critiquée par D. Godard 1980a pour le français, qui montre que l'extraction hors d'une complétive complément de nom n'a rien à voir avec l'extraction hors d'une relative. J. Ross lui-même notait pour l'anglais des exceptions pour les complétives des "modal constructions" de Z. Harris (en fait des nominalisations à verbe support) ou des "picture nouns".

Comme on l'a vu au chapitre 4, l'extraction hors des compléments de noms est assez régulière avec les phrases à verbe support et certaines phrases figées. Elle est également possible avec certaines combinaisons de verbes causatifs et noms prédicatifs telles:

¹ L'acceptabilité de cette phrase est variable selon les locuteurs (Cf supra p. 157 et infra p. 260). Elle est pour nous douteuse.

*Tu lui as donné l'espoir d'aller au Japon
Où lui as-tu donné l'espoir d'aller ?*

b) Contrainte de l'îlot nominal

La contrainte du SN complexe a été étendue à tous les composants du syntagme nominal, en particulier aux compléments de nom.

En fait, comme dans le cas des complétives, les SP compléments de nom peuvent être extraits facilement dans le cas des phrases à verbe support et de certaines combinaisons verbe-objet comme par exemple:

*Tu as vu le portrait de Marie
De qui as-tu vu le portrait ?*

Le complément peut parfois être plus enchâssé :

*Tu as vu une copie du portrait de Marie
De qui as-tu vu une copie du portrait ?*

Il semble qu'on puisse en fabriquer en augmentant la cohésion sémantique verbe-complément (qui devient un "prédicat naturel" Cf N. Erteschik 1977)².

Nous ne traitons de telles extractions dans notre grammaire que dans le cadre de phrases à verbe support et phrases figées et nous les interdisons généralement dans le cas des phrases libres.

c) Îlots sujets

L'idée qu'il existerait des contraintes propres à la position sujet est ancienne: contrainte du sujet phrastique (J. Ross 1967), contrainte du sujet spécifié (N. Chomsky 1977), phénomènes d'opacité.... En fait, la particularité du groupe sujet, dans sa définition géométrique, ne s'observe en français que pour l'extraction hors d'une phrase, non pour l'extraction hors d'un SN.

* sujets nominaux

Considérons le cas des sujets nominaux. Certains compléments de noms peuvent être extraits mais ce n'est pas le fait que le nom tête soit complément, plutôt que sujet, qui règle ce mouvement. Les phrases à verbe support autorisent de façon assez systématique cette extraction d'un complément nominal, et ce, que le nom prédicatif soit sujet ou objet :

² D. Godard à paraître recense les différents types de compléments susceptibles d'être extraits hors du groupe nominal : il s'agit essentiellement des "génitifs" (arguments de noms prédicatifs ou possesseurs) et des partitifs : *une classe dont je connais les élèves, le gâteau dont j'ai mangé une partie*. Le dernier cas pourrait être traité dans notre grammaire en autorisant l'adjonction du déterminant complexe (*une partie (de) N*) sur le noeud correspondant à la trace de l'objet (et à condition de contraindre le pronom *dont* dans ce cas par un trait spécifique).

Luc a fait une enquête sur Max
C'est sur Max que Luc a fait une enquête
Une enquête sur Max a été faite
C'est sur Max qu'une enquête a été faite

Même si l'on tente de reformuler la contrainte en terme de fonction sujet (au sens de la structure argumentative) et non de position sujet (au sens géométrique), excluant ainsi les sujets passifs, la contrainte ne tient pas. Il existe en effet d'autres exemples d'extraction tout aussi naturels où il est difficile de ne pas faire du groupe sujet un sujet de base, ainsi dans des phrases à verbe support où le nom prédicatif est sujet à l'actif:

Jean a le souvenir de ces faits
La mémoire de ces faits est revenue à Luc
Une envie soudaine de s'en aller a pris Luc

Comme pour les autres phrases à verbe support, le complément du nom prédicatif peut en général être extrait:

C'est de ces faits que Jean a le souvenir
C'est de ces faits que la mémoire lui est soudain revenue
C'est de s'en aller que l'envie l'avait pris (?E + pas de trouver un compromis)

Il n'y a donc pas lieu de faire de différences entre phrases à verbe support à nom prédicatif sujet ou objet pour l'extraction du complément de nom.

En revanche, si l'on considère l'extraction hors d'une complétive, la position sujet joue un rôle: l'extrapolation hors du SN complexe sujet est exclue; elle est un peu meilleure quand la complétive est extraposée:

Max a donné à Paul l'ordre de recevoir Marie
C'est Marie que Max a donné à Paul l'ordre de recevoir
L'ordre de recevoir Marie a été donné à Paul (par Max)
** C'est Marie que l'ordre de recevoir a été donné à Paul (par Max)*
? C'est Marie que l'ordre a été donné à Paul de recevoir

* sujets phrastiques

Si l'on considère l'extraction hors des arguments phrastiques, l'asymétrie entre sujets et objets est bien attestée :

On affirme qu'il aime Marie.
Qui affirme-t-on que Jean aime ?
Que Jean aime Marie a été affirmé.
** Qui que Jean aime a-t-il été affirmé ?*

Mais, comme dans le cas des phrases à verbe support, l'extraction a une acceptabilité

supérieure, si la complétive sujet est extraposée:

Il a été affirmé que Jean aimait Marie.
?? Qui a-t-il été affirmé que Jean aimait ?
Il plaît à Jean d'aller à la plage
? Où plaît-il à Jean d'aller ?

De même pour les complétives sujets de phrase copulatives:

Aller à Bilbao est intéressant.
? Où aller est-il intéressant ?*
Il est intéressant d'aller à Bilbao.
Où est-il intéressant d'aller ?

Les contraintes d'extraction obéissent donc à des règles complexes, devant prendre en compte la nature et la fonction du groupe hors duquel on extrait, et sont soumises à d'importantes variations sémantiques et lexicales (verbes supports, "verbes ponts", Cf infra).

1.2. Représentation en TAG

A. Kroch 1987 et 1990 étudie les îlots *wh*- et nominaux dans le cadre d'une mini grammaire TAG de l'anglais³ et conclut que la sous-jacence au lieu d'être un principe stipulé par la théorie linguistique, comme en grammaire générative, peut être considérée comme découlant naturellement du formalisme TAG.

C'est un résultat important, qu'il faut cependant nuancer si l'on considère les phénomènes d'îlots dans toute leur variété. Certains, mais pas tous, découlent en effet de la seule topologie des arbres élémentaires, c'est-à-dire des seules contraintes sur la bonne formation des structures TAG. Pour d'autres, l'utilisation de traits (c'est-à-dire de stipulations extérieures au formalisme TAG proprement dit) est également nécessaire.

Notre définition de structures "étendues" permet également de rendre compte naturellement (et sans qu'il y ait contradiction) de nombreux cas qui sont apparemment des violations de la contrainte de la sous-jacence.

Nous pouvons donc reformuler le traditionnel principe A/A en termes de structures élémentaires (non de structures dérivées) et en en précisant certaines limites.

Nous allons successivement examiner l'îlot de la question indirecte, l'îlot nominal (et les relatives), l'îlot adjectival et l'îlot prépositionnel. Les extractions par question, relativation ou clivée (qui correspondent toutes à des arbres élémentaires à tête verbale ou adverbiale) sont en général traitées suivant les mêmes principes. Pour ne pas alourdir la lecture, nous prenons en général un seul exemple, celui de la question. Quand les traitements (ou les contraintes) sont différents, nous les exposons tous.

³ Cette mini-grammaire n'était pas lexicalisée, et surtout ne comportait que l'adjonction, pas la substitution. C'est pourquoi notre traitement, s'il aboutit pour partie aux mêmes conclusions, est sensiblement différent.

1.2.1. Ilôts des questions indirectes

Les verbes à question indirecte prennent des compléments phrastiques comportant déjà l'élément question (si la question porte sur un complément du verbe) où un élément question (ou *si*) a été adjoit:



Figure 1

On ne peut générer une phrase telle que:

* *Qui Paul demande-t-il qui aime ?*

car elle supposerait un arbre initial *qui qui aime*, c'est-à-dire un arbre à deux éléments *qu* antéposés pour *aimer*, avec lequel combiner *Paul demande-t-il P*. Or l'arbre de la question est un arbre élémentaire défini sur un seul complément du verbe à chaque fois (Cf chapitre 3).

A partir de la phrase : *Tu demandes qui construit sa maison (ici + depuis si longtemps)*, on ne peut non plus générer:

* (*Où + Depuis combien de temps*) *demandes-tu qui construit sa maison ?*

Car cela supposerait d'adjoindre *où* ou *combien de temps* sur *qui construit sa maison*, ce qui est exclu non pas pour la même raison que supra (pas de double question avec antéposition en français) mais par une contrainte de trait, puisqu'il s'agit ici d'une contrainte sur la combinaison de deux arbres élémentaires :

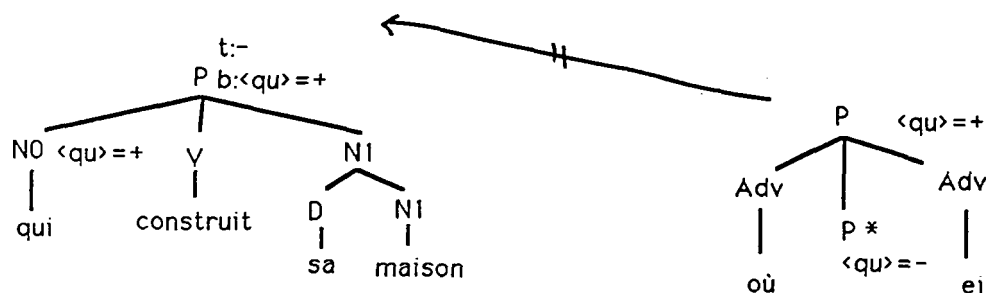


Figure 2

C'est la même contrainte locale qui interdit⁴:

- * *Où qui construit sa maison ?*
- * *Qui où construit sa maison*

Ce n'est donc pas non plus une contrainte en termes de mouvement.

Ainsi, la traditionnelle contrainte d'îlot *qu-* peut être exprimée directement en termes de bonne formation des structures élémentaires (interdisant un arbre à deux arguments *qu-* préposés en français; Cf A. Kroch et A. Joshi 1985), ou bien comme une contrainte locale sur la combinaison de deux structures élémentaires. On peut donc faire l'économie de contrainte de mouvement à exprimer sur des structures non bornées (comme en grammaire générative classique) ou de propagation de traits (comme en GPSG ou en LFG, Cf infra p. 249). Les relatives sont étudiées en même temps que les autres modificateurs du nom (îlot nominal)

1.2.2. Ilots nominaux

a) Les modificateurs du nom

Les modificateurs du nom sont des arbres auxiliaires en N. Ils peuvent comporter un élément *qu-* mais ils ne peuvent s'adjoindre en tête de phrase:

- (1) *Jean loue un appartement de trois pièces*
- (2) *Jean loue un appartement de combien de pièces ?⁵*
- (3) * *De combien de pièces Jean loue-t-il un appartement ?* (ou autre sens)

C'est donc une contrainte structurale qui empêche la phrase (3). De même pour leur relativation ou leur extraction par *c'est que*:

- * *les trois pièces dont Jean loue un appartement*
- * *Ce sont les trois pièces dont Jean loue un appartement*

L'on retrouve la même contrainte d'îlot avec les relatives, hors desquelles on ne peut extraire aucun élément:

- (4) *Je connais l'homme qui aime Marie*
- (4a) * *Qui connais-tu l'homme qui aime ?*

4 Comme le note D. Godard 1980b, l'extraction en relative peut être meilleure, surtout si l'interrogative est à l'infinitif : *un livre que je ne sais pas (à qui donner + où acheter)*. Il faudrait d'une part ajouter aux familles d'arbres (ici à celle de *donner*) certaines structures relatives à double extraction (avec adjonction obligatoire d'une principale), d'autre part autoriser l'adjonction de *où* sur certaines structures relatives.

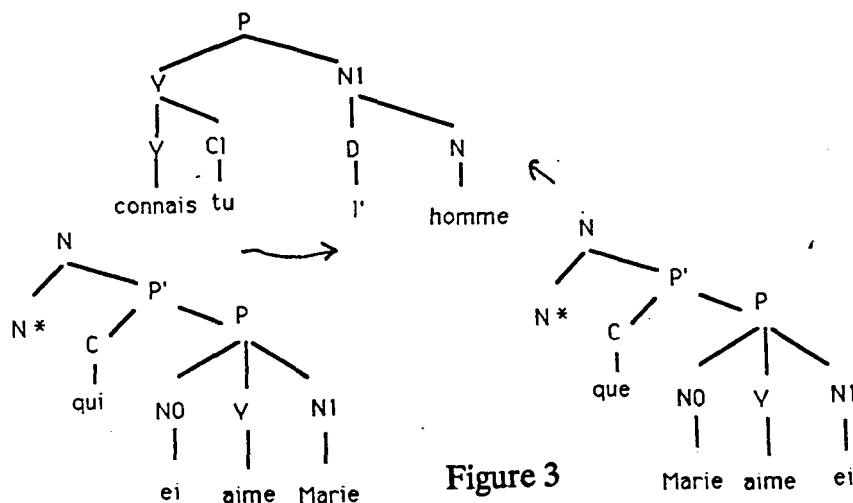
5 Nous appelons dans la suite du texte, et conformément à la tradition générative, ce type de questions "questions-écho". En fait, elles sont souvent employées comme questions à part entière, surtout à l'oral.

- (5) *Je connais l'homme que Marie aime*
 (5a) * *Qui connais-tu l'homme que aime ?*

Rappelons que les relatives sont des arbres auxiliaires en N à tête verbale. Elles correspondent aux sujets et compléments du verbe (ou du nom prédicatif pour les phrases à support) (Cf chapitres 3 et 4).

Les phrases 4a et 5a sont exclues pour raison formelle et pour raison linguistique :

- elles devraient être dérivées respectivement de: *Qui qui aime* et *Qui que aime*, qui ne font pas partie des structures bien formées correspondant à *aimer*, puisqu'un arbre élémentaire correspond à l'application d'une règle syntaxique à un syntagme à la fois. (contrainte linguistique, pour le français)
- la séquence *connais-tu l'homme* ne peut pas se combiner avec de telles structures puisqu'elle devrait s'adjoindre entre *qui* et *que* (ou *qui*) alors que c'est un arbre initial.
- il faudrait donc que *Qui qui aime* et *Qui que aime* soient des arbres auxiliaires s'adjoignant de part et d'autre de la racine phrastique de *connais-tu l'homme*, mais alors ce ne seraient plus des relatives.



La contrainte d'îlot est la même pour les relatives correspondants à certains ajouts à la phrase, par exemple *où* ou *dans lequel* :

- (6) *Je connais le magasin où Jean a acheté une montre*
 (7) * *Que connais-tu le magasin où Jean a acheté ?*

Les relatives correspondant aux adverbes sont aussi des arbres auxiliaires en N (parfois en Adv) et la principale est un argument substitué (Cf chapitre 3). La phrase 7 n'est pas possible dans notre système car :

- d'une part, une relative s'adjoignant à *magasin*, a par définition tous ses éléments après *magasin*. La seule solution serait que *connais-tu le magasin* s'adjoigne au milieu de la relative, mais ce ne peut pas être un arbre auxiliaire puisqu'il ne s'agit pas d'une structure à argument phrastique.
- d'autre part la structure *que où Jean a acheté* est mal formée: elle ne pourrait être dérivée que par l'adjonction de *où* dans la structure à question *que Jean a acheté* (mais il n'y a pas de noeud N où adjoindre sauf le sujet), et ce ne peut être un arbre auxiliaire

phrastique s'adjoignant de part et d'autre de *connais-tu le magasin*).

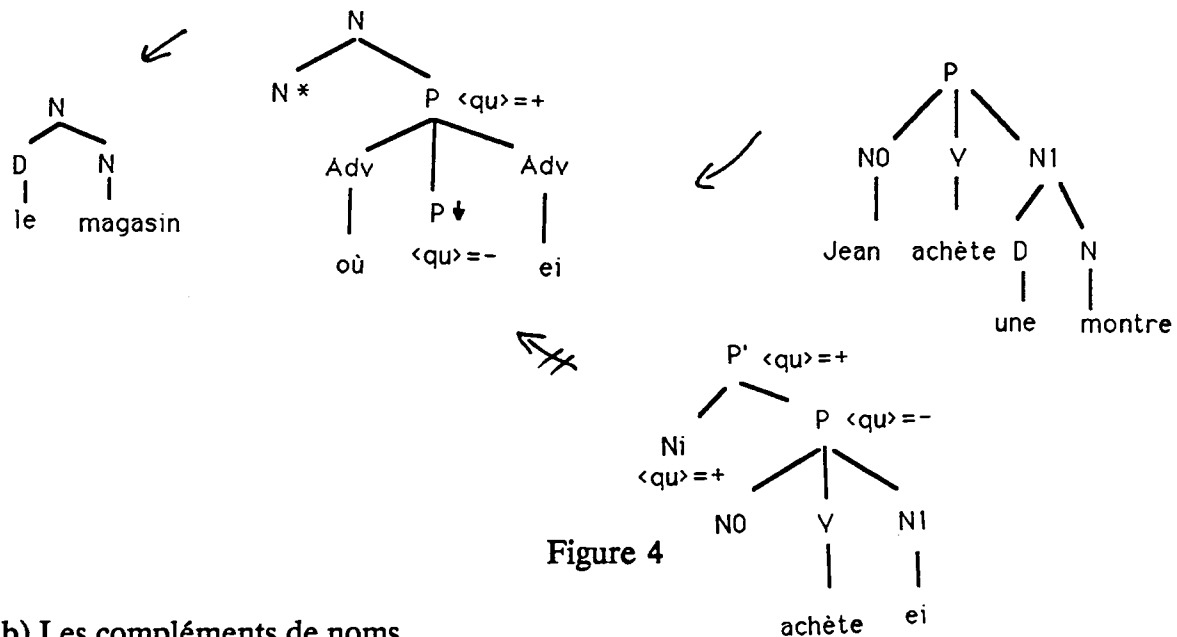


Figure 4

b) Les compléments de noms

Pour les compléments de nom, nous traitons l'extraction différemment selon qu'il s'agit d'une phrase à verbe simple ou d'une phrase à verbe support. Dans le premier cas, nous interdisons l'extraction (ce qui est bien sûr une grossière approximation) et nous l'autorisons dans le second :

- (1) *Jean a rapporté une agression contre Marie*
- (1a) * *Contre qui Jean a-t-il rapporté une agression ?* (ou avec autre sens)
- (2) *Jean a commis une agression contre Marie*
- (2a) *Contre qui Jean a-t-il commis une agression ?* (M. Gross 1976)

Les seules questions correspondant au SN complexe: *Det agression (de N₀) contre N₁* sont celles où l'on substitue un pronom interrogatif (où un SN avec un élément *qu-*) au noeud N₀ ou N₁ :

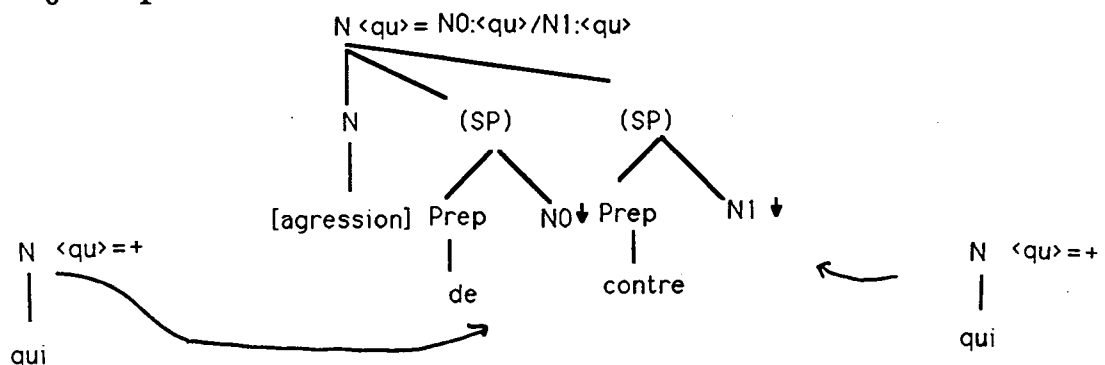


Figure 5

une agression (contre qui + quelle femme)
l'agression (de qui + de quel voisin) contre Marie

Ces SN doivent, comme tous les compléments (ou sujets) SN, être substitués dans l'arbre du verbe principal. Les seules phrases possibles dans notre grammaire sont donc les questions en écho:

Jean a rapporté l'agression (de Jean) contre qui ?

Jean a rapporté une agression de qui (contre Marie) ?

La phrase (1a) n'est donc pas possible, pas plus que la phrase (1b):

(1b) *De qui Jean a-t-il rapporté une agression (contre Marie) ?*

La seule question possible qui fasse passer un complément de *agression* en tête de phrase est celle liée à la phrase à support: *agression* est alors tête d'un arbre élémentaire phrastique (dont N_1 est une des feuilles).

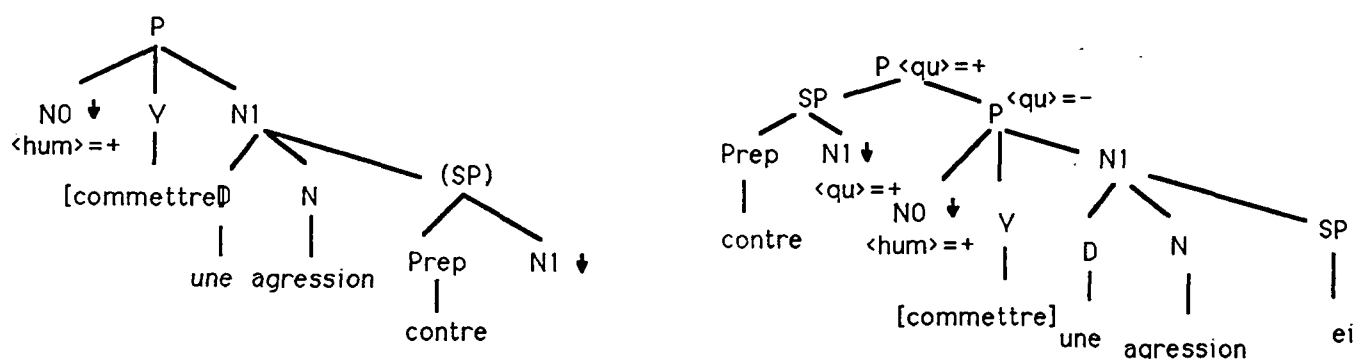


Figure 6

Il en est de même pour la relative et la clivée (Cf chapitre 4)

On retrouve le même contraste entre phrase à verbe simple et phrase à verbe support lorsque le complément de nom est une complétive:

(3) *Jean désapprouve le projet (E+ de Marie) d'aller à Bilbao*

(3a) ** Où Jean désapprouve-t-il le projet d'aller ?*

(4) *Jean a le projet d'aller à Bilbao*

(4a) *Où Jean a-t-il le projet d'aller ?*

Il y a bien un arbre élémentaire correspondant à la question *Où aller* mais cette structure ne permet pas de dériver la phrase (3a). Comme pour le complément de nom de l'exemple (1a), la complétive doit être substituée dans le groupe nominal complexe *le projet de Pinf*, lui-même substitué dans l'arbre de *désapprouver*. *Où* ne peut donc se retrouver en tête de phrase.

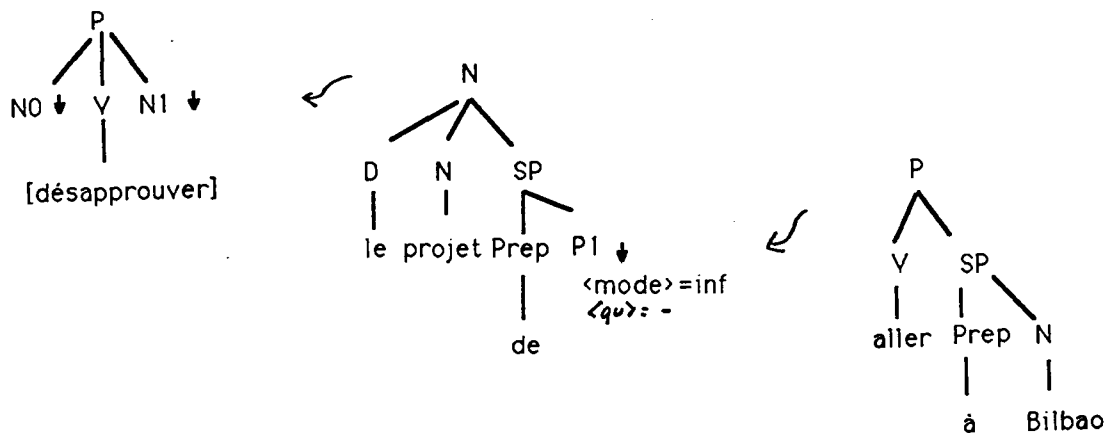


Figure 7

En revanche, dans la phrase (4), l'arbre phrastique élémentaire a une double tête *avoir* et *projet*, et la complétive est un noeud feuille. Il peut donc être un arbre auxiliaire et s'adjoindre entre *où* et *aller* pour générer la phrase (4a):

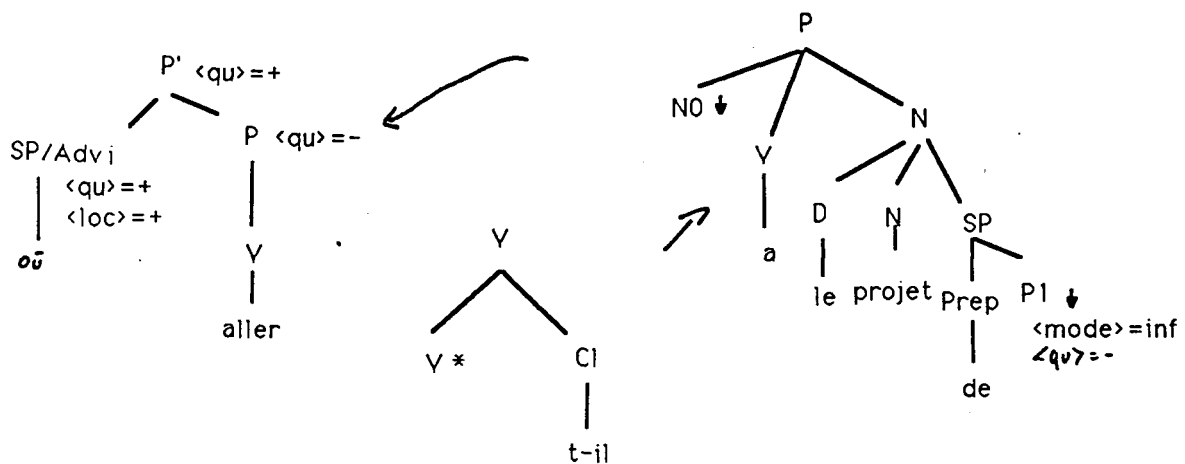


Figure 8

Il en est de même pour la relative et la clivée (Cf chapitre 4).

1.2.3 Ilôts adjectivaux

Nous traitons ici des groupes adjectivaux compléments (ou attributs). S'il s'agit d'adjectifs épithètes, la contrainte de l'îlot nominal vue supra interfère. Nous distinguons les ajouts à l'adjectifs, qui ne peuvent être extraits, des compléments (ie des arguments) de l'adjectifs qui autorisent l'extraction en phrase copulative.

a) Les modifieurs de l'adjectif

On peut avoir une question en écho sur certains modifieurs d'adjectifs (sans mouvement), mais, s'il y a question avec antéposition, c'est tout le groupe adjectif-modifieur qui doit se déplacer:

Jean était rouge de colère
Comment était-il ?
 * *De quoi était-il rouge ?*

La seule question possible est la question en "écho", c'est-à-dire sans mouvement, qui est une simple substitution d'un pronom (ou d'un SN) *qu-* à un noeud feuille SN:

Jean était rouge de quoi ? (de honte ou de colère ?)

Notons que le cas suivant serait différent:

Jean était rouge d'avoir couru
 ? *C'est d'avoir couru que Jean est rouge*

Car *d'avoir couru*, à la différence de *de colère*, n'est pas un modifieur de l'adjectif (Cf L. Picabia 1978) mais de toute la phrase:

C'est rouge de colère qu'il était.
 * *C'est rouge d'avoir couru qu'il était.*
D'avoir couru, il est tout rouge.
 ? * *De colère, il est tout rouge.*

Il ne fait donc pas partie du groupe adjectival et n'est pas concerné par la contrainte d'îlot (pour les îlots ajouts Cf infra).

b) Les compléments de l'adjectif

L'on ne peut extraire les compléments nominaux de l'adjectif qu'en phrase copulative. De même pour l'extraction hors de ses compléments phrastiques.

L'extraction n'est pas possible en phrase non copulative :

(1) *Paul a trouvé Marie contente de son nouveau chapeau*
 (1a) * *De quoi a-t-il trouvé Marie contente ?*

Il n'y a pas de structures interrogatives (ou relatives ou clivées) correspondant à *De quoi contente*, car ces structures sont définies uniquement pour les arbres à racine phrastique. C'est une contrainte linguistique, ajoutée au formalisme original. Mais, à supposer qu'il y en ait, la combinaison d'une telle structure avec celle du verbe *trouver* ne pourrait donner la phrase 1a) car celle-ci suppose de combiner simultanément *de quoi* au noeud P racine de l'arbre et *contente* au noeud Adj feuille du même arbre. C'est donc le formalisme qui l'interdit⁶.

⁶ Une variante a été proposée qui considère comme unités des ensembles d'arbres élémentaires ("TAGs multiples"); Cf chapitre 6.

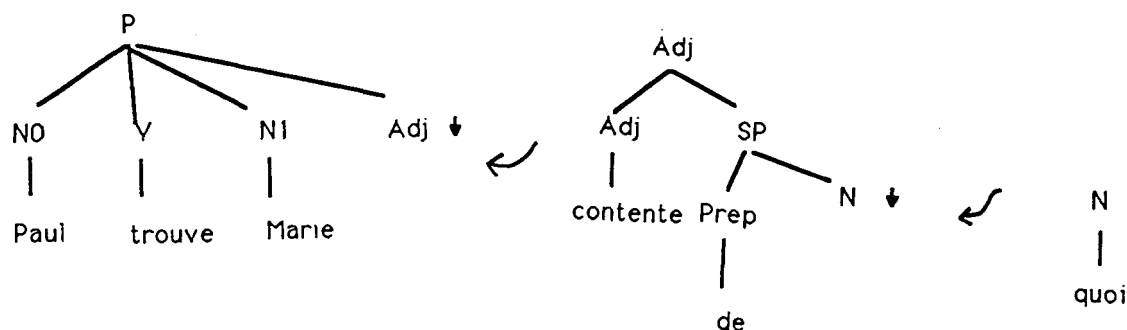


Figure 9

La seule question possible est donc celle en écho, par simple substitution de *quoi* dans l'arbre de l'adjectif:

(1b) *Tu as trouvé Marie contente de quoi ?*

La situation est différente en phrase copulative:

(2) *Marie est contente de son nouveau chapeau*

(2a) *De quoi Jean est-elle contente ?*

(2b) *Marie est contente de quoi ?*

Cette extraction est possible à partir du moment où l'on considère la phrase copulative comme la structure élémentaire de l'adjectif. Il s'agit alors d'une structure à tête double: *être* et l'adjectif à laquelle correspond une famille d'arbres phrastiques élémentaires, comme pour les autres prédicats (Cf chapitres 3 et 4). Le complément de l'adjectif est bien un syntagme des arbres phrastiques correspondants et des structures élémentaires sont définies dans la famille de l'adjectif pour son extraction :

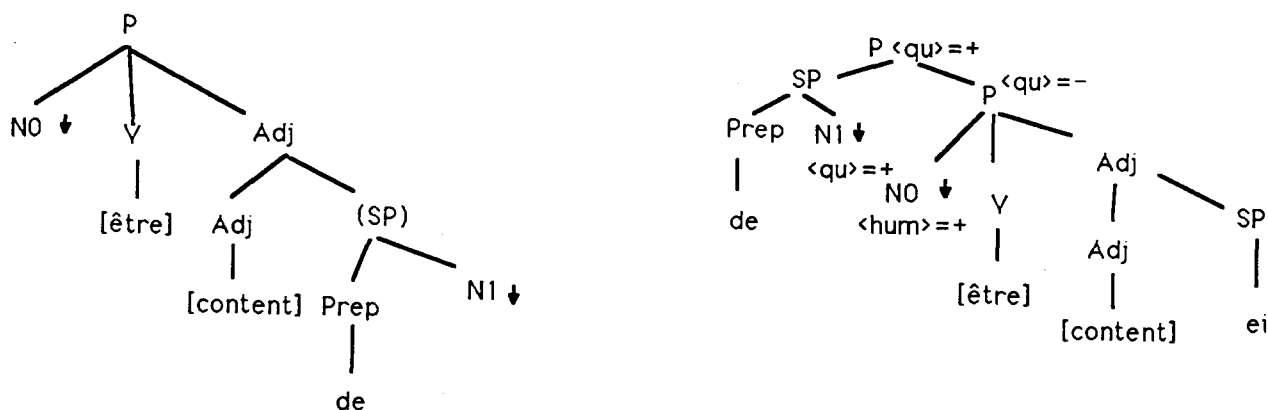


Figure 10

Comme dans le cas des phrases à verbe support, il y a donc une double question correspondant à la phrase (2): celle qui porte seulement sur le complément de l'adjectif et celle qui porte sur tout le groupe adjectif + complément⁷:

⁷ La tête adjectivale peut être optionnelle dans l'entrée lexicale pour permettre la sélection de l'arbre interrogatif en *comment* dans la famille ainsi que la pronominalisation en *le*).

(2d) *Comment est Jean ?*

Nous traitons les phrases attributives à verbe autre que *être*, par exemple *rester*, *devenir* ou *paraître*, comme des phrases à tête verbale où l'adjectif est substitué. L'extraction est donc exclue:

(3) *Jean devient dégoûté de son nouveau chapeau*

(3a) *?* De quoi Jean devient-il dégoûté ?*

Dans le cas d'un complément phrastique, le contraste est le même entre phrase à verbe simple et phrase copulative:

(4) *Jean a trouvé Marie contente que tu connaisses Paul*

(4a) ** Qui a-t-il trouvé Marie contente que tu connaisses ?*

Un arbre élémentaire correspondant à la question *Qui tu connais* existe bien mais, en tant que complément de l'adjectif, non du verbe principal, devrait être substitué dans l'arbre AP, lui-même substitué dans l'arbre de *trouver*⁸. *Qui* (correspondant à un complément de *connaître*) ne peut donc se retrouver en tête de phrase.

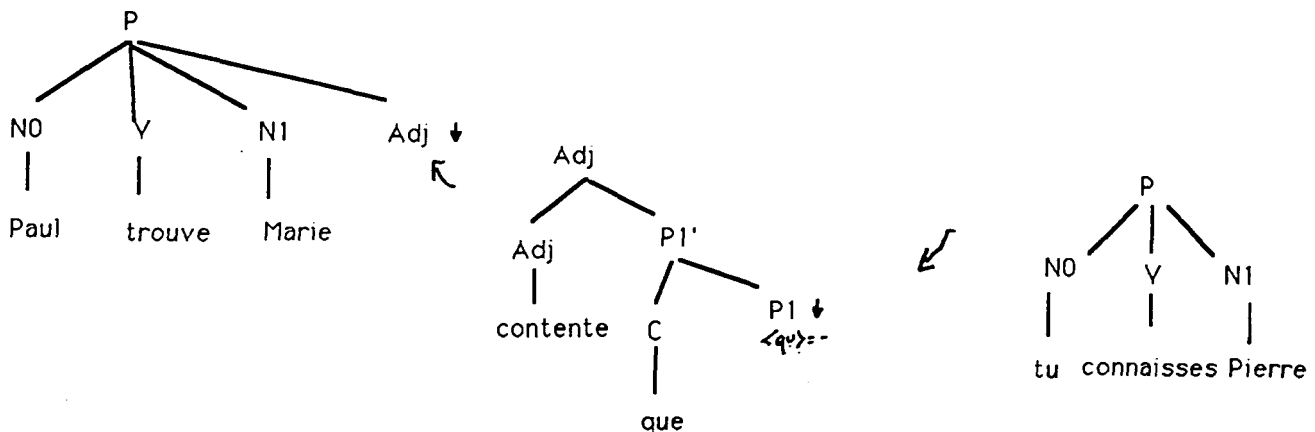


Figure 11

Le seul type de question possible est donc:

J'ai trouvé Marie contente que tu connaisses qui ?

La situation est différente en phrase copulative:

(5) *Marie est contente que tu connaisses Paul*

(5a) *Qui Marie est-elle contente que tu connaisses ?*

⁸ La substitution de l'arbre de la question dans l'arbre de *content* est par ailleurs exclue par un trait $\langle qu \rangle = -$. Le groupe *contente que qui tu connais* ne sera pas une dérivation bien formée, en revanche *contente que je sache qui tu connais(ses)* le serait.

L'adjectif est tête avec *être* d'un arbre élémentaire phrastique, dont la complétive est un noeud feuille. Il peut donc s'agir d'un arbre auxiliaire, qui comme tel s'adjoindra entre *Qui* et *tu connais*:

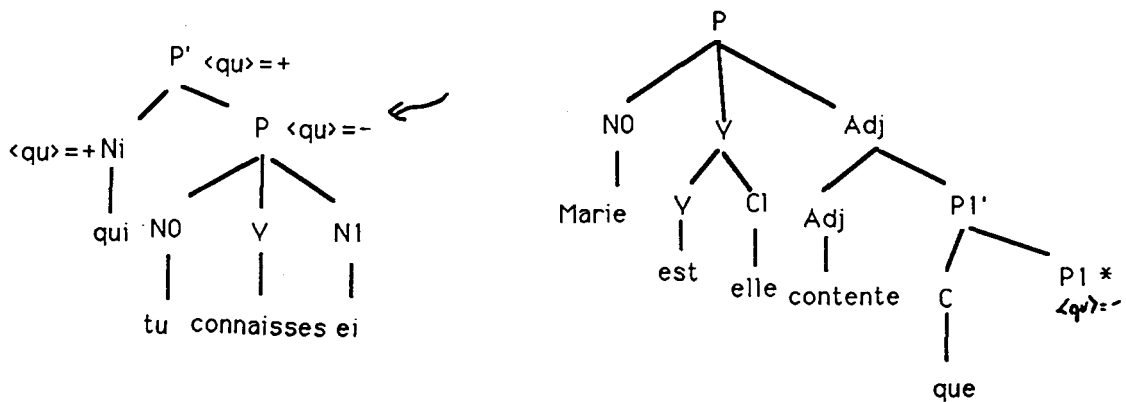


Figure 12

Il s'agit ici d'une possibilité structurale, qui ne présume rien de possibles blocages lexicaux. Pour certains adjectifs à complétive, l'extraction en phrase copulative est plus difficile (voir infra p. 223).

1.2.4. Ilôt prépositionnel

a) Syntagmes prépositionnels compléments

Nous ne nous intéressons qu'aux compléments de verbes, le cas des compléments de nom et d'adjectif ayant été vu plus haut.

Les SP étant représentés comme des syntagmes étendus, l'extraction est théoriquement possible de tout le SP ou du SN qu'il domine. C'est donc un principe supplémentaire qui oblige à ne pas déplacer le SN seul (à la différence de l'anglais par exemple).

En fait, dans le cas des phrases clivées, on trouve bien la double extraction prédite (Cf chapitre 3):

C'est à Marie que Jean pense

C'est Marie à qui Jean pense.

Dans le cas des relatives, le fait que l'antécédent soit nominal et non prépositionnel, est aussi une quasi-extraction hors de SP.

La contrainte supplémentaire concerne donc surtout les questions. Il s'agit d'un principe général (en amont de l'écriture de la grammaire), qui gouverne la bonne formation des structures élémentaires correspondantes (il n'a donc pas à être formulé explicitement): seuls les syntagmes appartenant à un arbre élémentaire phrastique et non dominés par SP peuvent donner naissance à un arbre question. Dans la famille de N_0 *penser à* N_1 , on aura donc les arbres 1, 2 et 3 mais pas 4:

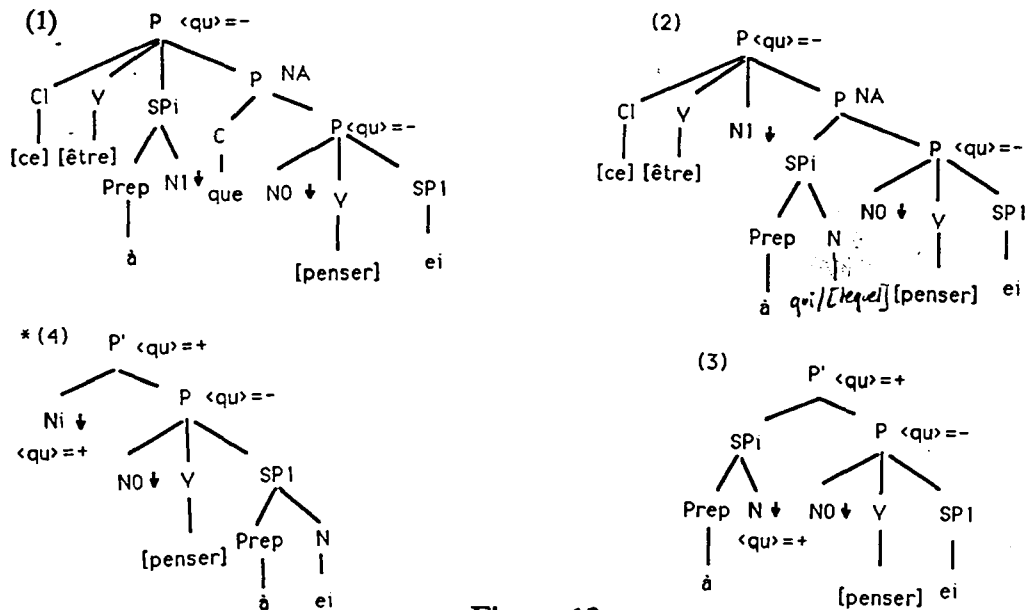


Figure 13

De manière générale, quand les raisons sémantiques (d'avoir un SP étendu) et les raisons syntaxiques (d'en faire une feuille, et donc un îlot) s'opposent, nous avons privilégié la vraisemblance sémantique (et le caractère naturel des unités définies) par rapport à l'élégance syntaxique.

Lorsque c'est un argument phrastique qui est dominé par SP, l'extraction est possible comme hors d'une complétive directe. La complétive, malgré son appartenance à SP est une feuille de l'arbre du verbe principal. Cet arbre phrastique peut donc être un arbre auxiliaire et peut s'adjoindre à l'intérieur d'un arbre de question par exemple:

J'ai rêvé de rencontrer Paul
Qui as-tu rêvé de rencontrer ?

Jean rêve de ce que Marie aille à Bilbao
Où Jean rêve-t-il de ce que Marie aille ?

Il y a néanmoins, comme dans le cas des verbes à complétives, des restrictions supplémentaires à ces extractions (voir infra).

b) Les syntagmes prépositionnels ajouts

Ils sont représentés par des arbres auxiliaires en P ou en V. Leur argument nominal (ou adverbial) ne peut être extrait:

- (1) *Je suis parti (pour + à cause de) Marie*
 (1a) * *Qui es-tu parti (pour + à cause de) ?*
 (1b) *Tu es parti (pour + à cause de) qui ?*

S'il s'agit d'ajouts à V, la structure même de l'arbre auxiliaire empêche la phrase (1a), puisque *Qui*, pour être en tête de phrase, doit être un ajout à P.

S'il s'agit d'ajouts à P, c'est un principe linguistique qui empêche la phrase (1a) et impose la phrase (1b):

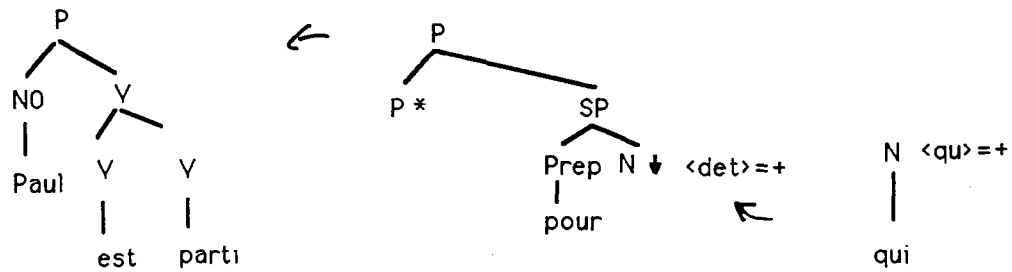


Figure 14

Les conjonctives, introduites par une conjonction ou par une préposition, ne permettent pas en général l'extraction:

Je pars à Bilbao (pour aider Marie + parce que Bob aime Marie)

* *Qui es-tu parti à Bilbao pour aider ?*

* *Qui es-tu parti à Bilbao parce que Bob aime ?*

Les arbres des questions existent bien : *qui aider* ou *qui Bob aimait*, mais ils sont, en tant qu'arguments phrastiques de la préposition ou de la conjonction, obligatoirement substitués. *Qui* ne peut donc se retrouver en tête de la conjonctive, et les seules questions possibles sont:

Tu es parti à Bilbao (pour aider qui + parce que Bob aime qui) ?

Ici, l'arbre élémentaire de la conjonction (ou de la préposition) est bien de racine phrastique. Le blocage n'est donc pas le même que pour les îlots nominaux et adjectivaux. C'est déjà un arbre auxiliaire (le noeud pied correspond à la phrase principale modifiée par la conjonctive). Le noeud où se combine la question ne peut donc être le pied de cet arbre auxiliaire, et la conjonction ne peut donc s'adjoindre entre *Qui* et *Bob aime*:

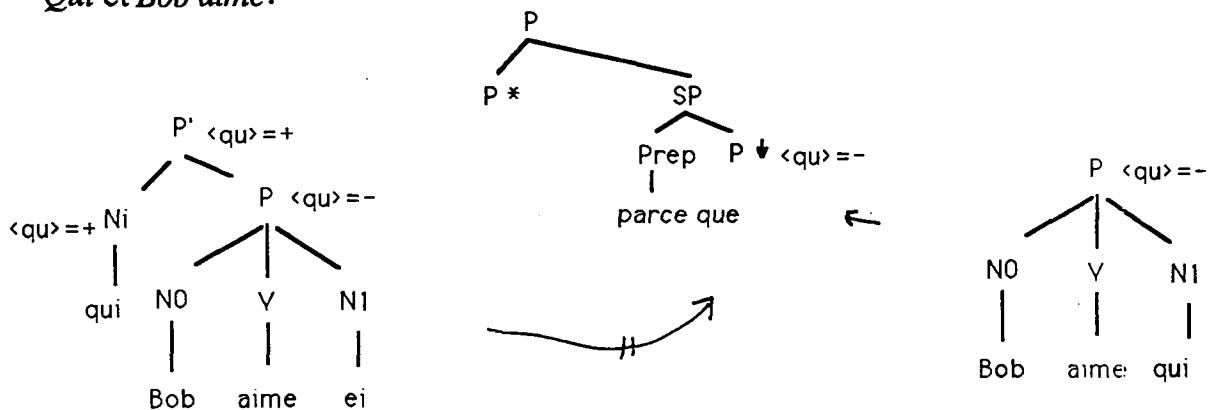


Figure 15

Comme pour les autres phénomènes d'îlots, nous n'avons donc pas besoin d'une contrainte particulière. C'est la structure même des arbres élémentaires qui impose le résultat souhaité.

Nous ne traitons pas des traces parasites, qui sont des cas où l'extraction hors de la subordonnée est rendue possible par une extraction parallèle hors de la principale.

Quel livre as-tu rangé sans lire ?

Il y a alors mise en facteur de l'élément *qu-* antéposé.

On traite selon le même mécanisme les îlots coordonnés (Cf J. Ross 1967). On ne peut extraire un élément hors d'un des syntagmes coordonnés:

- (1) *Jean aime le chocolat et Paul préfère la banane*
 (1a) * *Qu'aime Jean et Paul préfère la banane ?*
 (1b) * *Que Jean aime le chocolat et Paul préfère-t-il ?*

On représente les conjonctions de coordination comme des arbres à deux (ou plus) éléments enchâssés (l'étoile de Kleene * ne change rien à la capacité générative du formalisme).

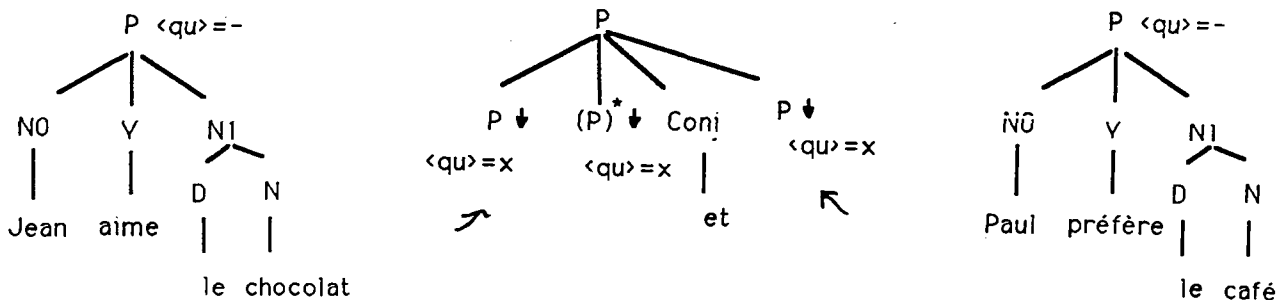


Figure 16

La phrase (1b) est impossible car le deuxième (comme le nième; $n > 1$) membre de la coordination est substitué. La phrase (1a) est impossible pour une autre raison: le parallélisme des traits *qu-* dans chaque structure coordonnée :

- (2) *Qui Marie aime-t-elle et qui Jean aime-t-il ?*
 (2a) * *Qui Marie aime-t-elle et Jean est parti ?*
 (2b) * *Jean est parti et qui Marie aime-t-elle ?*
 (3) *Qui penses-tu que Marie aime et qui crois-tu que Paul déteste ?*

Quand il s'agit d'une question indirecte, le trait *qu-* ne "remonte" pas au niveau de principale et le parallélisme peut être rompu:

- (4) *Je demande qui Marie aime et tu penses à Paul.*
 (5) *Je pensais à Paul et Marie a demandé qui Jeanne aimait.*

L'interdiction de la phrase (1b) est en fait plus sévère que celle de (1a), surtout si la coordination a un sens d'opposition:

? Quel examen; les étudiants peuvent-ils rater et avoir quand même leur diplôme en juin ? (traduit de J. Goldsmith 1985)

De fait, elle correspond dans notre représentation à une contrainte de trait, non à une contrainte structurale. L'on pourrait donc relâcher la contrainte sur les traits *qu-* des phrases coordonnées.

En revanche, nous ne traitons pas des cas de double extraction qui réclament un mécanisme particulier de mise en facteur de sous-arbres (Cf chapitre 6):

Qu'est-ce que Marie adore aujourd'hui et détestait autrefois ?

En conclusion, nous pouvons donc reformuler les contraintes d'îlots comme suit: l'extraction ne peut s'appliquer qu'aux syntagmes arguments d'arbres élémentaires de racine phrastique. A cela s'ajoute le principe de bonne formation des arbres élémentaires et la définition même des arbres auxiliaires qui ne peuvent avoir qu'un seul noeud pied.

Le principe A/A (ou principe de la sous-jacence) est utile en grammaire générative, où des transformations s'appliquent à des arbres dérivés, c'est-à-dire à des structures syntagmatiques de profondeur non bornées: il peut y avoir un nombre indéfini de SN emboîtés, ou de SP ... et il faut savoir à quels niveaux appliquer ces transformations. Il pose des problèmes en ce qu'il est trop contraignant puisqu'il est parfois possible d'extraire une catégorie hors d'un syntagme de même catégorie.

En TAG, il est inutile. Les transformations s'appliquent seulement aux arbres élémentaires, qui sont par définition des arbres de profondeur limitée. Mais rien n'empêche a priori ces arbres élémentaires de comporter des syntagmes enchâssés, et cela est même nécessaire pour pouvoir faire apparaître les différentes têtes lexicales des phrases à verbe support et des phrases figées, et garantir l'unité sémantique et argumentale de celles-ci (notion d'arbres élémentaires étendus, Cf supra chapitre 4). Or, dans ce cas, il est justement souvent possible d'extraire une catégorie hors d'un syntagme de même catégorie.

C'est donc le principe constitutif des arbres élémentaires (le fait qu'ils correspondent à une phrase simple) qui rend le principe A/A superflu. Mais cette reformulation permet également de prévoir naturellement des cas où le principe A/A serait pris en défaut.

2. Dépendances non bornées

Nous exposons le problème général que posent les extractions sur des distances non bornées. Nous présentons différentes solutions employées par différents formalismes (grammaires en chaîne, GPSG et LFG) avant de présenter notre représentation en TAG.

2.1. Discussion théorique

2.1.1 Problème général

Il s'agit de garder en mémoire l'élément antéposé de façon à:

- reconnaître comme complète une séquence autrement malformée, par exemple s'il y a mouvement d'un complément obligatoire :

*Jean ressemble (*E + à Paul)*

*A qui crois-tu que Jean ressemble (E + *à Marie) ?*

- vérifier les contraintes d'accord entre sujet et verbe ou entre objet et participe passé (selon que c'est le sujet ou l'objet qui est extrait)

*Marie semble devoir pouvoir être (arrêtée + *arrêté)*

*Les enfants que l'on pense que (..) Jean a (vus + *vu)*

- vérifier les contraintes de sous-catégorisation (pour l'extraction des sujets et des compléments, obligatoires ou non)

*Jean (aime + *pense) Marie*

*C'est Marie que tout le monde sait (...) que Jean (aime + *pense)*

*Jean compte (sur + *à) Marie*

*C'est (sur + *à) Marie que tout le monde sait (...) qu'il compte*

- vérifier les contraintes de co-sélection (pour l'extraction des arguments et des modificateurs):

*Il viendra (*hier + demain)*

*C'est (*hier + demain) que tout le monde pense (...) qu'il viendra*

- vérifier des contraintes lexicales, qui peuvent permettre ou interdire l'extraction hors de certains compléments :

*C'est Marie qu'il a (dit + ?*hurlé) qu'il aimait*

*C'est contre Marie que tout le monde pense (...) qu'il a (*rapporté + commis) une*

agression

2.1.2 Aspects informatiques

Il est bien connu (notamment depuis G. Gazdar 1981) que les dépendances non bornées, du type:

Qui croyez-vous que Marie pense que Jean aime ?

peuvent être représentées dans une grammaire hors contexte. Elles peuvent donc l'être a fortiori dans une grammaire contextuelle. Ce n'est donc pas le fait qu'elles puissent représenter ce genre de phénomène qui fait l'originalité des TAG, mais la façon dont elles le représentent par opposition aux autres grammaires contextuelles ou hors contexte.

a) M. Salkoff 1973

M. Salkoff 1973 utilise une catégorie $\emptyset_{om}$ pour la trace et une marque -X pour la catégorie qui manque. La même catégorie (et la même marque) est utilisée en cas de pronominalisation (et en cas d'inversion du sujet).

Il note (p. 45) que $\emptyset_{om}$ peut être à une profondeur non bornée dans le cas d'une trace objet, mais n'explique pas le principe de propagation de la marque. Il n'est pas non plus précisé ce qu'il en est des contraintes d'îlots.

Tous les compléments sont mis à plat et regroupés sous une variable Objet (Ω), les infinitives sont mises à plat. Il lui faut donc préciser dans quels cas c'est tout le groupe objet qu'il faut extraire:

Jean pense à Marie (Ω = Prep N)

C'est à Marie qu'il pense

** C'est Marie qu'il pense à,*

dans quels cas c'est seulement une partie de l'objet :

Jean donne ce livre à Marie (Ω = N Prep N)

** C'est ce livre à Marie qu'il donne*

C'est à Marie qu'il donne ce livre

C'est ce livre qu'il donne à Marie

Dans ce dernier cas, il faut également préciser quelle partie du complément peut s'extraire :

J'ai vu partir Jean (Ω = V_{inf} N)

C'est Jean que j'ai vu partir

** C'est partir que j'ai vu Jean*

** C'est partir Jean que j'ai vu*

L'expression des contraintes d'extraction est également compliquée par le fait que dans les phrases à temps composé, l'objet est constitué de toute la chaîne qui suit l'auxiliaire et inclut le verbe au participe passé.

Cette grammaire doit donc autoriser l'extraction de non syntagmes, et se prête difficilement à des généralisations sur les contraintes d'extraction.

b) La grammaire syntagmatique généralisée (GPSG)

L'on peut se reporter le traitement effectué pour l'anglais dans Gazdar et al. 1985, qui a d'ailleurs été repris pour le français par G. Bès 1988. GPSG définit un trait particulier Slash, qui a pour valeur la catégorie qui manque à sa place. Il est introduit par des règles syntagmatiques comme, la suivante (proposée pour la topicalisation en anglais) :

$$S \rightarrow X^2 H / X^2$$

La marque / note en fait ce trait particulier (noté Slash) à valeur catégorielle (c'est-à-dire dont la valeur est une structure de trait possiblement complexe). L'élément vide (catégorie nulle) est défini comme: X^2/X^2 (avec la même catégorie X) ou comme⁹ :

$$\left[\begin{array}{l} N = x \\ V = y \\ \text{barre} = 2 \\ \text{slash} = \left[\begin{array}{l} N = x \\ V = y \\ \text{barre} = 2 \end{array} \right] \end{array} \right]$$

L'unification opérant sur ce trait comme sur les autres, les sous-arbres suivants sont autorisés (1) ou rejetés (2, 3 et 4, notés:*):

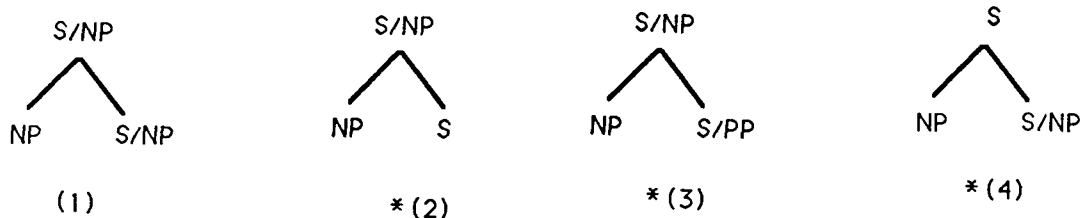


Figure 17

Ce trait SLASH a un mode de propagation particulier puisqu'il est à la fois un "trait de pied" (*foot-feature*) et un "trait de tête" (*head-feature*): en tant que trait de tête, il doit être propagé d'un syntagme à son syntagme tête (de la phrase au groupe verbal). On aura ainsi par exemple:

⁹ Les traits N et V servent à désigner les catégories Nom, verbe, adjectif et préposition selon la décomposition de N. Chomsky 1971.

S/NP -> NP VP/NP

VP/NP -> V NP/NP (avec NP/NP = $\emptyset$)

En tant que trait de pied, il peut être aussi propagé à d'autres syntagmes dominés (des principales aux complétives, et en anglais des groupes prépositionnels aux groupes nominaux enchâssés).

Ainsi les sous-arbres suivants sont autorisés (1 et 5) ou interdits (2, 3 et 4):

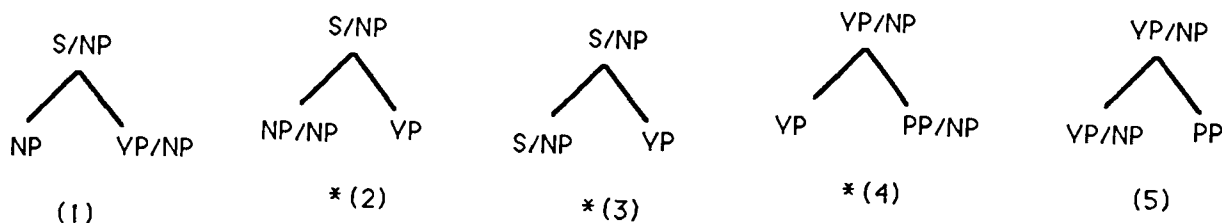


Figure 18

La plupart des contraintes d'îlot sont donc prises en compte par ces principes généraux sur la propagation des traits. L'interdiction du sous-arbre 4 équivaut à la définition d'un îlot adverbial en grammaire générative. L'interdiction de 2 et 3 équivaut à la contrainte de l'îlot sujet phrastique (VP étant tête de S, il doit porter le trait slash lui aussi). Il est à noter qu'elle repose crucialement sur l'interposition d'un syntagme SV entre V et P, défini comme tête de la phrase, ce qui serait difficilement justifié pour le français (Cf chapitre 2).

A ces principes de propagation s'ajoutent quelques contraintes supplémentaires: par exemple, que SLASH est impossible sur des préterminaux (ce qui force à extraire SN et empêche d'extraire N seul).

Le même mécanisme est utilisé pour la coordination à "gapping".

Cette représentation, de loin la plus élaborée pour ce qui est des contraintes d'extraction, pose plusieurs problèmes, linguistiques et informatiques:

- la propagation des traits peut se faire sur une longueur non bornée, ce qui rend les calculs lourds.
- les contraintes reposent sur des règles syntagmatiques parfois discutables, par exemple l'existence d'un syntagme VP entre la phrase et le verbe.
- il est difficile d'exprimer des contraintes lexicales sur les conditions d'extraction (verbes support, verbes ponts).

c) La grammaire lexicale fonctionnelle (LFG)

Au départ (J. Bresnan 1982b), les phénomènes d'extraction étaient traités au niveau de la structure en constituants, par un mécanisme proche de celui de GPSG.

R. Kaplan et A. Zaenen 1989 ont récemment défini les extractions de SN au niveau de

la structure fonctionnelle, ce qui permet de rendre compte plus aisément de contraintes sémantico-lexicales sur la topicalisation. La règle de topicalisation est par exemple, au niveau syntagmatique:

$$\begin{array}{l} S' \rightarrow XP \quad S \\ (\uparrow \text{Topic} = \downarrow) (\uparrow = \downarrow) \\ (\uparrow \text{Topic}) = (\uparrow \text{GF}) \end{array}$$

Ce qui signifie qu'une phrase topicalisée (ici S') a une fonction *topic* dans sa structure fonctionnelle dont la valeur est d'une part celle du syntagme XP antéposé (c'est-à-dire l'ensemble des traits de ce syntagme), d'autre part celle d'une fonction grammaticale quelconque de S (notée ici par la variable GF ; c'est-à-dire l'ensemble des traits associés à cette fonction) : sujet (Subj), objet direct (Obj), complément prépositionnel (Obl).... Cette formulation n'utilise pas de catégorie vide : le même syntagme XP remplit deux fonctions à la fois et c'est le principe de consistance fonctionnelle qui empêche (pour les fonctions sous-catégorisées) à un syntagme de même fonction d'être réalisé dans S . Avec $GF = \text{Obj}$, on aura donc bien:

John, Mary loves
* *John, Mary loves Peter*

Pour les dépendances à distance, il faut introduire un procédé particulier, l'Incertitude Fonctionnelle, qui permet d'exprimer une contrainte d'unification sur un élément enchâssé dans un nombre indéterminé de structures fonctionnelles. Si l'on considère que cet élément doit toujours, en anglais, être enchâssé dans une suite de compléments phrastiques (fonction *Comp*), mais pas de groupes sujets ou de compléments nominaux, l'équation fonctionnelle pour la topicalisation de l'objet par exemple devient :

$$(\uparrow \text{Topic}) = (\uparrow \text{Comp} * \text{Obj})$$

Ce qui condense les équations suivantes:

$$\begin{array}{ll} (\uparrow \text{Topic}) = (\uparrow \text{Obj}) & \text{John, Mary loves} \\ (\uparrow \text{Topic}) = (\uparrow \text{Comp Obj}) & \text{John, Bob thinks that Mary loves} \\ (\uparrow \text{Topic}) = (\uparrow \text{Comp} (\uparrow \text{Comp Obj})) & \text{John, Bob says that Sue thinks that Mary loves} \\ & \text{loves} \\ & \dots \end{array}$$

Ces équations sont des équations sur les valeurs des traits *Obj* et *Topic*, qui sont elles-mêmes des structures de traits. Toutes les contraintes portant sur la valeur de l'objet (par exemple celles imposées par le verbe de la complétive) doivent se retrouver en valeur du trait *Topic*. On obtient donc les schémas fonctionnels suivants pour les phrases : *John, Mary loves* et *John, Bob thinks that Mary loves*

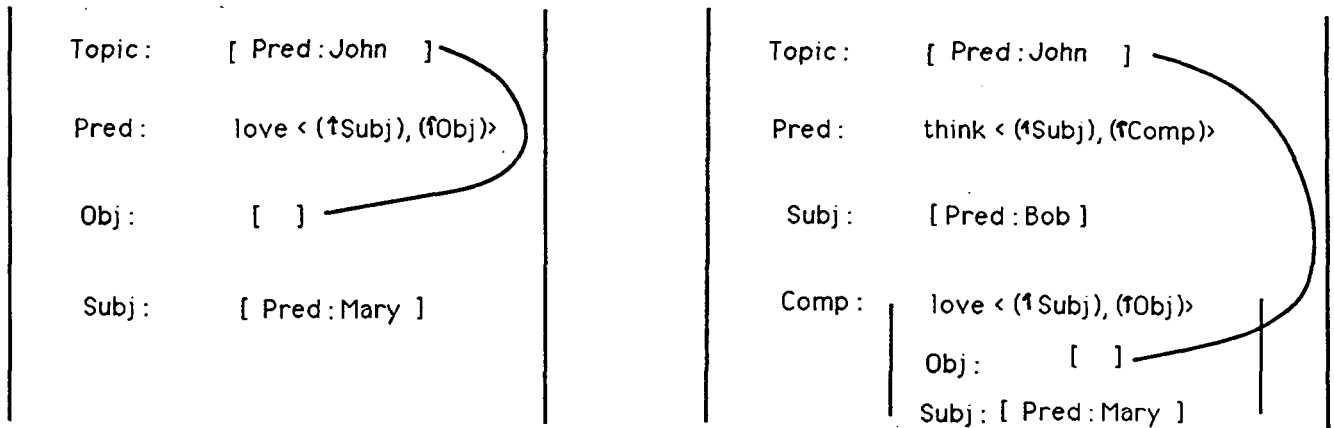


Figure 19

Il n'y a pas de traces (ou d'éléments vides) et ce sont des principes généraux de bonne formation des structures syntagmatiques et fonctionnelles qui assurent l'absence du SN normalement Objet dans la complétive la plus enchâssée, puisque la fonction Obj est déjà occupée par le syntagme topicalisé. (le même procédé assure les restrictions lexicales désirées sur ce syntagme).

Il est à noter que la représentation repose sur une distinction entre fonctions grammaticales occupées par des catégories différentes: ainsi pour les compléments, les complétives ont la fonction Comp, tandis que les SN (passivables) ont la fonction Obj. En revanche tous les sujets sont représentés par la même fonction Subj, et sont des îlots comme les SN objets.

Il est vraisemblable que le même traitement devrait être retenu pour les "mouvements" d'éléments *qu*- également.

Comme en GPSG, l'unification doit se faire sur une longueur non bornée (la structure fonctionnelle a autant d'enchâssements que la phrase a de compléments phrastiques).

Certaines contraintes lexicales peuvent être plus faciles à exprimer qu'en GPSG, par exemple en notant Obj (et non Comp) la fonction des complétives des verbes non ponts (voir infra p. 268).

Il est à noter que le procédé d'incertitude fonctionnelle, qui nécessite un algorithme de calcul particulier, serait un simple corollaire en TAG (A. Joshi et K. Vijay-Shanker 1989).

2.2 Représentation en TAG

Nous allons montrer comment nous utilisons l'adjonction pour représenter les dépendances à distance, et comment un usage alternatif de la substitution permet d'empêcher ces extractions sur des longueurs non bornées.

Notre représentation repose crucialement sur la distinction entre les deux opérations du

formalisme: la substitution et l'adjonction, c'est-à-dire sur la distinction entre arbre initial (substitué) et arbre auxiliaire (adjoint).

Rappelons que, pour qu'une entrée puisse être représentée par un arbre auxiliaire, il faut qu'elle ait un noeud pied de la même catégorie que son noeud racine. Ainsi les verbes (ou tous les items têtes de structures phrastiques) ne peuvent être représentés par un arbre auxiliaire que s'ils ont un argument phrastique. Les verbes à arguments nominaux sont donc nécessairement représentés par des arbres initiaux (c'est le formalisme qui l'impose).

La condition sur l'identité des catégories entre noeud feuille et noeud racine est une condition nécessaire mais non suffisante: un verbe à argument phrastique peut très bien être représenté par un arbre initial. Ce sont des raisons linguistiques qui vont justifier, pour les verbes à arguments phrastiques, le choix entre arbre initial et arbre auxiliaire. Nous serons amenée à faire des distinctions entre familles de verbes à argument phrastique et aussi entre verbes d'une même famille. Le principe général est le suivant: les structures permettant les dépendances à distance sont représentées par des arbres auxiliaires, les structures bloquant de telles dépendances sont représentées par des arbres initiaux.

Rappelons aussi que la combinaison d'un arbre auxiliaire et d'un arbre initial donne un arbre initial et que seule la combinaison de deux arbres auxiliaires donne un arbre auxiliaire.

2.2.1 Adjonction et dépendances à distance

Contrairement aux autres grammaires d'unification, nous n'utilisons pas de traits pour les extractions à distance, qui peuvent être représentées directement grâce aux propriétés de l'opération d'adjonction.

En effet, en TAG, nous considérons les questions, les relatives, les phrases clivées ... comme des arbres élémentaires sélectionnés par le verbe tête. Ainsi les verbes simples à deux arguments nominaux libres (par exemple *aimer*) sélectionnent l'ensemble des arbres élémentaires correspondants à leurs différentes réalisations syntaxiques de surface (Cf chapitre 3):

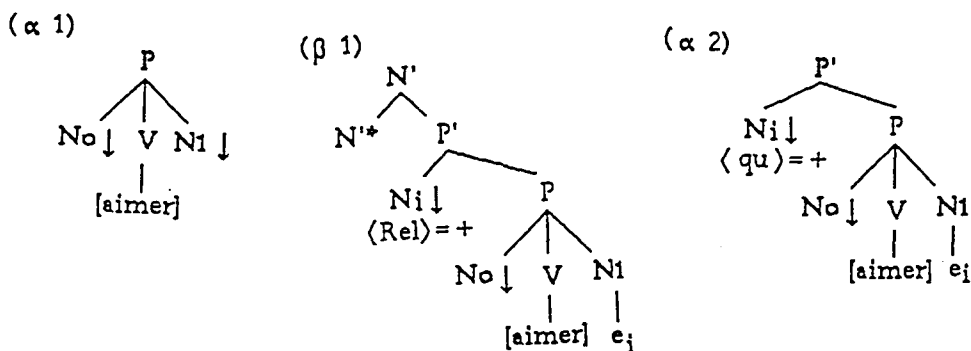


Figure 20

Deux cas de dépendances non bornées se présentent: avec les modaux et auxiliaires, avec les verbes à complétives. Dans les deux cas, c'est l'opération d'adjonction qui permet le résultat souhaité.

a) auxiliaires et modaux

Comme il a été vu au chapitre 3, ils sont représentés comme des ajouts au verbe et peuvent s'adjoindre de façon consécutive. Ces adjonctions se passent de la même façon, que l'arbre phrastique comporte une extraction ou non (modulo les éventuelles contraintes d'inversion expliquées infra p. 223):

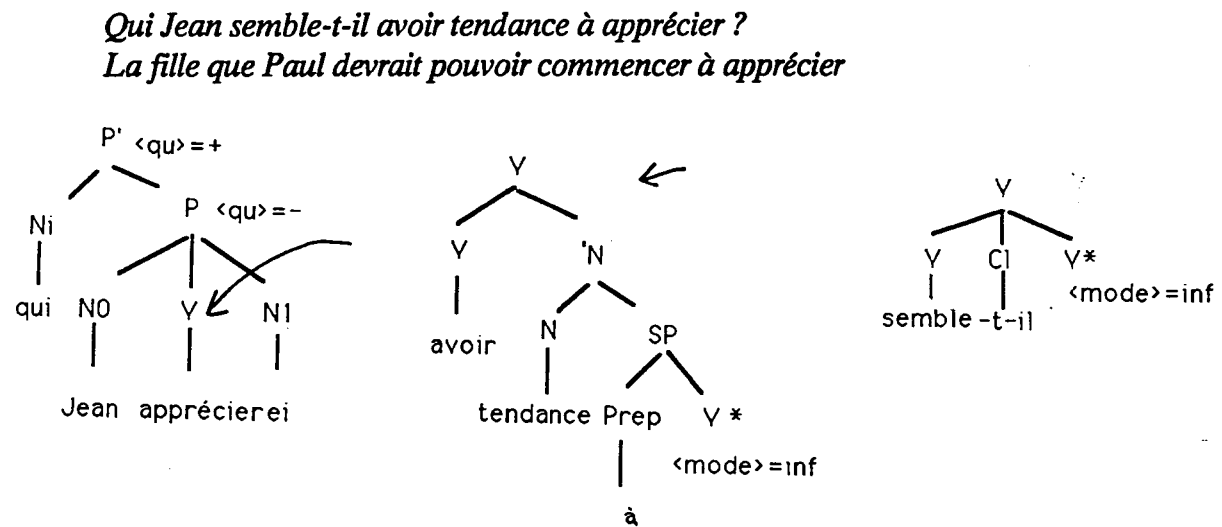


Figure 21

b) verbes à complétives

Les verbes à complétives ou infinitives sont représentés en général par des arbres auxiliaires à racine phrastique (Cf chapitre 3).

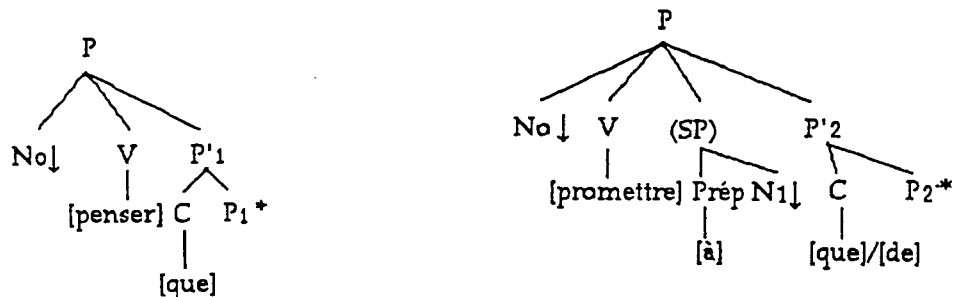


Figure 22

Leur adjonction est donc possible quand la complétive comporte une extraction, à condition qu'il y ait un noeud P intermédiaire (nécessaire également pour l'adjonction d'adverbiaux).

L'on génère donc bien par adjonctions successives:

Quelle fille Paul pense-t-il que Bob sait que (...) Jean aime ?

La fille que Paul pense que Bob sait que (..) Jean aime ...

C'est la fille que Paul pense que Bob sait que (...) Jean aime.

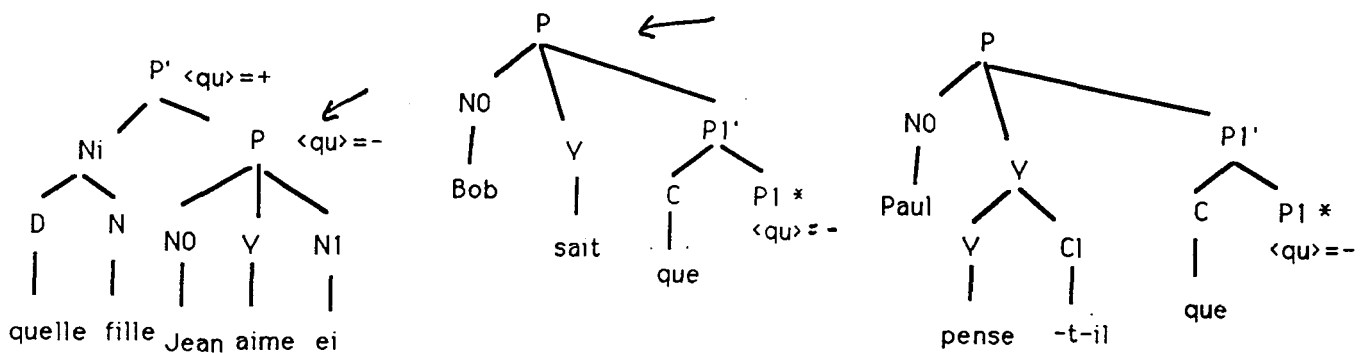


Figure 23

Les complétives en *ce que* P qui permettent les mêmes extractions que les complétives en *que* P sont représentées de la même façon que ces dernières, mais sont réservées aux structures à complément prépositionnel.

Jean cherche à ce que Marie aille à Bilbao

Où Jean cherche-t-il à ce que Marie aille ?

L'alternance *Prep ce que* P / *que* P , observée avec certaines prépositions (et certains verbes), est représentée par un arbre à complétive directe dans la même famille du verbe (qui ne sera sélectionné que si la préposition est marquée comme optionnelle dans l'entrée lexicale (cf chapitre 3)).

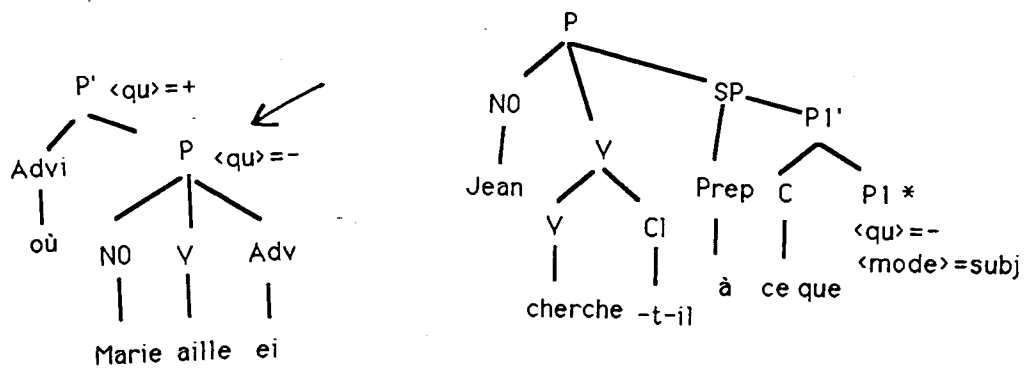


Figure 24

* infinitives et complétives

L'on représente les infinitives comme des réductions de complétives, c'est-à-dire par des arbres phrastiques (Cf chapitres 2 et 3). La particularité des arbres infinitifs est l'absence de sujet réalisé en français, sauf après certains verbes comme *voir*, *entendre* et dans les tournures factitives. Deux possibilités se présentent: soit dédoubler tous les arbres de toutes les familles (sauf ceux à extraction du sujet) en créant autant d'arbres sans sujets; soit noter le noeud sujet comme optionnel dans tous les arbres (toujours sauf ceux à extraction du sujet).

Il est plus satisfaisant d'utiliser les mêmes structures phrastiques quel que soit leur mode (et donc les mêmes arbres pour complétives et infinitives), mais cela complique le système de traits à employer, puisqu'il y a corestriction entre la présence du sujet et le mode de la structure. Une façon d'exprimer cette corestriction serait par exemple la suivante (pour tout noeud P)¹⁰:

$\langle \text{suj} \rangle = - \Rightarrow \langle \text{mode} \rangle = \text{imper/inf}$
 $\langle \text{suj} \rangle = + \Rightarrow \langle \text{mode} \rangle = \text{subj/ind}$

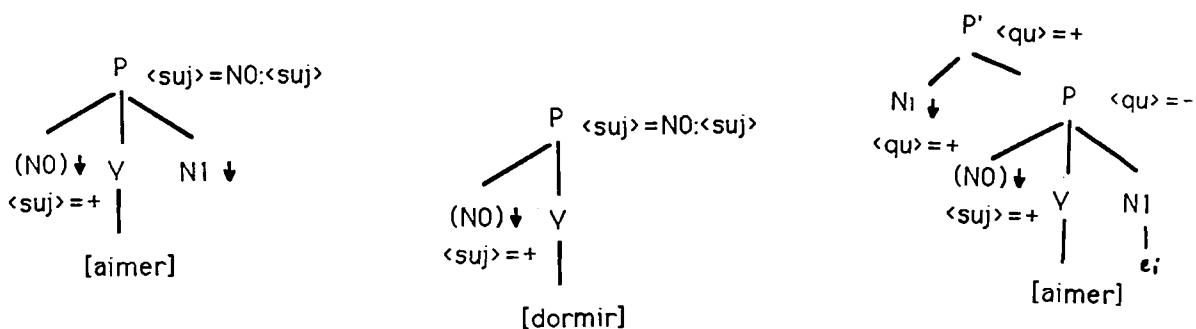


Figure 25

¹⁰ Si l'on voulait intégrer les verbes qui admettent une complétive à l'infinitif avec sujet réalisé (*entendre*, *voir*, *laisser*...), cette corestriction de traits, au lieu d'être générale, devrait être dépendante du verbe de la principale (c'est-à-dire notée au noeud pied dans la structure de ces verbes).

Le trait <subj> en P n'a la valeur + que si le noeud sujet est présent dans l'arbre. Sinon sa valeur est non spécifiée. Il pourra donc alors s'unifier avec un trait subj=-, comme celui qu'on trouve au noeud pied de *N₀ vouloir* P mais non à celui de *N₀ demander* P qui peut avoir des modes plus variés :

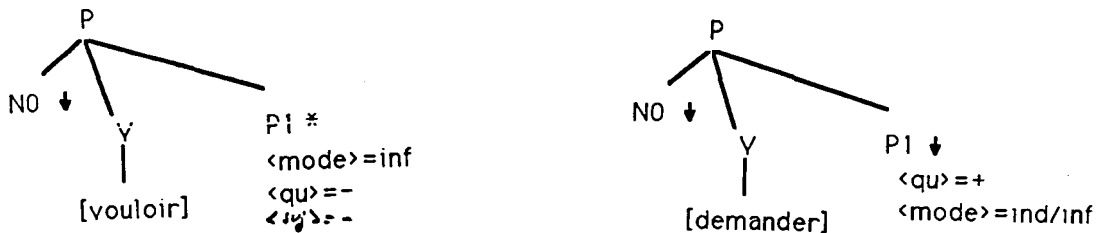


Figure 26

Dans l'implémentation actuelle de la grammaire, c'est la solution du dédoublement des arbres qui a été retenue, pour des raisons d'efficacité de traitement. Les valeurs du trait <mode> sont directement spécifiées au noeud P :

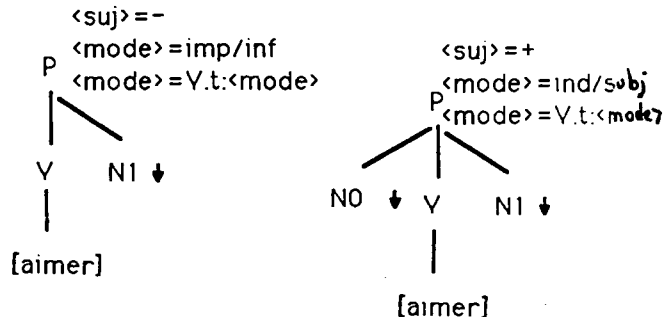


Figure 27

Les arbres des infinitives sont également utilisés pour les phrases à l'impératif.

* participiales et complétives

L'on considère les participiales aussi comme des réductions de complétives (Cf chapitre 3). Elles utilisent les mêmes arbres que les phrases à verbe conjugué sans problèmes particuliers.

Bob croit Jean parti

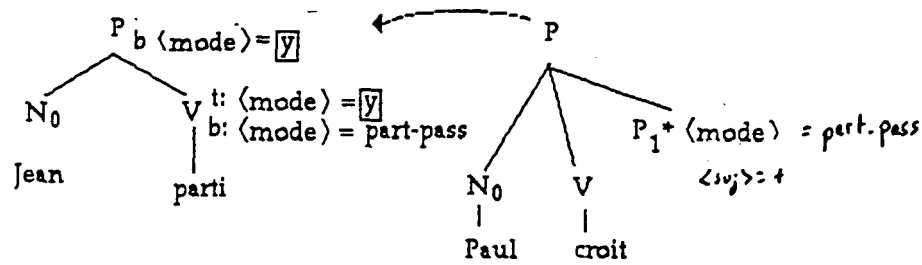


Figure 28

Les mêmes adjonctions s'effectuent donc quel que soit le mode de la complétive. Voilà aussi pourquoi les infinitives doivent être des phrases : si elles étaient des groupes verbaux, la principale ne pourrait plus être un arbre auxiliaire car elle n'aurait plus de noeud feuille de même catégorie que son noeud racine P; les infinitives devraient donc être substituées et comme telles seraient des îlots pour l'extraction, ce qui n'est pas conforme aux faits observés.

La plupart de ces adjonctions ont pour résultat d'éloigner l'élément questionné du verbe dont il est un argument et l'on aboutit aux cas de dépendances non bornées. Le caractère original de cette représentation est que cette relation entre l'élément préposé et son cadre argumentatif n'a pas à être calculée au cours de la dérivation: c'est une relation définie localement dès le départ et ce lien est maintenu au cours des dérivations ultérieures. L'arbre de dérivation associé à une phrase telle que: *Quelle fille Paul pense-t-il que Bob sait que ... (..) Jean aime ?* indique bien les différents cadres argumentatifs:

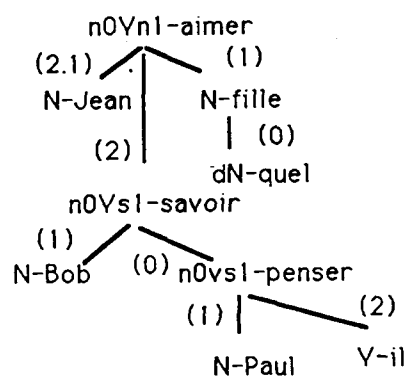


Figure 29

Notre traitement repose donc sur le fait que ce sont les mêmes arbres qui sont utilisés pour les principales, qu'il y ait extraction hors de la complétive ou non. Les principales peuvent donc "ignorer" l'extraction, modulo les contraintes d'inversion des arbres questions (voir infra).

c) autres phrases à complétives

Le mécanisme est le même avec les phrases à verbe support ou avec les phrases copulatives à complétive:

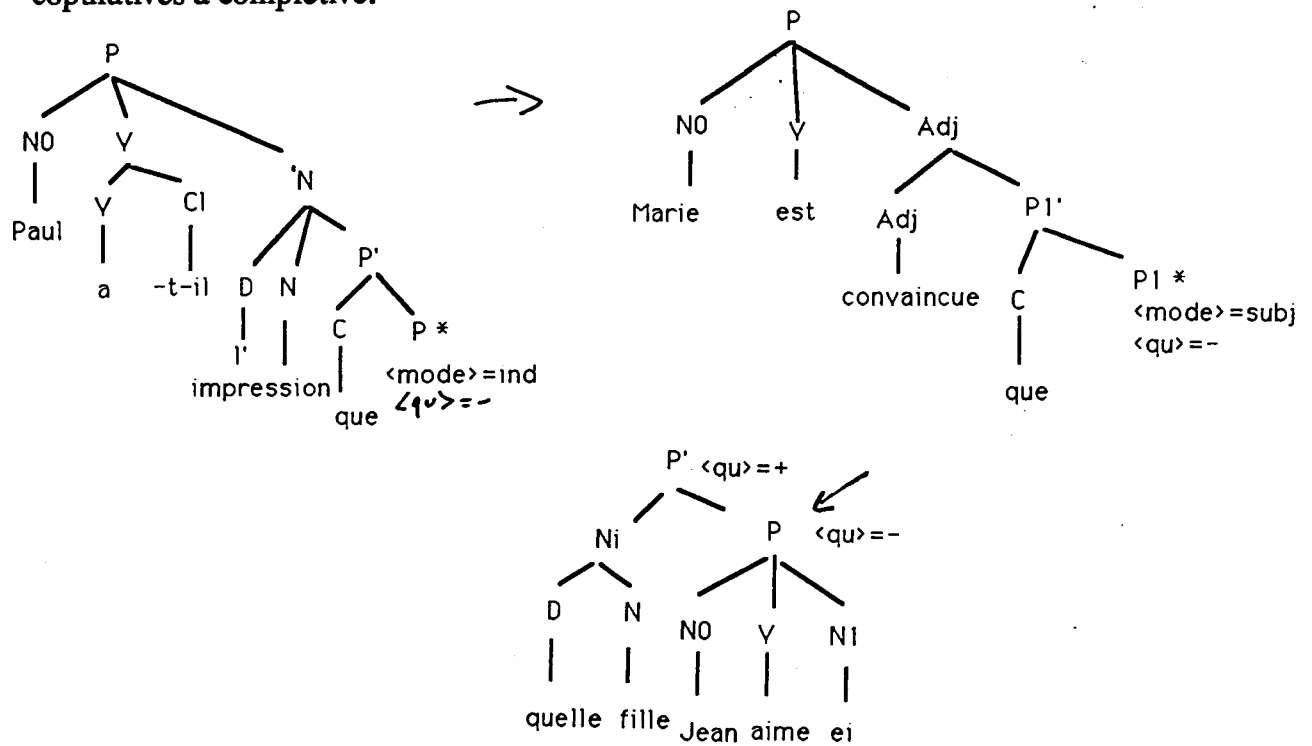


Figure 30

Toutes les combinaisons étant permises, on a bien par exemple:

Quelle fille Paul pense-t-il que Bob a l'impression que (...) Marie est convaincue que Jean aime ?

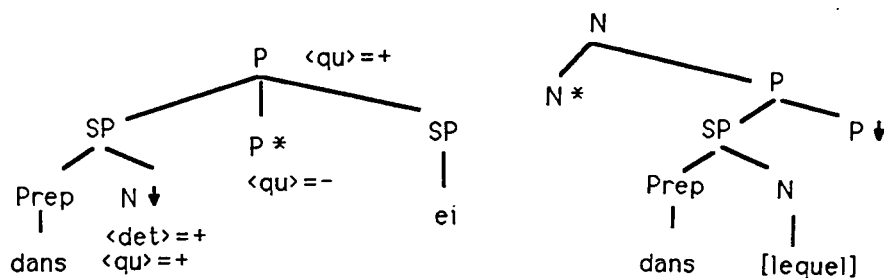
La fille que Paul a l'impression que Bob sait que (...) Jean aime ...

C'est cette fille que Paul pense que Bob a l'impression que (...) Jean aime.

C'est cette fille de qui Paul pense que Bob a l'impression que (...) Marie est convaincue que Jean aime.

d) compléments et ajouts

Nous traitons de la même façon les dépendances non bornées en cas d'extraction d'ajouts. Considérons les arbres élémentaires de la famille correspondant à *dans* (Cf chapitre 3), par exemple les arbres de la question, de la relative et de la clivée:



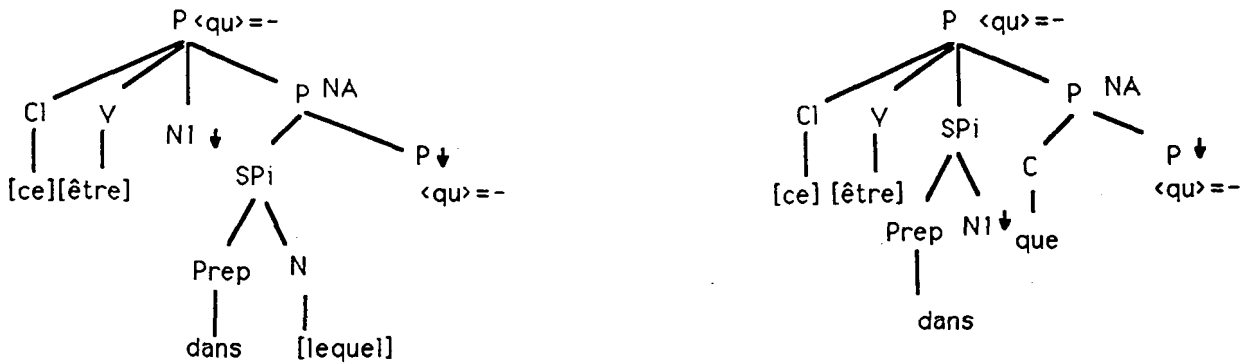


Figure 31

Nous avons donc bien:

Dans quelle maison Paul pense-t-il que Bob a l'impression que (...) Marie travaille ?
La maison dans laquelle Paul a l'impression que Bob sait que (...) Marie travaille ...
C'est dans cette maison que Paul pense que Bob a l'impression que (...) Marie travaille.

C'est cette maison dans laquelle Paul pense que Bob a l'impression que (...) Marie travaille.

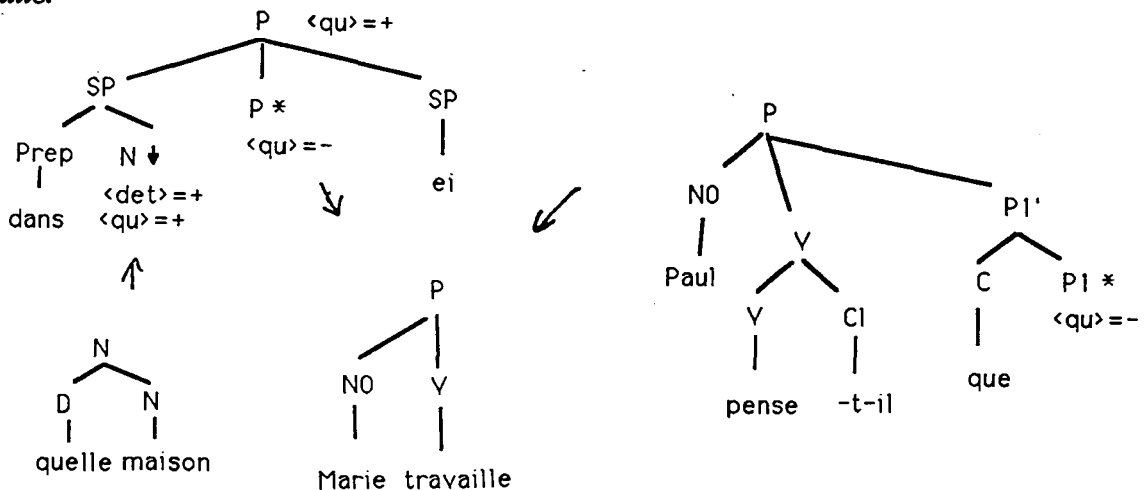


Figure 32

e) extraction des sujets

Nous ne traitons pas des dépendances non bornées portant sur des sujets, car elles semblent limitées en français à un très petit nombre de verbes (D. Godard 1980b). Les phrases sont souvent d'acceptabilité douteuse¹¹:

¹¹ Les acceptabilités sur ces phrases peuvent être assez variables (Cf supra p. 157). La relative est parfois mieux acceptée : ? *L'homme que vous savez qui aime Marie.*

* *C'est Paul que nous savons qui aime Marie*
 ?? *Qui est-ce que vous savez qui aime Marie ?*

Si l'on voulait intégrer ce type de construction dans notre grammaire, il faudrait prévoir une alternance *que/qui* au noeud pied des phrases à complétives et des traits (partagés avec la complétive) portant l'information que l'extraction est ou non celle du sujet. Cela semble une complication disproportionnée eu égard au faible nombre de verbes et de phrases concernés et à leur acceptabilité limitée¹².

Nous préférons donc interdire l'extraction, en ne prévoyant pas de noeud P intermédiaire dans les structures à sujet extrait¹³:

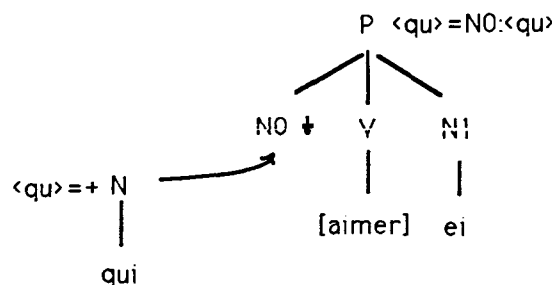


Figure 33

Nous avons donc rendu compte, grâce à l'opération d'adjonction, de nombreux cas de dépendances non bornées d'une façon purement locale.

Comme tel, cependant, ce mécanisme est trop général et des contraintes supplémentaires doivent lui être imposées. Pour bloquer l'extraction, on utilisera la substitution comme opération de combinaisons entre arbres élémentaires, à la place de l'adjonction, ce qui fait qu'on n'utilise pas toujours l'adjonction quand un arbre élémentaire de racine phrastique a une feuille de catégorie Phrase.

2.2.2. Substitution et adjonction

Dans certains cas, c'est le formalisme qui impose de ne pas utiliser l'adjonction pour un argument phrastique; dans d'autres cas, c'est un choix purement linguistique.

¹² Ce trait <extrac-sujet> = +/- (ou <comp> = *qui/que*) serait un trait local car il règle la combinaison de la complétive avec la phrase immédiatement au dessus d'elle (l'avant-dernière en cas de complétives en cascades) qui est celle dont le complémenteur change. Nous n'aurions donc pas les complications du trait SLASH en GSPG qui doit parcourir toute la cascade des compléments.

¹³ Nous pourrions aussi avoir un noeud P intermédiaire mais y limiter l'adjonction aux ajouts adverbiaux ou incises (<princ> = +).

- C'est Jean qui pense que Paul aime Marie*
 * *C'est Marie que c'est Jean qui pense que Paul aime*
 * *Marie, que c'est Jean qui pense que Paul aime*
 * *Qui est-ce Jean qui pense que Paul aime ?*

On les représente donc par des arbres initiaux, où sera substituée la complétive.

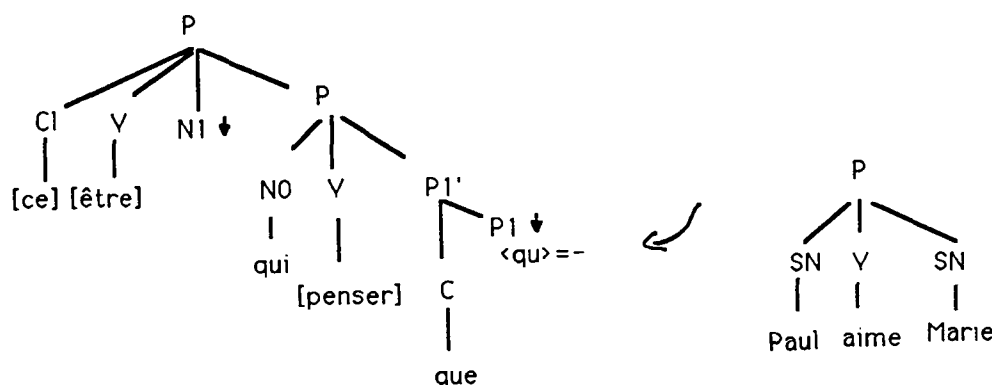


Figure 35

Les restrictions sont les mêmes pour les familles de phrases copulatives ou de phrases à verbe support pour les adjectifs et les noms prenant une complétive.

b) verbes à deux complétives

La complétive correspondant au noeud pied de l'arbre auxiliaire, qui est unique par définition¹⁴, un problème se pose pour les verbes à plus d'une complétive tels que *prouver*, *revenir à* ou *déduire*:

- Aller chez Marie revient à perdre une journée*
L'on ne peut déduire que Pierre est épris de Marie de ce qu'il a acheté ce livre.

C'est pourquoi nous avons introduit initialement l'opération de substitution pour pouvoir insérer dans l'arbre de la 'principale' cette deuxième complétive (A. Abeillé 1988b). Le choix entre substitution et adjonction pour introduire une complétive est motivé linguistiquement. En effet, l'opération d'adjonction permet la représentation de l'extraction à distance tandis que la substitution bloque cette extraction. Or, il y a bien dissymétrie entre les deux complétives en ce qui concerne l'extraction dans les deux phrases ci-dessus:

¹⁴ Si l'on définit des arbres auxiliaires à deux noeuds pieds (de même catégorie), l'on augmente le pouvoir génératif de la grammaire. L'on peut ainsi définir une hiérarchie de grammaires d'arbres adjoints avec des arbres auxiliaires respectivement à un pied, deux pieds, trois pieds ... La hiérarchie des langages générés va être respectivement $(a^n b^n c^n d^n)$, $(a^n b^n c^n d^n e^n f^n)$, $(a^n b^n c^n d^n e^n f^n g^n h^n)$... (K. Vijay-Shanker et al 1987)

Combien de temps est-ce qu'aller chez Marie revient à perdre ?

** Où est-ce qu'aller revient à perdre une journée ?*

? De qui peut-on déduire que Pierre est épris de ce qu'il a acheté ce livre ?

** Que peut-on déduire que Pierre est épris de Marie de ce qu'il a acheté ?*

Et aucun principe général n'interdit l'extraction hors des complétives en *de ce que* :

Où Jean rêve-t-il de ce que Marie aille ?

Le noeud pied de l'arbre auxiliaire phrastique correspondra donc à la deuxième complétive de *revenir à* et à la première de *déduire*, permettant ainsi l'extraction. Les autres complétives seront substituées, ce qui bloque l'extraction, la substitution combinant forcément un noeud feuille avec un noeud racine et empêchant ainsi l'insertion d'une structure 'à l'intérieur' d'une autre:

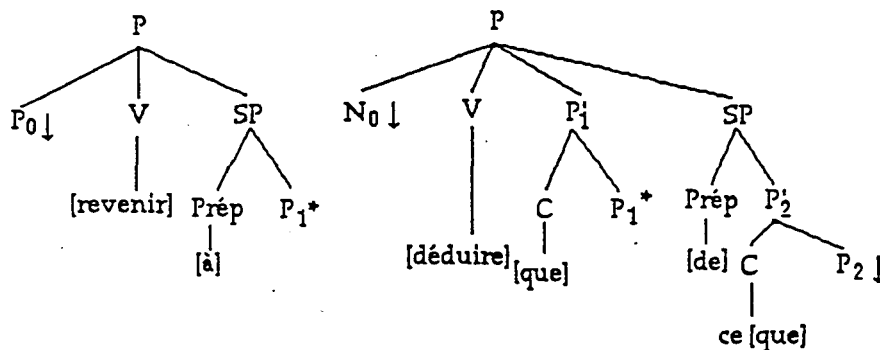


Figure 36

c) verbes à sujet phrastique

Cet usage de la substitution, nécessaire formellement dans le seul cas de verbes à plus d'une complétive, est étendu pour raison linguistique à tous les cas de complétives sujets (non extraposées) puisque l'extraction y est également interdite:

Aller en Espagne ennuie Jean

** Où aller ennuie-t-il Jean ?*

C'est pourquoi les verbes à complétive objet ont pour arbres passifs dans leurs familles des arbres initiaux et non des arbres auxiliaires. On aura donc bien le contraste suivant:

Jean a montré que la vie existe sur Mars
Où Jean a-t-il montré que la vie existe ?
Que la vie existe sur Mars a été montré par Jean
** Où que la vie existe a-t-il été montré par Jean ?*

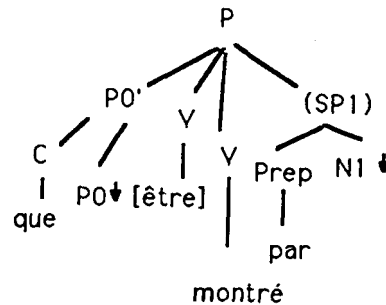


Figure 37

Ce n'est donc pas comme en GPSG un mode particulier de propagation d'un trait SLASH (et reposant crucialement sur la présence d'un constituant SV) qui rend compte du caractère d'îlot des complétives sujet par opposition aux complétives objets.

Il en va différemment quand le sujet phrastique est extraposé:

Cela ennue Jean d'aller en Espagne
Où cela ennue-t-il Jean d'aller ?

L'extraction étant autorisée, le (ou les) arbre(s) correspondant sont bien des arbres auxiliaires, non des arbres initiaux:

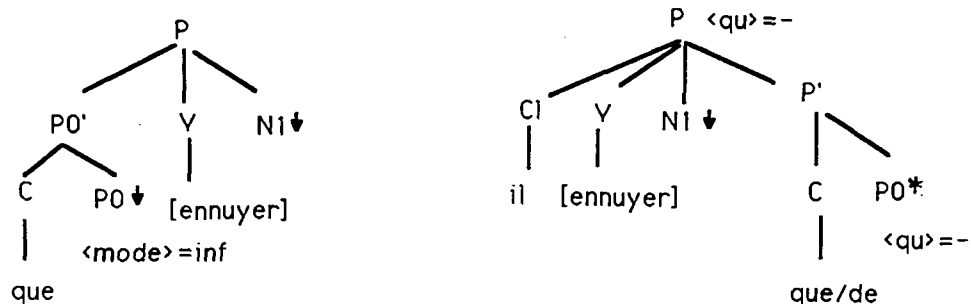
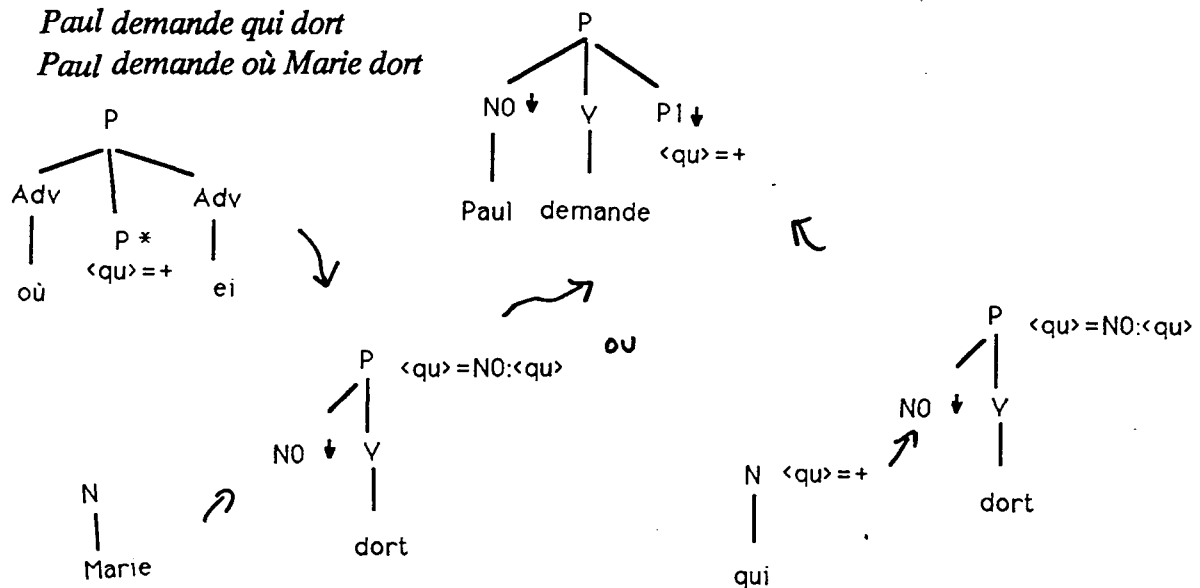


Figure 38

d) verbes à interrogatives indirectes

Les verbes à interrogative indirecte sont représentés comme des arbres phrastiques initiaux. L'élément *qu-* fait partie de l'arbre élémentaire de l'interrogative substituée s'il

correspond à un argument, ou s'y adjoint de façon autonome s'il correspond à un adverbial:



Il n'y a pas de contraintes lexicales de mode, les questions indirectes sont toujours à l'infinitif ou à l'indicatif.

Nous ne traitons pas des réductions d'interrogatives indirectes à SN, SP ou Adv (Cf *Il est parti mais je ne sais pas où*) ni des "questions cachées" (*N₀ demander l'heure* = *N₀ demander quelle heure il est*).

Expliquons maintenant pourquoi nous traitons les interrogatives indirectes comme des complétives substituées. En effet, nous avons vu supra que l'extraction hors d'une interrogative indirecte est exclue en vertu des principes de formation des interrogatives elles-mêmes, non en vertu de contraintes particulières dues au verbe principal.

L'on exclut donc :

* *Qui demandes-tu qui aime ?*

Il importe donc peu ici que *demande* soit un arbre auxiliaire ou un arbre initial, puisque c'est l'arbre *qui qui aime* qui est impossible.

De même, l'on exclut toutes les phrases où un verbe à complétive domine le verbe à interrogative, parce qu'elles supposeraient le même arbre à deux éléments interrogatifs antéposés que ci-dessus:

* *Qui penses-tu que Marie demande qui aime ?*

En revanche, la structure du verbe principal joue un rôle dans les cas où l'interrogative indirecte est elle-même une phrase à complétive:

- (1) *Jean demande qui pense que Paul aime Marie*
 (2) * *Qui Jean demande-t-il qui pense que Paul aime ?*

L'on ne peut exclure l'extraction sur les mêmes principes que ci-dessus puisque les deux éléments interrogatifs appartiennent à deux structures différentes: *qui pense que P* et *qui Paul aime* sont tous deux des arbres bien formés. Ils ne peuvent se combiner entre eux directement (contraintes sur les traits <qu> de leurs noeuds P).

En revanche, si l'on adjoint *Jean demande-t-il* sur *qui pense que P*, l'on obtient un arbre auxiliaire bien formé et qui extérieurement ne présente plus de trait <qu> incompatibles avec ceux du noeud intérieur P de *qui Paul aime*.

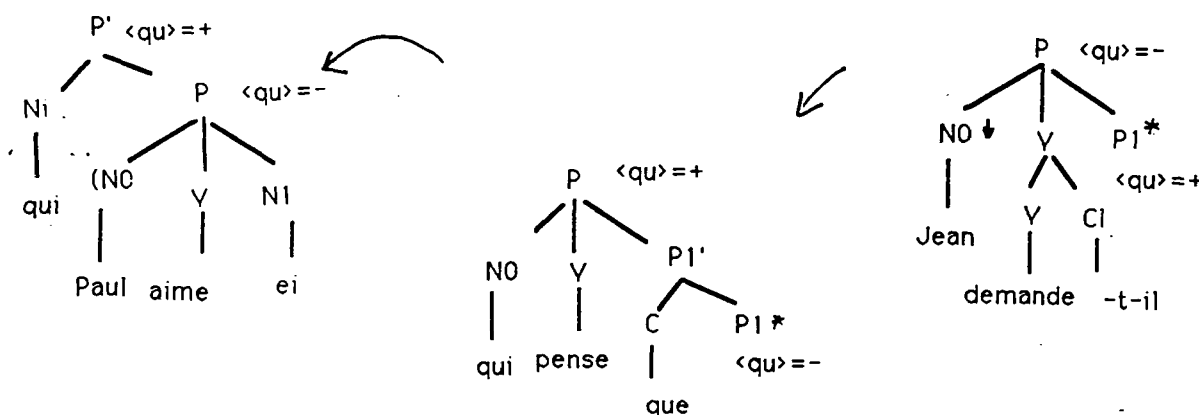


Figure 40

La solution, pour ne pas générer la phrase (2) est donc de considérer que N_0 demander P est un arbre initial. L'arbre dérivé *Jean demande-t-il qui pense que P* sera donc de type initial et ne pourra s'adjoindre entre *qui* et *Paul aime*.

En fait, l'interdiction de l'extraction hors d'une interrogative indirecte est moins générale, en particulier si l'interrogative commence par un ajout:

- (3) *Paul se demande comment résoudre ce problème*
 (3a) ? *Paul se demande comment résoudre quel problème ?*
 (3b) ? *Quel problème Paul se demande-t-il comment résoudre ?*

Il semble que l'antéposition du complément en *quel* ne gêne pas vraiment l'acceptabilité. Comme le note notamment D. Godard 1980b, le contraste est net entre les phrases (3b) et (4) :

- (4) * *Comment Paul se demande-t-il quel problème résoudre ?*

L'on pourrait rendre la phrase (1a) possible en définissant un second type d'arbre pour *demander* qui soit un arbre auxiliaire (dont l'emploi sera limité à certaines interrogatives à l'infinitif). La phrase (2) serait toujours impossible car il n'y a pas de noeud P

<qu> = + où adjoindre l'arbre de *demander*.

e) contraintes des verbes non "ponts"

On peut représenter de la même façon des verbes qui ont une complétive mais "bloquent" l'extraction hors de celle-ci, c'est-à-dire par des arbres initiaux, forçant ainsi la complétive à être substituée. Cela suppose que toutes les extractions (clivées, questions, relatives ..) soient impossibles de la même façon.

Il est intéressant de noter que la présence d'un tel verbe à un quelconque niveau d'enchâssement bloque l'extraction hors de la dernière complétive:

Paul s'est exclamé qu'il aimait Marie

- (1) * *Qui s'est-il exclamé qu'il aimait ?*
- (2) * *Qui pense-t-on qu'il s'est exclamé qu'il aimait ?*
- (3) * *Qui s'est-il exclamé qu'on a tort de penser qu'il aime ?*

C'est ce que fait précisément la substitution. Des verbes à complétives peuvent s'adjoindre à l'intérieur de l'interrogative (*qui on pense qu'il aime*), c'est le tout qui doit être substitué après *s'exclamer*. *Qui*, appartenant à la complétive, ne peut donc être antéposé et la seule question possible est donc¹⁵:

Il s'est exclamé qu'il aimait qui ?

De même, si l'on adjoint des complétives sur l'arbre de *s'exclamer* (*pense-t-on qu'il s'exclamait*), le résultat est un arbre initial, non un arbre auxiliaire et l'adjonction n'est pas non plus possible à l'intérieur de l'interrogative. Celle-ci ne peut être que substituée et l'élément *qu-* ne peut se retrouver en tête de phrase. On ne peut donc générer la phrase (2).

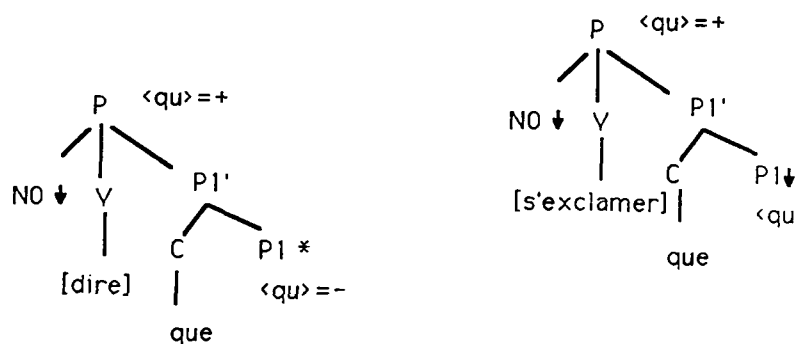


Figure 41

Nous ne traitons pas en revanche de la gradation d'acceptabilité entre verbes ponts et non ponts (D. Godard 1980a). Il semble y avoir un continuum entre *dire*, *taire*, *hurler*, *s'écrier*, *s'exclamer* ... Le principe même de la formalisation oblige à une distinction

¹⁵ dans la mesure où l'on refuse *Il s'est exclamé qui il aimait*, c'est-à-dire une interrogative indirecte après *s'exclamer*.

tranchée entre les deux extrêmes.

Un autre problème vient du fait que le blocage de l'extraction peut venir indirectement de l'adjonction d'un adverbe sur le verbe de la principale:

Jean lui disait (E + à mots couverts) qu'il irait chez Marie.

Où Jean lui disait-il (E + ?? à mots couverts) qu'il irait ?

Nous n'avons pas actuellement de moyens de représenter cette induction de contrainte.

Des variations existent également avec les adjectifs en phrase copulative :

Quelle pièce étiez-vous (étonné + ? sidéré + ?? abasourdi) que les élèves n'aient pas encore lue ? (D. Godard 1980a)

La phrase copulative correspondant à un adjectif non "pont" est un arbre initial et non un arbre auxiliaire, avec complétive substituée.

On note de semblables différences avec les phrases à verbe support ou les expressions figées, et nous les représentons de la même façon:

Paul (a + détient) la preuve que Jean était allé à Bilbao ce jour là.

Où Paul (a-t-il + ?détient-il) la preuve que Jean était allé ce jour là ?

Paul a fait la preuve que Jean était allé à Bilbao ce jour là.

** Où Paul a-t-il fait la preuve que Jean était allé ce jour là ?*

f) complétives et conjonctives

Les circonstancielles sont représentées de façon similaire aux complétives des verbes non ponts. Leur complémenteur éventuel fait partie de l'arbre auxiliaire de la préposition (ou conjonction) qui les sélectionne. Elles sont substituées dans l'arbre de la préposition (ou de la conjonction) ce qui garantit leur statut d'îlot vis à vis de l'extraction:

Jean rit pendant que Marie lit le journal

** Quel journal Jean rit-il pendant que Marie lit ?*

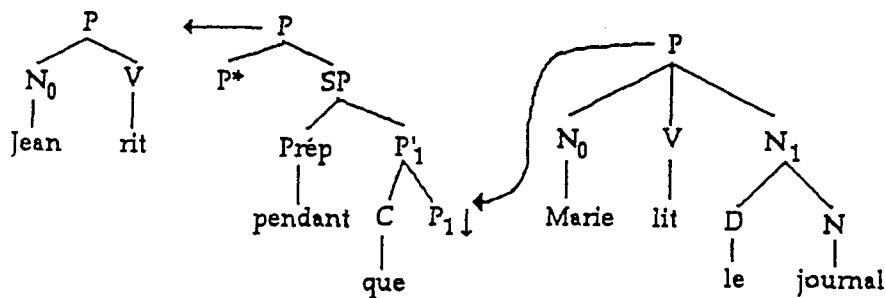


Figure 42

Avec certaines conjonctions, l'extraction n'est pas possible non plus hors de la principale:

Jean aime Marie, (pourtant + par conséquent) il rougit.

* *Qui Jean aime-t-il, (pourtant + par conséquent) il rougit ?* (traduit de J. McCawley 1989)

Elles sont représentées comme les conjonctions de coordination.

2.2.3 Un exemple détaillé

Nous allons maintenant détailler la construction d'une phrase bien formée à partir d'un arbre question. Il nous permettra d'expliquer les contraintes d'inversion et de donner une nouvelle illustration de la propagation dynamique de contrainte que permet la bipartition des traits top et bottom en TAG. Si l'on prend par exemple l'arbre *Qui Jean aime*:

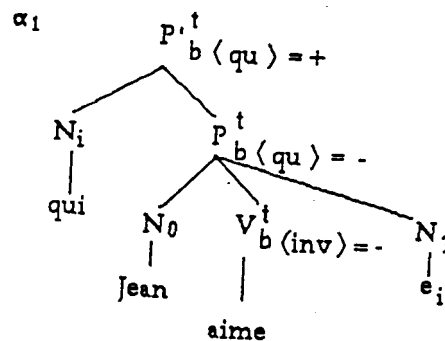


Figure 43

Il peut être substitué dans une phrase à complétive. Une phrase à complétive peut également s'adjoindre au noeud P dominé par P'. Dans le premier cas, le résultat sera une construction à interrogative indirecte, dans le second une question directe. Les deux noeuds phrastiques P' et P portent des traits différents, qui différencient les structures qui peuvent s'y adjoindre. A la racine, des structures verbales du type *demander* (mais non *craindre* ou *penser*) peuvent se combiner. En P, toute phrase à complétive peut s'adjoindre à condition qu'elle ne soit pas elle-même une question. Supposons donc que l'on dispose dans la grammaire des arbres suivants:

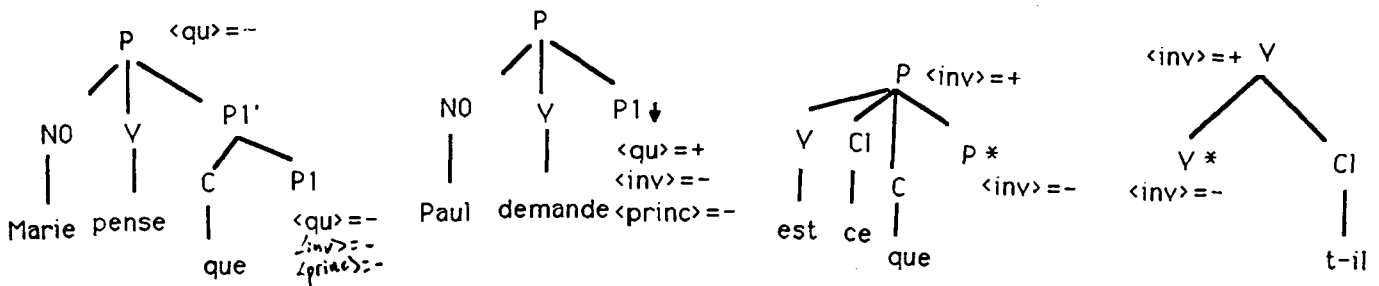


Figure 44

Il y a plusieurs façons d'engendrer une phrase bien formée :

- on peut adjoindre en V le clitique inversé *il* (ici *-t-il*), obtenant ainsi l'inversion complexe : *Qui Jean aime-t-il ?*
- on peut adjoindre en P l'arbre auxiliaire correspondant à *est-ce que*, obtenant ainsi : *Qui est-ce que Jean aime ?*
- on peut adjoindre en P l'arbre auxiliaire *Marie pense-t-elle* (dérivé lui-même de *Marie pense* à quoi s'est adjoint *elle*), obtenant ainsi : *Qui Marie pense-t-elle que Jean aime ?*
- on peut adjoindre en P un (ou plusieurs) arbre(s) auxiliaire(s) phrastique(s) non inversé(s) et au noeud racine du dernier un arbre avec inversion, générant par exemple : *Qui pensez-vous que Paul regrette que Jean aime ?*
- on peut substituer l'arbre initial dans un arbre à interrogative indirecte ($Qu = +$), générant par exemple : *Paul demande qui Jean aime.*
- on peut substituer l'arbre initial dans un arbre à interrogative indirecte et adjoindre en P un autre arbre à complétive : *Paul demande qui Marie croit que Jean aime.*

De façon plus générale, les contraintes d'inversion sont représentées comme suit :

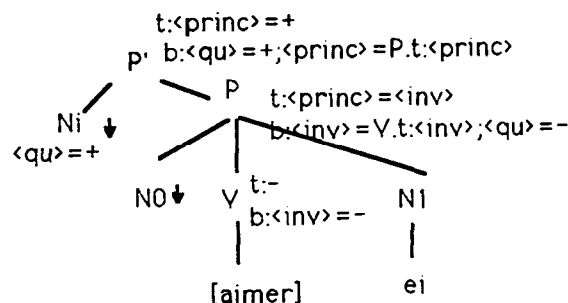


Figure 45

On utilise un trait <princ> pour distinguer entre phrases enchâssées et principales et un trait <inv> pour noter quand il y a inversion. Dans le cas des arbres questions, il y a corrélation entre ces deux traits : une question directe doit être inversée, une question indirecte ne peut pas l'être¹⁶. Si c'est une question directe, la contrainte d'inversion crée une contrainte d'adjonction (d'un clitique sujet ou de *est-ce-que*):

- * *Qui Jean aime ?*
- Qui Jean aime-t-il ?*
- Qui est-ce que Jean aime ?*
- * *Qui est-ce que Jean aime-t-il ?*

L'inversion peut aussi venir de l'adjonction d'une principale. La contrainte d'inversion peut donc se transmettre de phrase en phrase. S'agissant d'un trait top, elle concernera toujours la phrase la plus élevée:

- Qui Marie pense-t-elle que Jean aime ?*
- * *Qui Marie pense que Jean aime-t-il ?*
- Qui Paul croit-il que Marie pense que Jean aime ?*
- * *Qui Paul croit que Marie pense-t-elle que Jean aime ?*

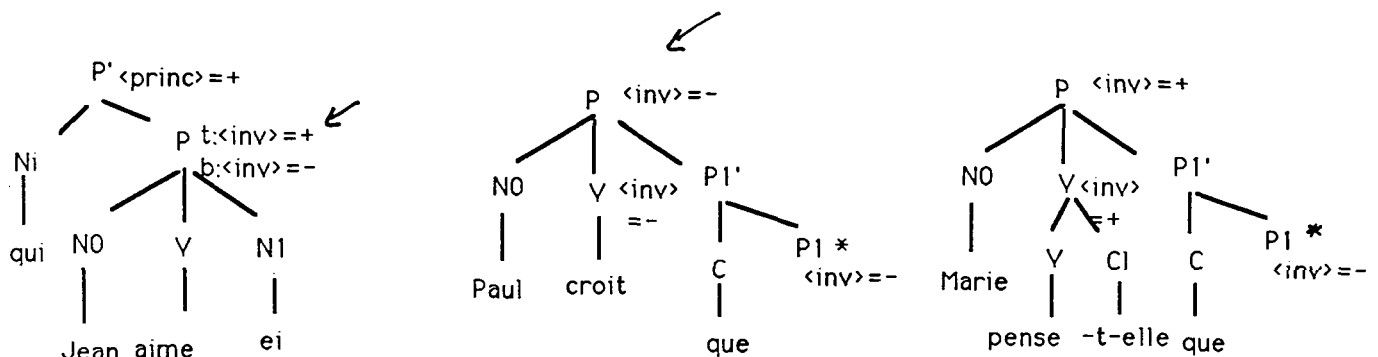


Figure 46

Lorsqu'il s'agit d'une question indirecte, ces adjonctions sont impossibles.

¹⁶ Nous assimilons inversion complexe et question en *est-ce-que*.

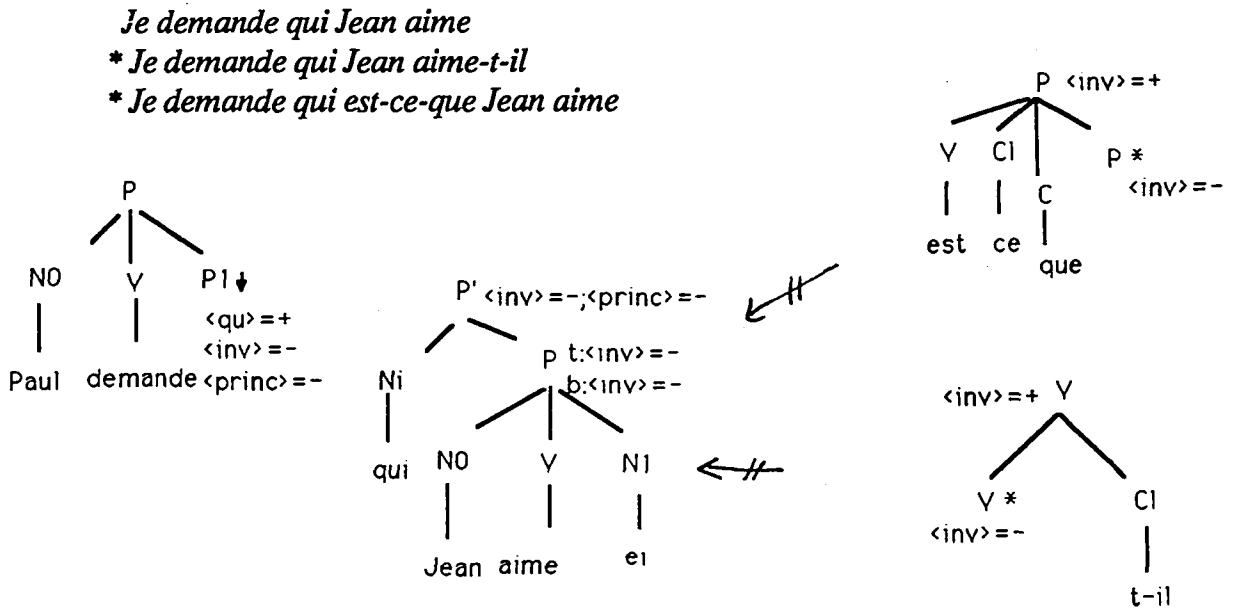


Figure 47

Il y a donc là un exemple de propagation dynamique de contrainte: l'arbre de départ n'est pas auto-suffisant, car si l'on unifie top et bottom à chaque noeud, il en résulte à au moins un noeud des traits à valeur incompatible. Il doit donc recevoir au moins une adjonction, mais elle peut se produire à des noeuds variés.

Cette incompatibilité entre enchâssement et inversion est propre aux questions, il en va différemment de l'inversion en relative ou en clivée (Cf chapitre 6).

Chapitre VI

Recherches en cours

Nous présentons dans ce chapitre quelques extensions en cours de la grammaire TAG du français. Certaines nécessitent des traits supplémentaires, d'autres nécessitent des structures nouvelles ("TAG multiples" ou mise en facteur de sous-arbres). Nous examinerons successivement certaines corestrictions entre ajouts et arguments ainsi que des phénomènes de contrôle (dans les infinitives), puis des phénomènes de variations d'ordre des mots et enfin quelques cas de coordination.

1. Les "ajouts-arguments"

Nous empruntons ce terme à J. Grimshaw 1988 qui l'utilise pour les déterminants possessifs et les compléments d'agent en anglais. Nous traitons ici des clitiques et déterminants. Les adjectifs équivalents à des compléments en *de N*, dont des exemples ont été donnés à la fin du chapitre 4, seraient d'autres cas d'ajouts-arguments.

Les liens entre ajouts et arguments sont multiples. Certains adverbes ont une portée préférentielle sur le sujet (Cf chapitre 2) et lui imposent des restrictions (*intelligemment* impose la présence d'un sujet ou complément d'agent humain). Les adverbes "symétriques" sont un autre exemple, où la contrainte porte sur le nombre (*conjointement*, par exemple, impose qu'un argument du verbe principal soit au pluriel). Nous n'envisageons ici que des cas de distribution complémentaire entre ajouts et arguments, c'est-à-dire où la présence de l'ajout empêche la réalisation de l'argument et vice-versa.

1.1 Les clitiques

Les structures des clitiques ont été présentées aux chapitres 3 et 5. Leurs corestrictions mutuelles aussi, qui ne posent pas de problèmes particuliers (elles sont représentées par un système de traits), de même que certaines restrictions venant du verbe principal (e.g. le trait $\langle \text{acc} \rangle = +/-$ pour les verbes à complétive). Nous traitons ici des corestrictions entre clitiques et arguments du verbe. Nous ne traitons pas des clitiques non arguments, qui présentent aussi des phénomènes de corestrictions avec des adverbiaux (*y* ou *en*) ou des compléments de nom (*en*). Nous ne traitons pas non plus des cas de "montée" de clitiques (dans les factitives).

1.1.1. Clitiques et optionalité des arguments

Nous représentons les clitiques comme des ajouts au verbe, et non directement comme des arguments substitués (Cf chapitre 3). Nous ne définissons donc pas de structures phrastiques supplémentaires (il en faudrait autant que de combinaisons de clitiques arguments possibles). Il faut donc exprimer des corestrictions entre l'adjonction des clitiques et la réalisation des arguments correspondants.

On utilise les trait nominatif (noté <nom>), accusatif (noté <acc>), datif (noté <dat>), locatif (noté <loc>), génitif (noté <gen>) (Cf chapitre 3). On utilise aussi l'opération d'attachement direct, qui attache la chaîne vide aux noeuds des arguments concernés (cette opération est implémentée en TAG, c'est la même que celle utilisée pour attacher les têtes lexicales à leurs arbres élémentaires). L'on note directement dans les arbres les rapports d'implication, par exemple: V.t:<acc>=+ => N₁=e ou V.t:<nom> => N₀=e. Au départ, c'est-à-dire avant adjonction d'un clitique, tous les traits casuels ont la valeur - au noeud V (bottom).

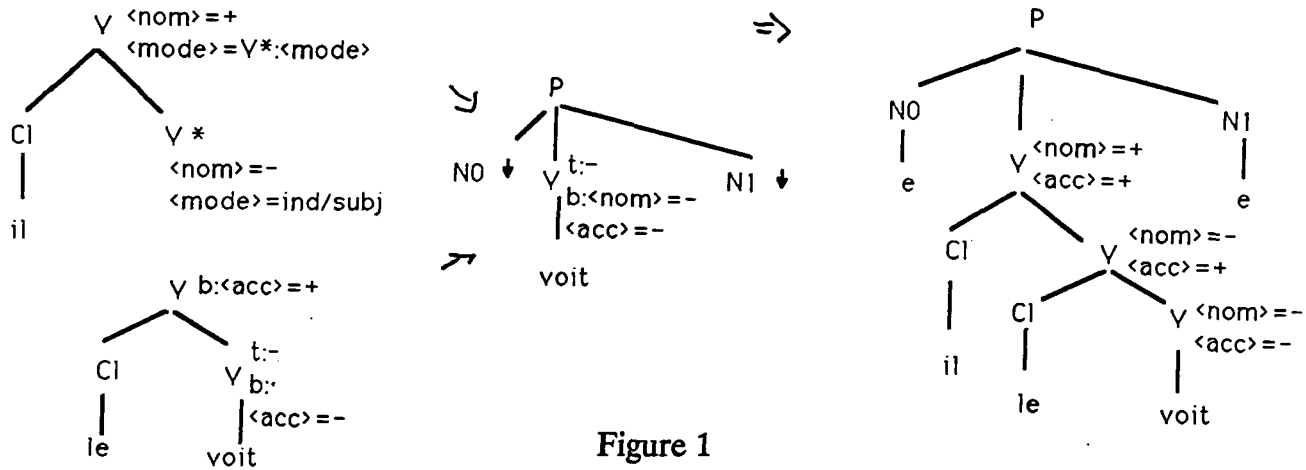


Figure 1

Les phénomènes d'alternance entre complément à *N* et clitique datif, ou entre complément *de N* et clitique *en* se règlent de façon similaire. C'est tout le sous-arbre SP qui doit être mis à zéro¹ : V.t:<dat>=+ => SP(dat)=e ou V.t:<gen>=+ => SP(gen)=e

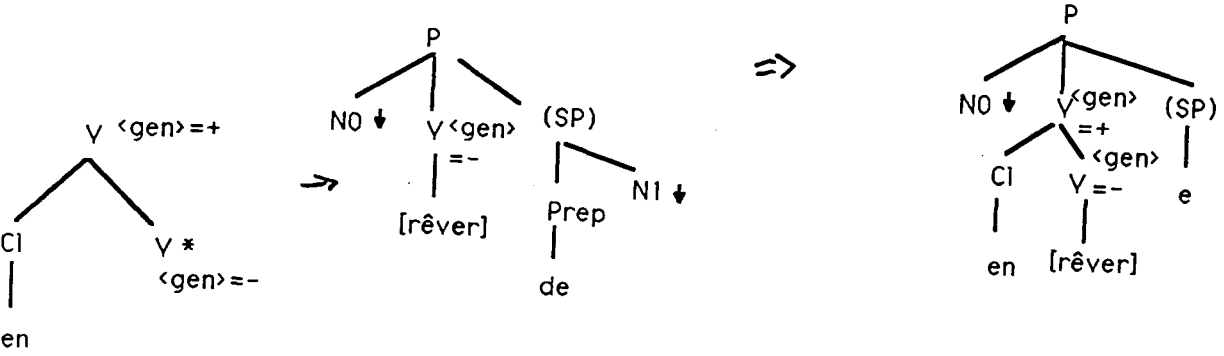


Figure 2

Il y a donc deux *en*, l'un accusatif (qui contrôle la réalisation du complément direct partitif ou pluriel), l'autre génitif, et deux *y* l'un datif, l'autre locatif. De même pour les réfléchis *me, te, se, nous, vous* qui peuvent être accusatifs ou datifs (Cf chapitre 3).

¹ Les verbes dont le complément *en* à ne peut donner matière à un clitique *lui* ou *leur* (N₀ penser à N₁) ont le trait V.t:<datif-h>=- dans le lexique.

On trouve les mêmes alternances pour les arguments libres des phrases figées, et on les traite de la même façon que pour les phrases libres :

$$N_0 \text{ promettre la lune (à } N_2) = N_0 \text{ Clit}_{\text{dat}} \text{ promettre la lune}$$

Notons que les sujets pourraient être traités différemment: comme des arguments N substitués à condition d'interdire leur modification par des relatives, leur coordination avec un autre N et l'insertion d'un adverbe entre eux et le verbe. Cette dernière condition étant plus difficile à exprimer si le clitique fait partie d'un syntagme séparé du verbe, nous avons préféré les traiter comme les autres clitiques (pour un traitement unifié des clitiques voir aussi M. Gross 1967, R. Kayne 1975).

Notons aussi que l'opération d'attachement direct de la chaîne vide n'est pas forcément nécessaire pour les cas des arguments optionnels, et l'on peut forcer la non réalisation de ces arguments par simple corestriction de traits (si on les marque de la même façon que les clitiques). Par exemple, $V.t:\langle \text{acc} \rangle = N_1:\langle \text{acc} \rangle$ et $N_1:\langle \text{acc} \rangle = -$.

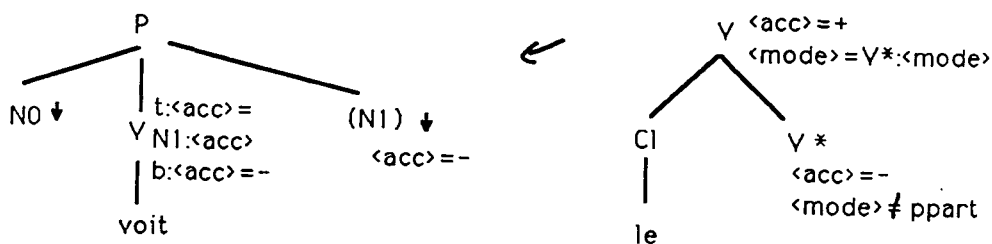


Figure 3

L'adjonction d'un clitique comme *le* ou *en* entraîne $\langle \text{acc} \rangle = +$ en V, ce qui est incompatible avec l'équation ci-dessus. La seule manière de rendre l'unification possible est que N_1 ne soit pas réalisé et donc le trait $\langle \text{acc} \rangle$ ne sera pas spécifié à ce noeud. C'est ce que nous utilisons pour l'instant pour les compléments prépositionnels optionnels, le remplacement d'un sous-arbre par une chaîne vide n'étant pas implémenté.

Il a été souvent noté (R. Kayne 1975, J. Grimshaw 1982), que la combinaison clitique (accusatif) + verbe peut avoir certaines propriétés de verbe intransitif, par exemple pour la factitive à sujet inversé ou pour l'inversion simple (normalement réservés aux intransitifs):

Je ferai partir les enfants

* *Je ferai laver les poupées les enfants*

Je ferai se laver les enfants

"Le séduisait surtout le fait de s'enfermer dans une cellule, de travailler "en tête à tête avec la pellicule" (Le Monde, janvier 1990)

?* *Séduisait Jean le fait de s'enfermer dans une cellule...*

Il serait relativement aisé de rendre compte de ces faits dans notre système, puisque les traits portés par le verbe sont différents selon que son complément est un clitique ou un

N (ou un SP).

1.1.2. Cas particuliers

Dans un certain nombre de cas, la coprésence du clitique et de l'argument est possible voire obligatoire.

a) Clitiques sujets inversés

On les a vus au chapitre 5. Ils sont représentés par des structures spécifiques:

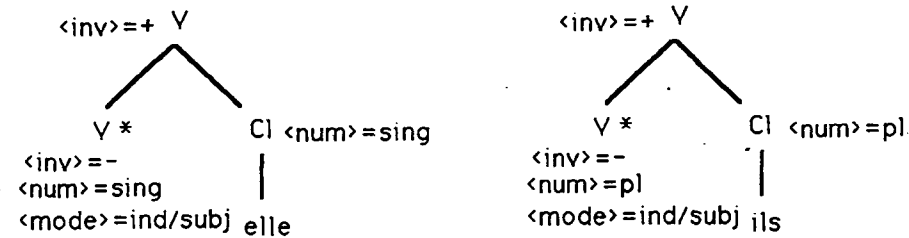
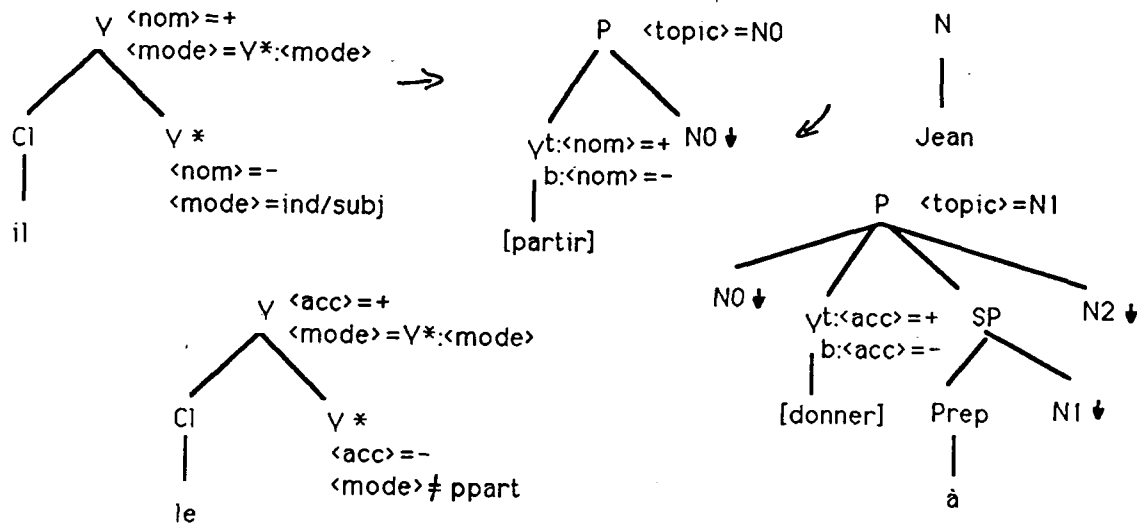


Figure 4

Ils ne portent pas de traits <nom> car ils sont compatibles avec un sujet nominal.

b) Constructions disloquées (et topicalisations)

La dislocation à droite donne naissance à des structures élémentaires supplémentaires, pour chaque verbe simple (ou figé) concerné. L'adjonction du clitique est obligatoire :



Il part, Jean

Jean le donne à Marie, le livre

Figure 5

Le clitique qui s'adjoint peut être inversé ou non:

Il est parti, Pierre ?
Est-il parti, Pierre ?

Pour la dislocation à gauche, on peut utiliser les mêmes structures que pour la question, en ajoutant des corestrictions de traits.

1.2. Les déterminants

Comme on l'a vu au chapitre 3, il y a sélection conjointe du déterminant par le verbe et par le nom.

Un cas particulier est celui des noms prédicatifs (ou des expressions à compléments de nom figé) où l'on remarque une distribution complémentaire entre complément et déterminant possessif:

(Son + L') agression (E + de Jean) contre Marie*
*Son agression (E + *de Jean) contre Marie*

L'on pourrait considérer un traitement similaire à celui des corestrictions entre clitiques et arguments (avec un trait <subj> et l'attachement direct de *e*). Un problème est celui de l'interprétation de *de N* qui peut toujours se faire comme possessif, si le nom prédicatif peut aussi avoir une acception concrète :

? Son portrait de la Grande Odalisque d'Ingres

Nous n'avons pas considéré les phénomènes d'alternance parfois cités entre déterminants possessifs et compléments en *par*, qui sont généralement d'acceptabilité douteuse :

*L'achat (*de + par) Jean de cette robe*
?? son achat de cette robe

Il conviendrait par ailleurs d'étendre les phénomènes d'alternance vus supra au système des noms de partie du corps pour lesquels le clitique datif peut alterner avec un complément de nom (Cf R. Kayne 1975 et J-P. Boons, A. Guillet, C. Leclère 1976). On le trouve aussi dans les phrases figées (Cf chapitre 4).

Jean casse du sucre sur le dos de Marie
Jean casse du sucre sur son dos
Jean lui casse du sucre sur le dos

2. Phénomènes de contrôle

Les phénomènes de contrôle sont variés et ont suscité une littérature abondante, tant en grammaire générative que dans les grammaires d'unification. Nous allons traiter essentiellement des sujets des infinitives. Nous mentionnerons une extension possible au sujet des noms prédicatifs.

Il convient de marquer le sujet implicite des infinitives essentiellement pour :

- contrôler les phénomènes d'accord (avec les infinitifs passés ou passifs, avec les clitiques réfléchis ou les déterminants possessifs figés):

*Jean veut être (parti + *partis + *partie) à midi*
*Marie veut être (*parti + *partis + partie) à midi*

*Jean veut (*m' + *t' + s') en aller*
*Je veux (m' + *t' + *s') en aller*

- contrôler les restrictions de sélection. Par exemple, les contraintes de sujet pluriel (N_0 se réunir) ou de sujet concret (N_0 casser):

*Jean (mange + *se réunit + *casse)*
*Jean veut (manger + *se réunir + *casser)*

- préparer l'analyse sémantique en rapprochant des phrases potentiellement synonymes:

Jean espère partir
 = *Jean espère qu'il partira*

Les grammaires d'unification (et toutes les théories qui considèrent les compléments à l'infinitif comme des SV) marquent aussi le sujet des infinitifs après verbes à montée ou modaux. Pour GPSG, LFG ou HPSG, la différence entre des verbes comme *sembler*, *commencer à*, *finir de...* et des verbes comme *vouloir*, *espérer...* est essentiellement une différence sémantique (ou de structure fonctionnelle) et non une différence syntaxique (ou de structure en constituant).

La différence entre verbes "à montée" et verbes "equi" est pour nous une différence structurale. Les premiers sont représentés comme des ajouts au verbe, les seconds comme des têtes d'arbres phrastiques avec une structure argumentale complète (sujet et complément phrastique). Pour les premiers, il n'est pas besoin de calculer le sujet de l'infinitif, qui appartient au même arbre élémentaire et est contraint par lui indépendamment de l'adjonction du verbe "à montée":

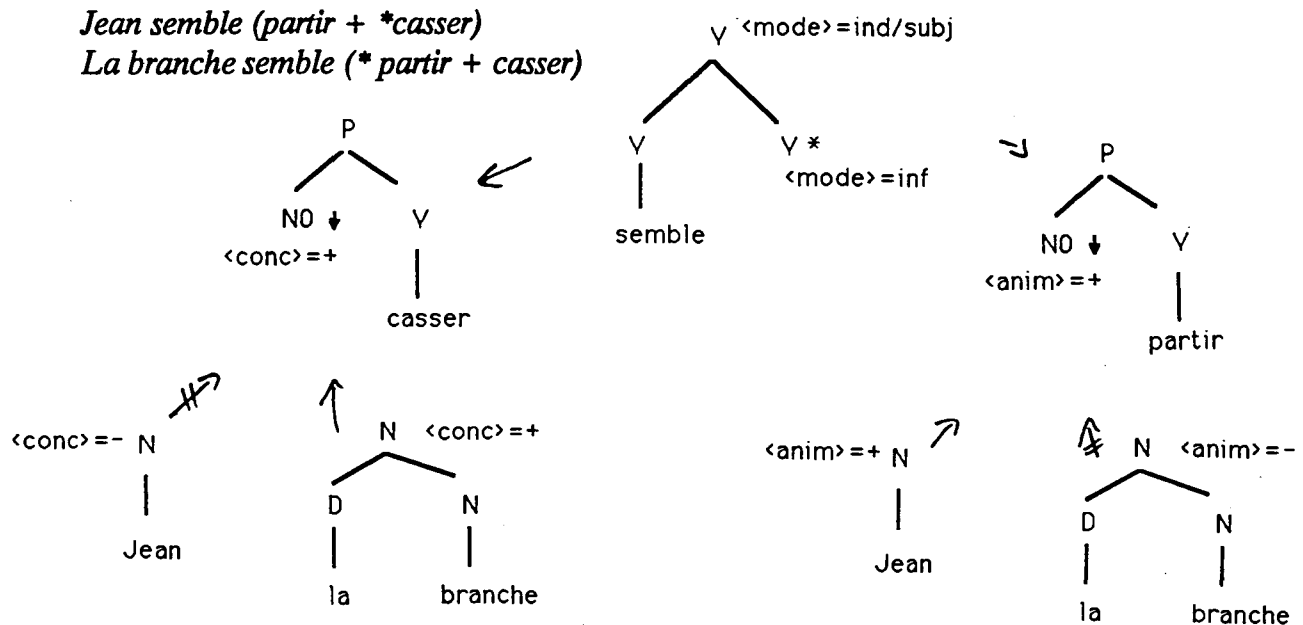


Figure 6

Pour nous, la restriction du sujet par l'infinitif dans les structures à montée est donc automatique et se fait de la même façon que pour les phrases simples :

*Jean (part + *casse).*
*La branche (*part + casse).*

Pour les verbes "equi" en revanche, il y a corestriction du sujet par le verbe principal et par le verbe à l'infinitif et un calcul particulier est nécessaire. Il faut faire intervenir des contraintes structurales et lexicales.

2.1. Contraintes structurales et lexicales

La valeur du sujet de l'infinitive dépend de la principale. Dans certains cas, il est structuralement déterminé; dans d'autres, il varie selon le verbe de la principale. Il est structuralement déterminé par exemple pour les familles n_0Vs1 (N_0 vouloir P_1), n_0Vps1 (N_0 rêver de P_1) ou s_0Vpn_1 (P_0 déplaire à N_1)² :

Jean_i veut aller_i au cinéma
Jean_i rêve d'aller_i au cinéma
Aller_i au cinéma déplaît à Jean_i

Il est lexicalement déterminé pour les familles $N_0VP_1PrepN_2$ ou $N_0VN_1PrepN_2$:

*Jean_i permet à Paul_j de venir_i/*j*
*Jean_i avoue à Paul_j être coupable_i/*j*

² On note dans les exemples la corréférence par coindage mais cette représentation n'est pas incorporée à la grammaire TAG.

Jean_i convainc Paul_j de partir_{*i/j}
Jean_i menace Paul_j de le battre_{i/*j}

On définit un trait $\langle \text{cont} \rangle$ (pour contrôleur) ayant pour valeur le faisceau des traits que peut avoir un argument : $\langle \text{num} \rangle$, $\langle \text{genre} \rangle$, $\langle \text{pers} \rangle$, $\langle \text{hum} \rangle$, $\langle \text{abs} \rangle$... En utilisant le système de traits actuels, il faut prévoir tous les traits que peut avoir le contrôleur et écrire une équation pour chaque trait, par exemple:

$\langle \text{cont num} \rangle = N_0: \langle \text{num} \rangle$
 $\langle \text{cont genre} \rangle = N_0: \langle \text{genre} \rangle$
 $\langle \text{cont hum} \rangle = N_0: \langle \text{hum} \rangle$...

Il serait plus satisfaisant d'avoir une équation "dynamique" (à la manière de celles de LFG, $(\uparrow \text{Subj}) = (\uparrow \text{Obj})$ qui identifie la structure fonctionnelle (c'est-à-dire tous les traits) de Subj et celle de Obj). D'une part, parce qu'il n'y a pas de limite au nombre de traits que peut avoir N_0 ; d'autre part parce qu'en cas d'ambiguïté entre deux contrôleurs possibles, il s'agit de tester en bloc les valeurs de l'un et de l'autre et non un mélange des deux. Si l'on note $\langle \text{cont} \rangle = N_0$, N_0 est interprété comme la valeur lexicale du trait $\langle \text{cont} \rangle$. Il faudrait donc une autre notation, par exemple³ $\langle \text{cont} \rangle = N_0:$.

Le trait $\langle \text{cont} \rangle$ est défini au noeud racine P des arbres des infinitives et instancié par le verbe tête de l'infinitive (qui comporte généralement des restrictions sur N_0 dans son entrée lexicale). On a ainsi avec N_0 *casser* $\{N_0: \langle \text{conc} \rangle = +\}$ ou N_0 *se réunir* $\{N_0: \langle \text{num} \rangle = \text{pl}\}$:

dans le fichier d'arbres

avec têtes lexicales spécifiées

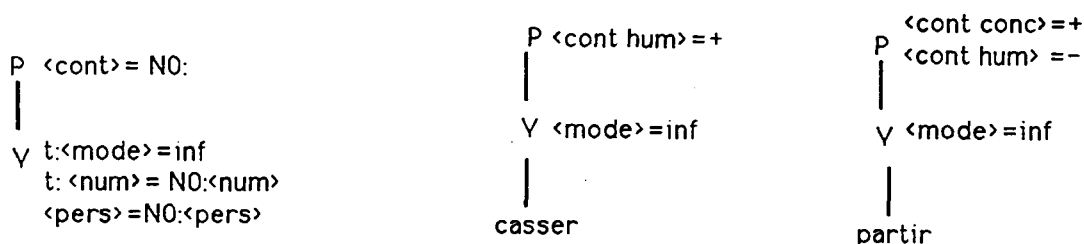


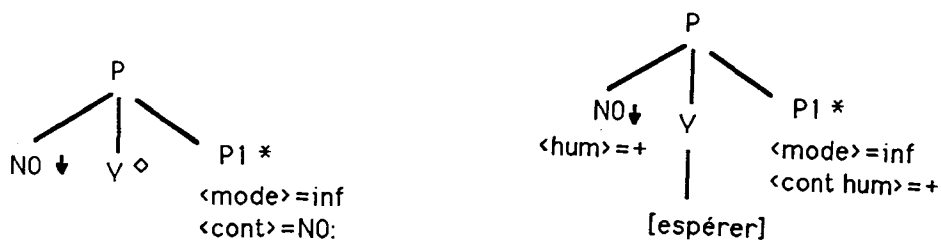
Figure 7

Dans l'arbre phrastique auxiliaire, dont la tête est le verbe "equi", le même trait $\langle \text{cont} \rangle$ est présent au noeud feuille de la complétive. Sa valeur est soit directement marquée dans l'arbre, soit instanciée lexicalement par le verbe tête (par exemple, N_0 *permettre* P_1 à N_2 $\{P_1: \langle \text{cont} \rangle = N_2:\}$ ou N_0 *avouer* P_1 à N_2 $\{P_1: \langle \text{cont} \rangle = N_0:\}$ dans le dictionnaire).

³ Le problème est que ce type d'équations entre traits peut, comme pour les grammaires LFG, augmenter considérablement le pouvoir génératif de la grammaire.

dans le fichier d'arbres

avec têtes lexicales spécifiées



dans le fichier d'arbres

avec têtes lexicales spécifiées

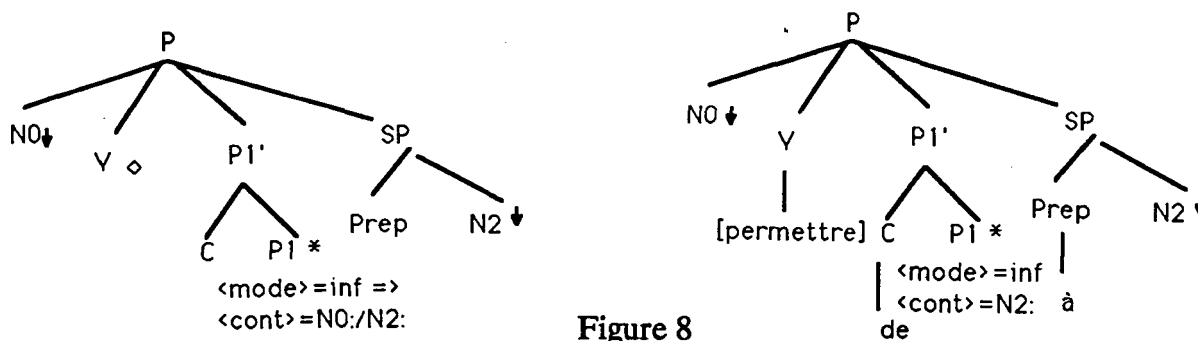


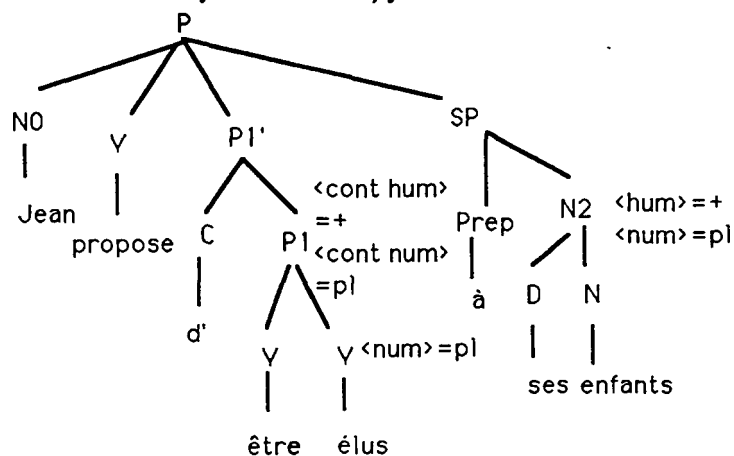
Figure 8

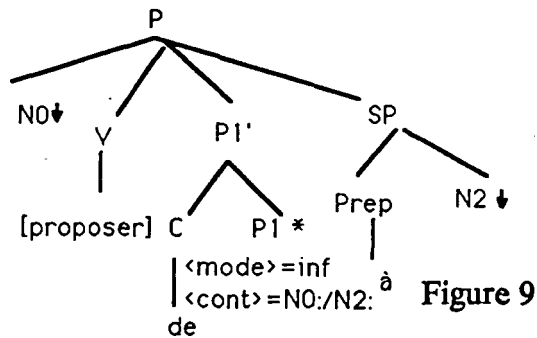
Lorsqu'il y a ambiguïté possible, on note (dans l'arbre de la "principale") pour le <cont> la disjonction des valeurs possibles, par exemple : $N_0 \text{ proposer à } N_1 P_2 \{P_2: \langle \text{cont} \rangle = N_0: / N_1:\}$

Jean_i a proposé à Paul_j de venir_{i/j} demain.

Bien entendu, quand les deux contrôleurs possibles ont des traits différents, seul celui dont les traits correspondent à ceux de l'infinitif sera retenu:

Jean_i propose à ses enfants_j d'être élus_{i/j} à sa place





Pour les familles n_0Vs_1 , n_0Vps_1 , s_0Vn_1 , s_0Vpn_1 ... (et de la même façon pour les figées), nous notons ce trait $\langle \text{cont} \rangle$ directement dans les arbres élémentaires. Pour les familles $n_0Vs_1pn_2$, $n_0Vps_1pn_2$... il est marqué dans le dictionnaire des verbes.

2.2. Cas particuliers

a) Avec les contrôleurs optionnels, le trait $\langle \text{cont} \rangle$ existe même si l'argument n'est pas réalisé :

Le général a ordonné (E + aux soldats_i) de tirer_i sur l'ennemi
Paul a dit (E + à Jeanne_i) de partir_i à midi

L'équation $\langle \text{cont} \rangle = N_2$: (dans les arbres de N_0 dire P_1 (à N_2) ou N_0 ordonner P_1 (à N_2)) n'instancie que les contraintes imposées par les verbes *ordonner* ou *dire* si N_2 n'est pas réalisé, puisque les traits venant de N_2 ne sont pas spécifiés. Ainsi le sujet de l'infinitif sera marqué pour certains traits sémantiques ($\langle \text{hum} \rangle = +$) et non pour les traits de nombre ou de genre. Il conviendrait peut-être d'ajouter alors une valeur par défaut :

Nous étions en retard. Le général avait ordonné d'être (parti + ??partis) à midi.

b) Pour des verbes à deux arguments phrastiques à l'infinitifs, il peut y avoir corestriction entre les contrôleurs de chacun des deux arguments:

$\text{Suj}_i \text{ Vinf} \cap \text{V (Prep) Suj}_i \text{ Vinf} \cap$
 $=:$ *Aller_i à la pêche revient à perdre_i son temps*

On écrit alors l'équation entre les variables des traits $\langle \text{cont} \rangle$: $P_0 \langle \text{cont} \rangle = P_1 \langle \text{cont} \rangle$:

c) On peut laisser la valeur de $\langle \text{cont} \rangle$ non spécifiée dans l'arbre de la "principale", par exemple avec: P_0 entraîner N_1 .

Rouler_x ivre peut entraîner la mort de milliers de gens.
Acheter_x ce terrain a conduit à un désastre.

Les phénomènes d'ambiguïtés entre deux contrôleurs possibles ont été vus supra. Mais il peut y avoir ambiguïté entre contrôleur argument et contrôle extérieure (non marqué), par exemple avec P_0 tuer N_1 :

2.4. Infinitives ajouts

Il faut étudier chaque conjonction (ou préposition). Avec *sans*, le contrôleur est généralement le sujet de la phrase, avec *pour* la situation est plus ambiguë⁵ :

- (1) *Jean_i travaille pour faire_j plaisir à Marie*
- (2) *Jean_i a engagé Marie_j pour faire_j/?*_i la vaisselle*
- (3) *Jean_i a donné l'appareil à Paul_j pour réparer_j/??_i la bicyclette*

On ne peut exclure pour les phrases (2) et (3) l'interprétation, certes marginale, où Jean est le sujet de l'infinitive (et se décharge d'autres tâches sur Marie et Paul respectivement). Il semble qu'on sorte ici des phénomènes réglables par la seule analyse syntaxique (en termes bresnaniens, on quitte le *functional control* pour entrer dans le *anaphoric control*). Nous ne traitons pas des phénomènes de contrôle dans les relatives (ex: *Je cherche une chemise à laver*) ou après les adjectifs de type *facile à*.

3. Variations d'ordre des mots

Les structures élémentaires présentées jusqu'ici encodent un ordre des mots fixes, qui ne correspond pas toujours à la variété des faits observés. Nous avons vu au chapitre 3 les permutations possibles entre compléments. Nous allons mentionner ici d'autres phénomènes.

A. Joshi 1987⁶ a présenté une variante des TAG qui distingue le composant structural des règles d'ordre linéaire. Dans ce système, les arbres n'encodent que les relations de dominance, toutes les relations de précédence (pour des noeuds appartenant au même arbre élémentaire) sont a priori possibles. Le système semble bien adapté à des langues à ordre des mots libre, comme le coréen ou le japonais. Il est linguistiquement trop expressif pour les langues à ordre plus fixe comme le français ou l'anglais. De plus, comme pour le format ID/LP en GPSG, il rend les grammaires formellement trop puissantes pour une analyse efficace.

Nous mentionnons ici quelques cas de variations d'ordre des mots plus contraints: inversion du sujet, adverbes mobiles, extraposition.

3.1. Inversion du sujet

Il s'agit ici de l'inversion simple de N, l'inversion complexe et l'inversion de clitiques sujets ayant déjà été vues (chapitre 5 et ici supra).

On la trouve dans les relatives, les phrases clivées ou les interrogatives indirectes avec des verbes intransitifs ou au passif:

⁵ Nous donnons ici des exemples avec *pour* final, mais les ambiguïtés potentielles sont les mêmes avec *pour* causal (pour une étude systématique voir M. Piot 1978):

Pour avoir_i frappé son frère, Marie_i a été punie
On l_i'admirait beaucoup pour avoir_i résisté à toutes ces épreuves

*C'est à Marie que pense Jean.
Marie, à qui pense Jean, arrive demain.
Je sais à qui pense Jean.*

S'agissant de structures élémentaires spécifiques, on peut ajouter les arbres élémentaires correspondants à l'inversion du sujet. Cependant, ce procédé ne convient pas si l'élément extrait est un ajout.

*C'est à Paris que dort Jean.
La ville où dort Jean est surpeuplée.
Je sais où dort Jean.*

Il faut alors une règle d'ordre particulière activée par les traits apportés par l'ajout.

L'inversion simple se trouve aussi avec l'antéposition de certains compléments prépositionnels ou de certains adverbes:

*Parmi les exportations figure l'alcool (M. Salkoff 1973)
Ici apparaissent des difficultés.*

Pour les compléments, l'on peut ajouter l'arbre élémentaire correspondant (ou une règle d'ordre) puisque peu de verbes sont concernés (Cf chapitre 3). Pour les adverbes, l'on peut aussi considérer les verbes intransitifs comme comportant une structure élémentaire supplémentaire inversée avec adjonction obligatoire de ce type d'adverbes.

Les inversions en incises font partie de la définition même des structures élémentaires des incises (qui appartiennent aux familles des verbes à complément phrastique).

3.2. Adverbes mobiles

Les adverbes de phrase doivent être aussi représentés comme des arbres auxiliaires de racine V, pour pouvoir s'interposer entre sujets et verbes ou entre verbes et compléments:

*Jean probablement partira demain
Jean verra probablement Marie*

On considère que c'est un même arbre auxiliaire avec disjonction sur les catégories pour ne pas générer de fausses ambiguïtés avec les verbes intransitifs:

Jean partira probablement

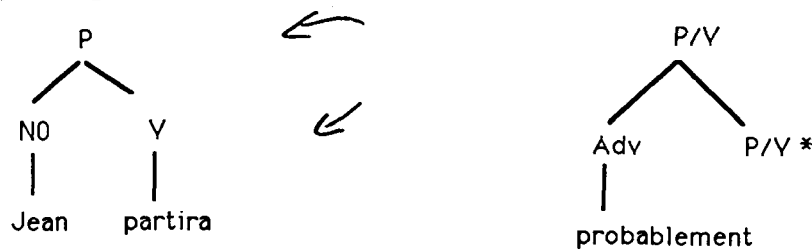


Figure 11

Cette solution ne s'applique pas si l'on veut interposer l'adverbe entre deux compléments:

Jean proposera à Paul certainement de venir
Jean a donné le livre probablement à Paul

Les grammaires en chaîne distinguent les ajouts à gauche, les ajouts à droite, et ceux qui sont non spécifiés, c'est-à-dire qui peuvent se mettre entre n'importe quels constituants d'une chaîne donnée. Les adverbes de phrase sont donc naturellement traités comme ajouts non spécifiés (M. Salkoff 1979). Un problème risque néanmoins de se poser pour l'interposition entre deux compléments puisque ceux-ci sont regroupés sous un noeud Objet (Ω).

On pourrait définir un mécanisme particulier, noté $\oplus$, à certains noeuds pieds, permettant cette souplesse. L'interposition entre deux compléments, ou entre nom prédicatif et complément de nom dans les phrases à verbe support ne poserait pas ici problème puisque les noeuds concernés appartiennent bien au même arbre élémentaire :

Jean a donné l'ordre hier de partir
Jean a donné des fleurs hier à Marie

Les incises, qui sont des arbres auxiliaires de racine phrastiques, devraient être traitées de la même façon, aussi bien celles à inversion du sujet (du type *pensait Jean*,) que celles où la complétive est pronominalisée (*Jean l'avait bien dit*,).

3.3. Quelques cas d'extraposition

L'extraposition des sujets a été vue aux chapitres 3 et 4. Elle repose sur des arbres élémentaires particuliers. Nous allons mentionner ici d'autres phénomènes d'extraposition (relatives ou compléments de nom). On pourrait les représenter par des règles de variations d'ordre des mots. Mais il serait préférable de les traiter comme l'on traite les questions ou les autres phénomènes d'extraction.

Une variante des TAG est de considérer comme unités de la grammaire non plus

seulement des arbres élémentaires mais aussi des ensembles d'arbres élémentaires. Ces TAG multiples (*multicomponent TAG*) ont été principalement étudiées par D. Weir 1988. A. Kroch et A. Joshi 1989 propose de les utiliser pour les extrapositions de relatives ou de compléments de noms en anglais. Des exemples pour le français seraient :

- (1) *Un homme est venu, qui portait un chapeau.*
- (2) *Un portrait a été fait hier de cette fille qui habite près d'ici.*
- (3) *Ordre avait été donné aux soldats de ne pas tirer.*
- (4) *? Un portrait est arrivé au musée, de cette fille qui habite près d'ici.*
- (5) *? On avait mis plusieurs copies sur le bureau de son rapport sur les troubles en Ouganda*
- (6) *? Une décision est intervenue, qui a été prise sans consultation préalable.*

Dans notre grammaire, la phrase (1) nécessite seulement une relative s'adjoignant à la phrase, et un système de traits permettant de la relier au nom sujet. Les phrases (2) et (3) non plus ne nécessitent pas de mécanisme particulier puisque l'extraposition des compléments de noms est prévue dans les arbres élémentaires pour les phrases à verbe support (Cf chapitre 4).

Les phrases suivantes en revanche, seraient des exemples de TAG multiples :

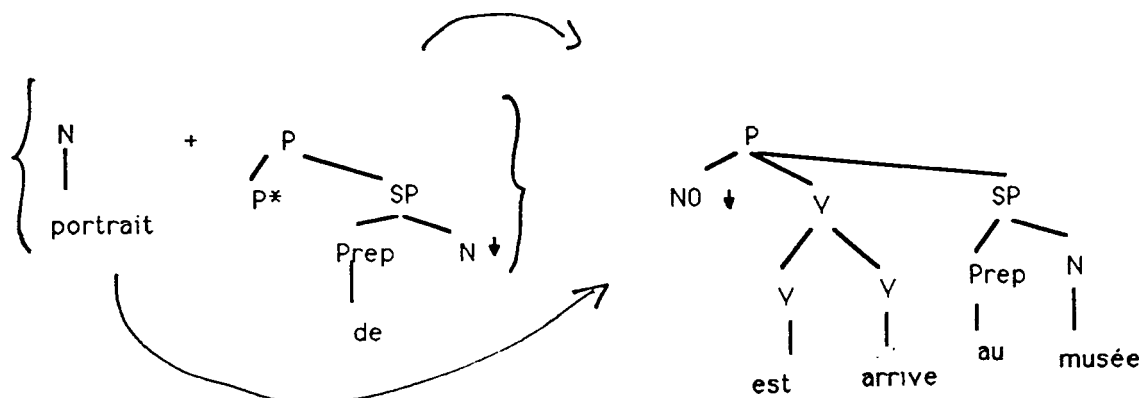


Figure 12

Le problème des restrictions (notamment lexicales) portant sur ces cas d'extraposition reste entier.

On pourrait utiliser le même mécanisme pour les extractions en *combien* hors du groupe nominal:

Combien Jean vend-il de tableaux ?

En effet, le déterminant discontinu *combien de* ne peut ici former un seul arbre élémentaire, puisqu'il doit s'adjoindre à la fois en P et en N.

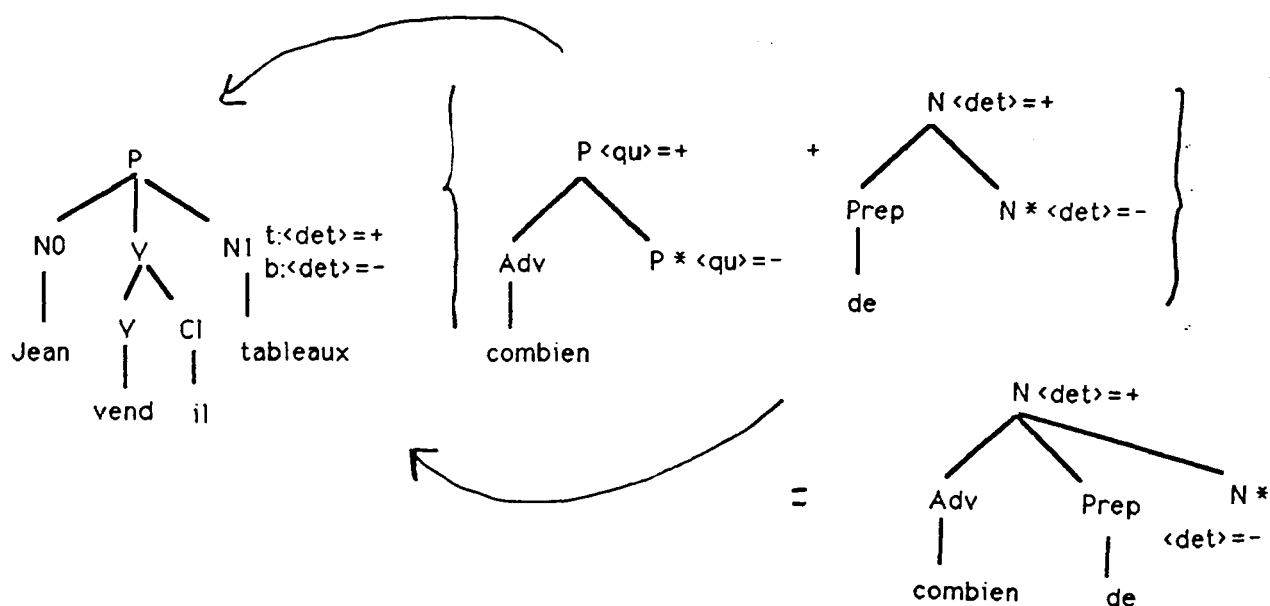


Figure 13

Si l'on limite un tel ensemble d'arbres élémentaires à se combiner toujours avec un seul arbre (ou ensemble d'arbres) élémentaire(s), le pouvoir génératif du formalisme reste le même, et l'analyseur que nous utilisons peut les analyser (Cf chapitre 7).

Nous n'avons pas intégré cette variante dans notre grammaire pour l'instant. Même si elle n'augmente pas le pouvoir génératif formel du formalisme, elle risque de donner, linguistiquement, un pouvoir expressif trop grand au système et de fournir trop de possibilités de choix un peu arbitraires.

En effet, définir deux arbres élémentaires indépendants avec un système de corestrictions de traits peut en général permettre le même résultat qu'un arbre élémentaire "multiple". La différence se fera dans l'arbre de dérivation, puisque le nombre d'unités indépendantes combinées sera différent, c'est-à-dire au niveau de l'interprétation sémantique. Les TAG multiples permettent en cela de rester plus proches de l'intuition initiale selon laquelle un arbre élémentaire doit être une structure argumentale complète.

4. Phénomènes de coordination

Comme le note M. Salkoff 1979, il s'agit là d'un des problèmes les plus ardues de l'analyse automatique. Les types de coordination sont multiples: coordination de constituants de même catégorie ou de catégories différentes, coordination de plus d'un constituant à la fois ou coordination à ellipse... , et les contraintes linguistiques les gouvernant n'ont pas toujours été décrites avec précision.

D'un point de vue formel, l'on risque d'augmenter considérablement le pouvoir génératif de la grammaire si l'on autorise des duplications de séquences quelconques de catégories.

La plupart des grammaires se limitent donc à certains cas de coordination. M. Salkoff 1979 présente un système assez complet basé sur une stratégie d'analyse particulière comportant des retours en arrière assez coûteux: quand l'analyseur rencontre un élément coordonnant, il fait l'hypothèse de toutes les sous-séquences des suites de

catégories vues jusque là, avec éventuellement mise à zéro de certains éléments. Ainsi, si la conjonction suit *Jean mange une pomme*, les différentes possibilités sont: *et une banane* (Objet), *et épluche une orange* (verbe et Objet), *et Paul regarde la télé* (Phrase), *et Paul une banane* (Phrase avec effacement du verbe), *et Paul aussi* (Phrase avec effacement du verbe et de l'objet). La grammaire ne traite pas de coordinations de constituants de catégories différentes comme: *Paul est satisfait et au courant de tout* ou *La petite dit son nom et qu'on l'appelait Bibi*.

Il serait tout à fait possible d'adapter le mécanisme de M. Salkoff 1979 aux TAG. A. Joshi 1990 a proposé récemment un autre traitement de la coordination en TAG, inspiré de celui de M. Steedman 1985 dans le cadre des grammaires catégorielles.

En grammaire catégorielle, il a été montré que ce qui apparaissait comme constituants "superflus" produits par l'analyse servait en fait à la coordination. L'analyse de M. Steedman repose sur la notion de "type raising" qui assigne deux catégories différentes aux groupes nominaux (NP en tant qu'argument du verbe ou (S/S\NP) en tant que foncteur prenant le verbe comme argument). Prenons l'exemple d'un verbe transitif, on peut combiner d'abord le sujet avec le verbe ou d'abord l'objet. On obtient donc deux dérivations concurrentes (*spurious ambiguities*) et deux sous-constituants possibles: sujet-verbe ou verbe-objet.

John	likes	apples		John	likes	apples	
NP	(S\NP)/NP	NP		S/S\NP	(S\NP)/NP	NP	
		<u>S\NP</u>				<u>S/NP</u>	
<u>S</u>				<u>S</u>			

Par simple duplication de constituants, on obtient donc naturellement deux cas de coordination qui dans d'autres cadres seraient des coordinations de non constituants :

John likes apples and hates oranges
John likes and Bill hates apples

Le traitement que propose A. Joshi 1990 s'inspire de celui-ci tout en évitant les *spurious ambiguities*.

En TAG, l'on peut distinguer la coordination de constituants, qui se fait simplement en assignant à *et* (et aux autres conjonctions de coordination) les structures élémentaires suivantes:

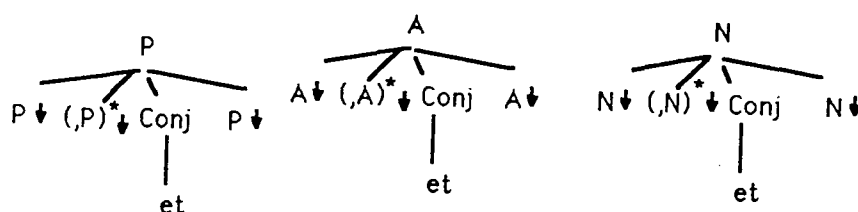


Figure 14

La grammaire étant lexicalisée, c'est seulement s'il y a une telle conjonction de coordination dans la phrase qu'on fera l'hypothèse de ce type de structures.

On peut noter que ce traitement n'est possible que si les constituants en question sont des noeuds racine de structures élémentaires. Nous ne pouvons pas traiter ainsi les coordinations à des noeuds intérieurs, comme les coordinations de verbes ou de SP arguments:

Jean dort et rêve

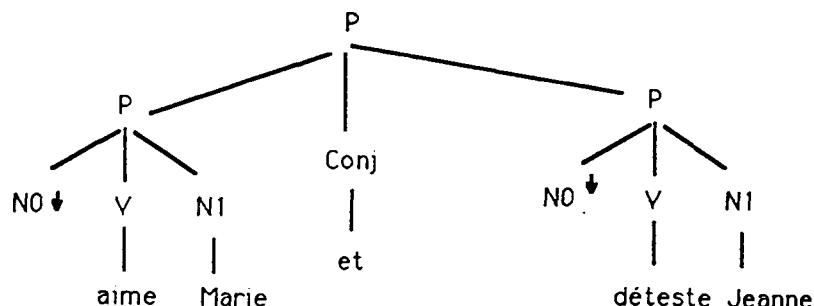
Jean pense à Marie et à Jeanne

Nous ne pouvons pas non plus traiter ainsi les coordinations de plus d'un constituant à la fois : *Jean a donné une pomme à Jeanne et une fleur à Suzanne*, ou les coordinations à ellipse: *Paul aime Marie et déteste Jeanne*.

Le traitement que propose A. Joshi 1990 pour l'anglais permet la mise en facteur de certains sous-arbres élémentaires ou dérivés. Nous adaptons ses exemples aux français.

L'on peut interpréter les arbres élémentaires comme des fonctions qui prennent des arguments (les noeuds feuilles) et retournent un résultat (le noeud racine). Ainsi l'arbre de *aimer* est une fonction à deux arguments $\text{Aimer}(N_0, N_1)$ ou en notation catégorielle $(P \backslash N) / N$, celui de *dormir* une fonction à un argument $\text{Dormir}(N_0)$ ou $(P \backslash N)$. Notons que l'on peut interpréter de la même façon des arbres en cours de dérivation. A la dérivation partielle *Jean aime N_1* correspond la fonction *Jean-aimer* (N_1) ou P / N , à *N_0 aime Marie* correspond la fonction *aimer-Marie* (N_0) ou $P \backslash N$, puisque l'ordre dans lequel on substitue les arguments de *aimer* est libre.

Il s'agit de combiner de telles fonctions en leur faisant partager leurs arguments. Prenons l'exemple de *Jean aime Marie et déteste Jeanne*. Commençons par substituer *Marie* en N_1 dans la première phrase *N_0 aime N_1* et *Jeanne* en N_1 dans la deuxième *N_0 déteste N_1* . L'on obtient deux fonctions de N_0 vers P , que l'on coordonne, avec mise en facteur de N_0 , c'est-à-dire effacement du second N_0 . Ce serait la même chose avec *Jean aime Marie mais préfère parler à Jeanne*, les deux dérivation partielles *N_0 aime Marie* et *N_0 préfère parler à Jeanne* étant de même type.



=>

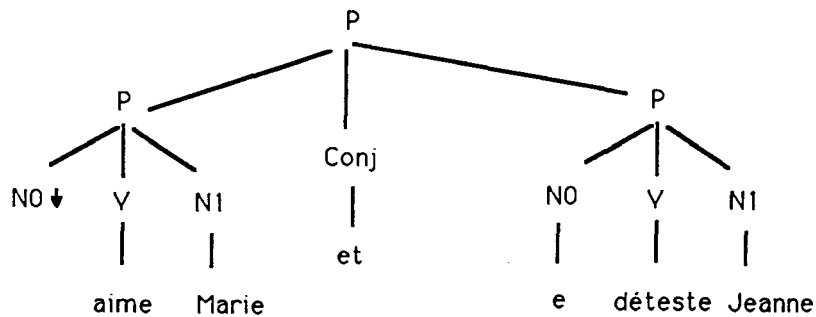


Figure 15

Le même traitement vaut pour *Jean dort et rit*, sauf que la coordination a lieu directement entre deux structures élémentaires et non en cours de dérivation, et qu'elle peut avoir lieu au noeud V (et non au noeud P):

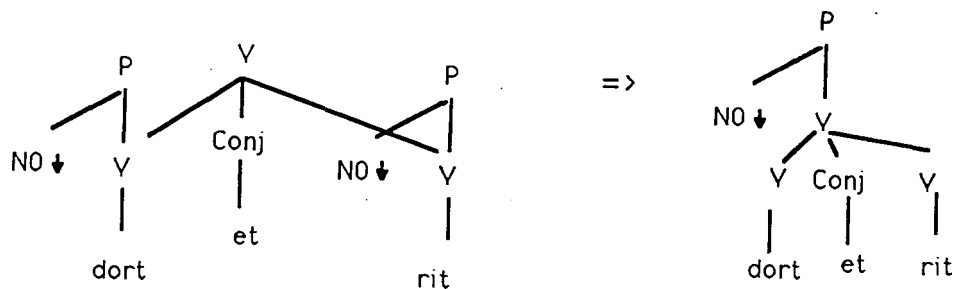


Figure 16

Le traitement serait le même pour la coordination de deux ajouts, par exemple deux adverbes s'adjoignant à P, la coordination ayant lieu dans les arbres auxiliaires au noeud Adv immédiatement dominé par P.

Le principe général est le suivant. Soient deux arbres élémentaires ou dérivés, correspondant à fonctions de même arité (et de même type):

- 1) repérer les arguments "partagés".
- 2) coordonner les noeuds les plus bas dominant les deux séquences lexicales réalisées à condition qu'elles soient de même catégorie.
- 3) "linéariser" en effaçant au besoin ce qui domine le noeud où a lieu la coordination.
- 4) choisir pour la réalisation de l'argument mis en facteur le noeud qui permet aux éléments lexicaux des fonctions coordonnées d'être contigus et attacher la chaîne vide à

l'autre noeud.

Ce processus n'est appelé que si l'on rencontre dans la phrase une conjonction de coordination.

La condition sur le plus bas noeud permet d'analyser *Jean dort et rit* comme une coordination de verbes et non de phrases, de même pour : *Jean regarde et admire Paul*.

La condition de contiguïté impose d'effacer la deuxième (et non la première) occurrence de N_0 dans N_0 aime Jeanne et N_0 déteste Paul:

Jean aime Jeanne et déteste Paul

* *aime Jeanne et Jean déteste Paul*

De la même façon, on efface la deuxième occurrence de N_0 et la première de N_1 dans N_0 acheter N_1 et N_0 envoyer N_1 à N_2 .

Jean a acheté et envoyé ces fleurs à Marie

* *Jean a acheté ces fleurs et envoyé à Marie*

Ce système est en cours de développement et n'a pas encore donné lieu à implémentation. Son interfaçage avec l'analyseur est délicat car il suppose de manipuler des arbres sans têtes lexicales, ou de pouvoir effacer une tête lexicale pour mettre par exemple un noeud V en facteur en cours d'analyse, comme dans : *Jean aime Marie et Paul Suzanne* ou *Jean a donné une pomme à Suzanne et un livre à Jeanne*.

Il est comme tel trop puissant car il autorise aussi bien (1) que (2):

(1) *I saw one man_i yesterday and two e_j today*

(2) * *I saw that man_i yesterday and this e_j today*

De plus, il autorise des coordinations existant en anglais mais non en français telles que:

* *Jean pense à et rêve de Marie*

Chapitre VII

L'analyse syntaxique automatique

Comme nous l'avions annoncé en introduction, nous avons fait le choix d'une grammaire la plus étendue et la plus explicite possible. Nous nous distinguons donc des systèmes d'analyse qui ne s'appuient pas sur des règles de grammaire, mais sont fondés sur des probabilités sur les séquences (n-grammes) de catégories (par exemple F. Débili 1977). L'intérêt d'une telle méthode est que l'analyseur peut construire sa propre grammaire à partir des suites de catégories rencontrées effectivement dans des textes, et classer ces suites par ordre de probabilité. Mais elle rencontre un taux d'erreur incompressible qui conduit généralement à lui associer un module fondé sur des règles de grammaire plus classiques (Cf A-M. Derouault 1985).

Nous nous distinguons également d'une autre optique d'apprentissage automatique où l'on fournit au système des phrases exemples dont l'analyseur extrait le moule (G. Francopoulo 1988).

Dans les deux cas, le système ne peut fournir que des analyses minimales (étiquetage catégoriel, groupements élémentaires...) et dont la fiabilité n'est jamais garantie.

Le va et vient entre l'écriture d'une grammaire et l'expérimentation sur des phrases variées est bien entendu nécessaire, mais pour nous seul un francophone linguiste peut établir les règles syntaxiques et lexicales avec toute la précision nécessaire.

Nous allons maintenant définir les conditions que notre système entend remplir. Rappelons que nous nous plaçons dans le cadre de l'analyse de phrases isolées, et non de l'analyse de textes.

a) Le programme est un analyseur, non un reconnaisseur.

Il ne doit pas seulement dire si la phrase est correcte ou incorrecte, il doit aussi en fournir une représentation, qui pourra ensuite être utilisée pour d'autres traitements.

Si la phrase est correcte, il s'agit de fournir plusieurs analyses si la phrase est syntaxiquement ou lexicalement ambiguë (voir infra).

Si la phrase est incorrecte, le constat d'échec de l'analyse ne suffit pas, et il peut être intéressant d'utiliser les fragments d'analyse construits (qu'il s'agisse de produire des explications dans le cadre d'un système d'aide, ou de tenter une interprétation dans le cadre d'un système "tolérant").

b) L'analyseur doit analyser les phrases correctes et seulement celles-là¹.

On peut se demander pourquoi une telle exigence à partir du moment où il s'agit d'analyser des textes écrits (dépêches, journaux...), a priori corrects. On pourrait donc

¹ A partir du moment où la correction peut être définie en termes purement formels. Les problèmes de référence et de coréférence restent entiers.

penser que, puisque les phrases à analyser seront toujours bien formées (une fois corrigées les éventuelles fautes typographiques), la question de l'exclusion des phrases incorrectes ne se pose pas, et qu'une grammaire plus lâche ferait l'affaire.

Il n'en est rien. Même si toutes les phrases à analyser sont correctes, une grammaire plus lâche que la grammaire des seules phrases correctes va créer des ambiguïtés artificielles pour les phrases correctes elles-mêmes.

Par exemple, si l'on accepte *avec N* comme modifieur de nom quel que soit le déterminant (et les traits sémantiques) de N, on acceptera comme groupe nominal la séquence: ** une fille avec gentillesse* mais on analysera également comme ambiguë la phrase : *Il a parlé à cette fille avec gentillesse.*

De même si on ne limite pas la "montée" des clitiques à certains verbes (*faire, laisser*), on accepte la phrase incorrecte: ** Je le sais manger* qu'on analyse comme *Je sais le manger*. Mais on analyse comme ambiguë la phrase suivante: *je le sais boucher*, soit comme *je sais Jean boucher* (= *je sais qu'il est boucher*) ou comme *je sais boucher ce trou* (= *je sais le boucher*) alors que seule la première analyse est pertinente.

Ce n'est donc pas parce qu'on se limite à une application sur phrases écrites et sur phrases correctes que l'on peut s'autoriser une grammaire plus floue.

Un deuxième argument est de pouvoir utiliser cette grammaire pour d'autres applications: si on veut l'utiliser pour l'enseignement, les phrases entrées ne seront pas toujours correctes; de même si on l'utilise pour la génération, il faut bien sûr éviter de générer des phrases incorrectes (d'autres contraintes sont aussi à ajouter dans cette optique, par exemple ne pas générer de phrases ambiguës, Cf infra).

c) L'analyseur doit donner toutes les analyses possibles d'une phrase.

Il ne s'agit donc pas pour nous d'utiliser un analyseur déterministe. Il ne s'agit pas non plus d'établir des règles de préférence ou de priorité chargées de privilégier une analyse parmi d'autres. Ce type de règles est souvent fragile, il est toujours limité à un contexte. Il ne se justifie donc que pour une application particulière.

Il s'agit en revanche de ne pas créer d'ambiguïtés artificielles en multipliant les analyses superflues. C'est un problème difficile, et nous n'y avons pas toujours réussi (Cf section 3).

d) Notre analyseur doit être suffisamment général pour s'adapter à diverses applications.

C'est pourquoi, tout en faisant intervenir certaines connaissances sémantiques lors de l'analyse syntaxiques (figement, restrictions de sélections), nous séparons analyse syntaxique et analyse sémantique. Nous pensons qu'il est souhaitable d'avoir un analyseur syntaxique autonome car la syntaxe fait intervenir des connaissances plus générales que la sémantique dont les connaissances sont souvent liées à un domaine d'application particulier. Notre grammaire se veut donc générale et nous en avons tenu compte pour constituer notre dictionnaire. Nous nous sommes basée sur les listes de mots les plus courants (A. Juilland et al. 1970, N. Catach 1984), complétées par des

expressions figées (tirées des listes de M. Gross) pour lesquelles il n'existe pas de listes de fréquences, et que nous avons donc sélectionnées de façon intuitive en nous fondant sur notre lecture des journaux (*Libération, le Monde*).

Nous ne construisons donc pas un système pour une application dans un domaine précis, mais un module suffisamment général (et modulable) pour être intégré (et adapté) à diverses applications.

Notre grammaire et notre système d'analyse doivent donc être comparés avec des systèmes ayant les mêmes ambitions.

Nous allons rappeler quelques systèmes d'analyse existant pour le français avant de présenter notre analyseur, et en particulier la stratégie d'analyse induite par la lexicalisation de la grammaire. Nous exposons ensuite brièvement quelques problèmes liés à l'ambiguïté potentielle des phrases à analyser, et enfin mentionnons d'autres applications en cours de notre grammaire.

1. Quelques analyseurs existant pour le français

Nous ne prétendons pas ici nous livrer à un inventaire exhaustif. Il s'agit simplement de donner quelques repères, en privilégiant les réalisations les mieux comparables à la nôtre, c'est-à-dire celles qui témoignent d'un réel souci de couverture lexicale et grammaticale dans la perspective d'une analyse syntaxique autonome. Nous ne mentionnons donc pas les analyseurs dont l'analyse syntaxique est "guidée" par une analyse sémantique simultanée (M. Rady 1983). Nous ne mentionnons pas non plus les modules d'analyse syntaxiques intégrés à des systèmes plus vastes et obéissant à des contraintes liées à une application spécifique, par exemple les analyseurs du Geta ou celui d'Eurotra, pour la traduction automatique, ou celui du système Vortex pour la correction d'orthographe.

La plupart utilisent des formalismes ou des théories linguistiques bien établis: grammaires en chaîne (M. Salkoff), Gouvernement et Liage (E. Wehrli), grammaires catégorielles (G. Bès et al.), grammaires logiques (A. Bérard-Dugourd). Il est à noter que les deux formalismes les mieux représentés pour l'analyse de l'anglais (GPSG et LFG) n'ont pas, à notre connaissance, donné lieu à des grammaires d'envergure pour le français.

1.1. M. Salkoff 1973 et 1979

Cette grammaire se base sur une liste de 150 chaînes syntaxiques, et environ 200 restrictions. Celles-ci ne sont pas exprimées par unification sur des structures de traits mais, de façon procédurale, par des conditions sur les suites de sous-catégories licites et interdites. M. Salkoff 1979 ajoute de très nombreuses restrictions de sélection sur les arguments des verbes, noms et adjectifs, qui limitent les ambiguïtés artificiellement créées par l'analyseur. Sont ainsi distinguées 15 classes de noms, 30 de verbes, 10 d'adjectifs

Ces chaînes sont combinées par remplacement (substitution d'un élément lexical ou d'une chaîne) ou ajout (qui ressemble à l'opération de furcation vue au chapitre 1 pour

les arbres). C'est donc une grammaire hors-contexte utilisant par commodité des règles dépendantes du contexte.

A l'analyseur des chaînes s'ajoute un composant spécifique pour la coordination: quand le programme rencontre une conjonction, il interrompt l'analyse pour engendrer toutes les chaînes conjonctionnelles possibles compte tenu des éléments qui précèdent.

La couverture de la grammaire est importante (c'est une des grammaires les plus complètes réalisées à ce jour), mais parfois au détriment de la finesse des descriptions linguistiques: ainsi des structures "plates" sont souvent préférées aux regroupements syntagmatiques linguistiquement justifiés. A la différence de la grammaire en chaîne construite pour l'anglais (N. Sager 1981), il n'y a pas non plus de module transformationnel proprement dit. La détermination d'une chaîne centrale par élimination des ajouts s'en rapproche, mais aucun lien n'existe entre phrases actives et passives, questions et déclaratives, etc...

La grammaire n'est pas lexicalisée: certaines chaînes comportent un ou plusieurs préterminaux, d'autres non (la mise à plat de nombreuses structures fait que beaucoup de chaînes comportent un préterminal). Le lexique est donc un composant séparé, il est de petite taille. Il ne comporte pratiquement pas d'expressions figées et de phrases à verbe support.

L'analyseur est de type descendant (en profondeur d'abord) et procède de gauche à droite. Il a été testé sur des textes de biologie.

1.2. J-H. Jayez 1982, A. Gall et al. 1986

Ces deux systèmes, d'ambition beaucoup plus limitée, s'inspirent également des grammaires en chaînes. A la différence de M. Salkoff, ils opèrent une analyse transformationnelle des groupes nominaux complexes (nominalisations), ou des questions simples, afin de reconstituer des relations entre phrases simples, par exemple :

entrée: achat de marchandises par des clients douteux
sortie: les clients achètent des marchandises
les clients sont douteux

Le premier système comprend un module d'analyse et un module de paraphrase. L'analyse est basée sur un lexique de catégories (avec des indices pour les sous-catégories) et un automate à état fini qui divise les intitulés en sous-chaînes et contrôle les suites de catégorie acceptables. La paraphrase utilise des chaînes (moules), un lexique complémentaire (verbes) et une base de connaissances spécialisées.

Le deuxième système analyse des questions variées portant sur des phrases simples. Il autorise une grande flexibilité: questions avec ou sans inversion, avec ou sans topicalisation, avec ou sans négation, clivées ou non, à l'actif ou au passif ... Il ne prend pas en compte, en revanche, les dépendances non bornées.

Le lexique est à chaque fois un composant séparé, et comprend dans les exemples de programme cités environ 30 mots. Les deux systèmes sont définis pour des applications de gestion.

1.3. E. Wehrli 1984, 1988

Ce système s'inspire de la théorie générative récente (GB). Il s'agit d'une approximation puisque cette théorie ne correspond pas à un formalisme directement utilisable de manière automatique.

Comme dans la théorie chomskyenne, l'approche est modulaire: module X-bar, module des rôles theta, module de liage (pour les traces).

La couverture de la grammaire est plus limitée que celle de M. Salkoff. Elle comprend notamment les complétives et les infinitives, passifs, auxiliaires et verbes à montée du sujet, constructions de type *facile à*, coordination de constituants sans ellipses, questions (et dépendances non bornées), clitiques, quelques expressions figées. Il est à noter que la représentation est très contrainte pour ces dernières: aucune variation syntaxique, ni insertion d'adverbes.

D'un point de vue informatique, le module syntagmatique (X-bar) est peu efficace: il laisse passer beaucoup de phrases incorrectes car il n'a accès à aucune information lexicale (comme la sous-catégorisation). C'est le module theta qui a charge d'éliminer ces analyses superflues. Il fait également la différence entre phrases idiomatiques et autres.

Le lexique est un composant séparé qui comprend plus de 2000 formes.

1.4. K. Bashung et al. 1987, G. Bès et C. Gardent 1989, C. Gardent et al. 1989

Il s'agit d'un système fondé sur une grammaire catégorielle d'unification (inspirée du formalisme UCG défini par J. Calder et al. 1988, avec certaines modifications) et qui comporte une composante particulière pour éliminer les analyses superflues (analyses différentes pour une phrase non ambiguë)². A la différence d'autres grammaires catégorielles, ce formalisme n'utilise que l'application (avant et arrière) et non la composition, ou les combinateurs.

La couverture syntaxique est relativement limitée: auxiliaires, clitiques (qui sont traités comme des arguments SN déplacés), complétives, questions (dépendances non bornées), inversion, négation, accords...

Une attention particulière a été portée à certaines variations d'ordre des mots en français (essentiellement celles des arguments SN, *qu-* ou clitiques), qui sont traitées par un système de traits inspiré de celui de L. Karttunen 1989.

L'analyseur est un *chart parser* qui procède de façon ascendante.

1.5. A-M. Bérard-Dugourd 1990

Ce système utilise un algorithme et une grammaire spécifique. Le formalisme est celui des grammaires logiques (grammaires à structure de rôle). La grammaire est constituée d'environ 350 règles de réécriture sensibles au contexte. Elle couvre différents types d'adverbiaux, les principaux types de phrases simples, les interrogatives, les relatives

² Un calcul sémantique est effectué en parallèle, inspiré de la *Discourse Theory* de H. Kamp 1984.

(non extraposées), les passifs, différents types de coordination et de comparatives.

Une attention particulière a été portée aux locutions, qui sont repérées dans une étape de prétraitement, ce qui suppose de se limiter à celles qui sont susceptibles de peu de variations (Cf chapitre 4). 650 locutions ont été ajoutées au dictionnaire originel.

La grammaire utilise un dictionnaire de 30 000 lemmes qui lui est préexistant, et qui comporte peu d'informations syntaxiques. Ainsi, les informations concernant la sous-catégorisation des complétives et les restrictions de sélection des verbes (et autres prédicats) sont très incomplètes, ce qui entraîne des analyses erronées.

Des heuristiques (règles de préférence) permettent de réduire le nombre d'ambiguïtés artificiellement créées par l'analyseur (pour les attachements prépositionnels ou l'analyse des coordinations), mais entraînent également un taux d'erreur incompressible

L'analyseur est de type ascendant et procède de droite à gauche (essentiellement pour optimiser la prise en compte du contexte dans les règles d'attachement). L'intérêt de ce système est dans son attention aux phénomènes rencontrés sur textes réels. Il s'agissait d'analyser des textes journalistiques (Le Monde, AFP), administratifs (Parlement canadien) et techniques (IBM).

Tous ces systèmes ont peu de liens, celui de M. Salkoff mis à part, avec une étude systématique du lexique. Tous, sauf celui d'A. Bérard-Dugourd, s'appuient sur des petits dictionnaires. Les expressions figées sont rarement prises en compte, leur variabilité ne l'est pratiquement pas. Aucune n'est lexicalisée au sens où nous l'entendons.

Nous ne mentionnons pas de nombreuses autres réalisations: P. Zweigenbaum et M. Cavazza 1990, F. Segond 1990... dont l'accent est davantage mis sur la conception de l'algorithme d'analyse que sur celle de la grammaire ou qui se donnent explicitement pour but d'analyser des textes avec une grammaire minimale (et un dictionnaire minimal: J. Vergne 1990).

2. Une stratégie d'analyse originale

La lexicalisation de la grammaire permet de définir de nouvelles stratégies de traitement automatique (voir Y. Schabes, A. Abeillé, A. Joshi 1988). Notre grammaire a été testée avec l'analyseur défini dans Y. Schabes et A. Joshi 1988 et en améliore considérablement les performances (par rapport à une grammaire non lexicalisée).

L'avantage de la lexicalisation pour l'analyse automatique est que le parseur n'a pas à connaître l'ensemble de la grammaire pour analyser une phrase. Une première lecture de la phrase lui permet d'aller chercher dans le lexique les arbres correspondants aux mots de la phrase à analyser, ensuite il procède à l'analyse syntaxique proprement dite, mais avec seulement ce sous-ensemble pertinent de la grammaire. Il est à noter que notre grammaire ne contraint pas le type d'analyse utilisée dans la seconde passe: toutes



les techniques classiques sont disponibles (Cf infra section 4).

Nous avons donc défini une stratégie d'analyse en deux temps :

- 1) L'analyseur lit la phrase et sélectionne les arbres élémentaires correspondant aux têtes lexicales rencontrées.
- 2) L'analyseur cherche à combiner ces structures entre elles (à adjoindre les arbres auxiliaires, à substituer les arbres initiaux aux noeuds marqués à substituer) de façon à obtenir un arbre dérivé de racine phrastique.

Dans la première passe, on attache donc les têtes lexicales aux arbres élémentaires et l'on note également leur place dans la phrase.

C'est une stratégie *off-line* (une stratégie *on line* serait capable de dire au fur et à mesure de la lecture d'une phrase et à tout moment si les mots rencontrés jusque là peuvent appartenir à une structure correcte ou non).

Notre stratégie en deux passes présente plusieurs avantages. Elle diminue considérablement la taille de la grammaire à explorer et est généralisable à la plupart des analyseurs existants. En particulier la plupart des problèmes dûs à des phénomènes de récursion peuvent être éliminés. Elle permet également de simuler des techniques de "regard en avant" (*look ahead*). Nous allons examiner ces avantages successivement.

2.1. Réduction de la taille de la grammaire

La grammaire effectivement utilisée pour l'analyse d'une phrase donnée (que nous appelons la grammaire de l'analyseur) est toujours un sous-ensemble de la grammaire totale. Si on a une phrase qui a autant de mots différents qu'il y a de structures différentes dans la grammaire, et si chacun de ces mot correspond à une structure élémentaire différente, les deux sont égales. Ce dernier cas est largement théorique (il y a 2000 structures différentes dans la grammaire).

Par exemple, pour la phrase: 0 Jean 1 aime 2 Marie 3, l'analyseur trouve trois entrées dans le dictionnaire syntaxique, et les arbres sélectionnés sont :

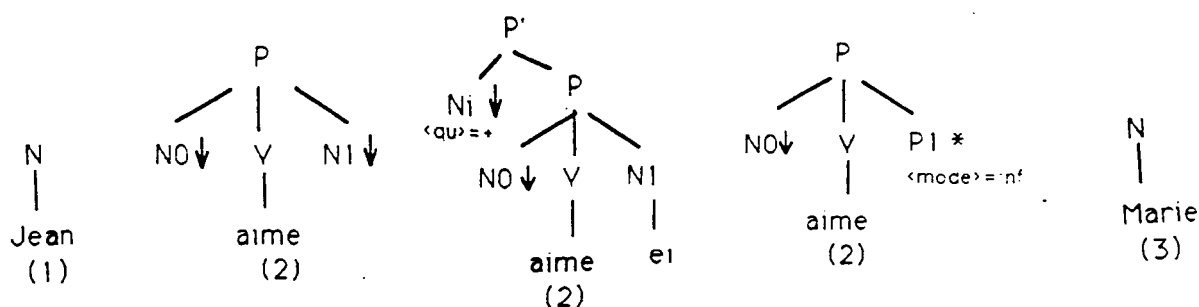


Figure 1

L'analyseur regarde dans le dictionnaire quels sont les arbres possibles pour les têtes lexicales de la phrase et attache ces têtes lexicales aux arbres correspondants. Quand les arbres ont des têtes multiples, il doit également vérifier que les autres têtes sont dans la phrase. Lorsqu'il attache une tête lexicale, il y a unification des traits au noeud feuille correspondant. Ainsi avec l'exemple précédent, on obtient :

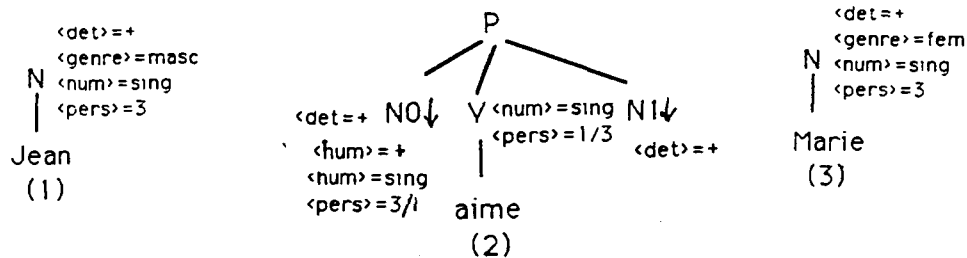


Figure 2

Ainsi les arbres de la grammaire du français correspondant aux adjectifs, aux adverbes, aux groupes prépositionnels ... ne sont pas sélectionnés puisqu'aucune tête lexicale de la phrase ne correspond à ces catégories. De même aucune structure phrastique à deux compléments ou à un complément adjectival n'est sélectionnée.

Tous les arbres de la famille de *N₀ aimer N₁* ne sont pas non plus sélectionnés: ainsi les arbres des relatives ne le sont pas car il n'y a pas de pronoms relatifs dans la phrase. De même, les arbres des clivées, puisqu'aucune forme de *ce* et *être* n'apparaissent dans la phrase. Les arbres passifs ne le sont pas non plus, car la tête lexicale n'étant pas au participe passé, il y a incompatibilité des traits au noeud V feuille.

Les arbres correspondant à *N₀ aimer que P_{subj}* ou *N₀ aimer à P_{inf}* ne sont pas sélectionnés car il n'est pas possible d'attacher les co-têtes lexicales (*que* ou *à*) aux arbres correspondants puisqu'elles ne figurent pas dans la phrase. En revanche les arbres de *N₀ aimer P_{inf}* le sont puisqu'ils n'exigent pas d'autre tête que *aimer*.

Cet avantage informatique est lié à certaines propriétés des langues naturelles, à savoir d'une part que les phrases sont relativement courtes et comprennent peu de mots par rapport à l'ensemble des mots existant dans la langue, d'autre part que la plupart des mots sont ambigus mais que leur taux d'ambiguïté (le nombre de structures où ils entrent) est nettement inférieur au nombre de structures existantes dans la grammaire (il en irait donc différemment si les TAG étaient utilisées pour les langages de programmation).

2.2 "Elimination" de la récursion

Une grammaire lexicalisée a toujours un degré fini d'ambiguïté (Cf chapitre 1). A chaque mot (ou groupes de mots) correspond donc un ensemble fini de structures élémentaires. La longueur de la phrase à analyser est également finie, par définition. L'analyseur ne peut donc pas boucler indéfiniment, puisqu'il "consomme" au moins un élément lexical à chaque utilisation d'une règle. Il n'y a pas d'utilisation itérative

possible des structures potentiellement récursives du formalisme (les arbres auxiliaires) puisqu'après la première passe elles sont différenciées par leur tête lexicale et par la position de cette tête dans la phrase. Chacune ne peut être utilisée qu'une fois au plus. Une conséquence est qu'on peut utiliser un analyseur purement descendant dans la deuxième passe, sans restrictions particulières. Les problèmes liés à l'analyse descendante (beaucoup d'hypothèses de syntagmes inutiles, risque de récursion infini avant d'atteindre le niveau lexical) disparaissent en effet avec une grammaire lexicalisée, puisque les têtes lexicales (avec leurs positions) fournissent une information de type ascendant³.

Cette information de type ascendant peut fournir l'équivalent d'un regard en avant dans la seconde passe (analyse proprement dit).

2.3 L'équivalent d'un "regard en avant"

Le fait d'encoder la position de la tête lexicale dans son arbre élémentaire dès la première passe contraint la façon dont l'analyseur peut combiner les arbres dans la seconde passe. En particulier, toute combinaison qui ne placerait pas les têtes lexicales en ordre croissant (conformément à l'ordre linéaire de la phrase) est exclue d'emblée. Par exemple, dans la phrase précédente (*Jean aime Marie*), *Jean* doit être substitué en N_0 , non en N_1 , puisque son indice précède celui du verbe *aimer*.

Cette information peut être équivalente aux techniques de "regard en avant" (*look ahead*), surtout utilisées par les analyseurs déterministes. Si l'on veut éviter les retours en arrière, le regard en avant est nécessaire car nombreuses sont les phrases dont l'analyse n'est donnée que par le dernier mot. Cette technique évite donc de faire de multiples hypothèses qui se révéleront ensuite erronées.

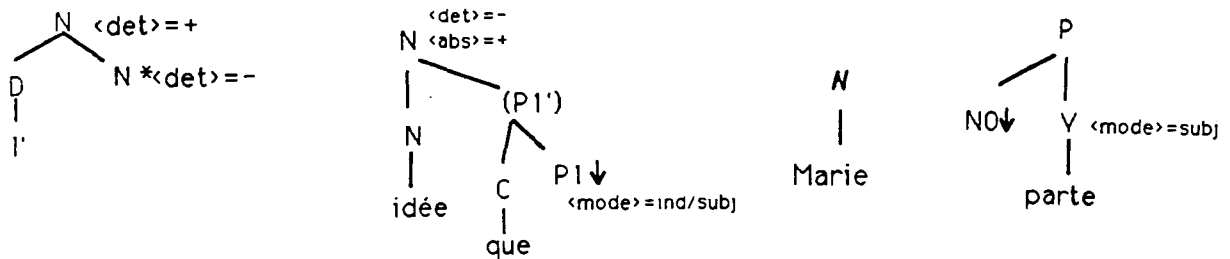
Prenons l'exemple des ambiguïtés possibles entre relatives et complétives. Pour savoir si *que* doit s'analyser comme relatif ou comme complémenteur, il faut connaître non seulement le nom qui le précède mais aussi le verbe de la phrase enchâssée (qui peut être à une certaine distance). Ainsi, *l'idée que Marie parte* doit s'analyser comme un SN à complétive, *le gâteau que Marie rejette* comme un SN à relative et *l'idée que Marie rejette ...* comme ambigu (selon que *rejette* est suivi d'un SN ou non). Pour les deux premiers exemples, aucune ambiguïté ne va apparaître et l'analyseur n'aura pas de choix à faire entre relative et complétive.

Le verbe enchâssé étant tête (avec le pronom relatif) de la relative (Cf chapitre 3), il n'y a pas de relative en *que* existante pour le verbe *partir*. Le complémenteur étant tête (avec le nom prédicatif) de la structure du SN complexe, il ne peut être sélectionné dans le cas de *gâteau* (qui n'est pas un nom à complétive)⁴.

3 Un analyseur de ce type, écrit en Prolog, a d'ailleurs été implémenté par Y. Schabes et B. Daille.

4 Cette utilité des arbres élémentaires à têtes multiples est détaillée section suivante pour le cas des expressions figées. Il faut noter que dans les exemples présentés ici, un arbre élémentaire (en N) pour *que* pronom interrogatif serait aussi sélectionné. Il sera éliminé lors de l'analyse et ne change rien au problème d'ambiguïté qui nous occupe. De même pour les arbres cliques sélectionnés par *le* et *l'*.

Arbres élémentaires sélectionnés dans la première passe pour *l'idée que Marie parte*



Arbres élémentaires sélectionnés dans la première passe pour *le gâteau que Marie rejette*

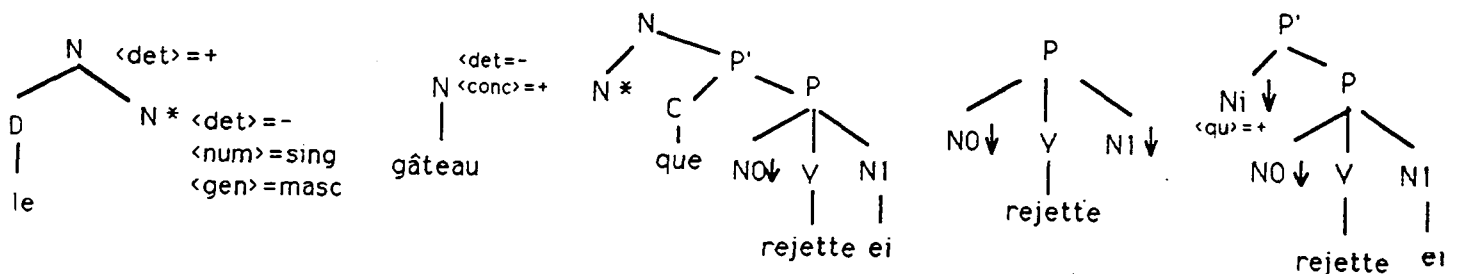


Figure 3

La première structure de *rejette* est forcément sélectionnée, car c'est la seule qui intègre à la grammaire le mot *que*. Les deux dernières structures sont donc automatiquement exclues.

Pour *l'idée que Marie rejette* (sans rien après *rejette*), trois structures concurrentes seront sélectionnées pour *rejette* (la relative en *que* N_0 *rejette*, la question et N_0 *rejette* N_1). Les deux dernières seront éliminées car aucun arbre ne peut être substitué en N_i ou en N_1 .

Un autre exemple, de G. Sabah 1990, est celui de l'attachement de la relative dans:

J'ai vu le cahier de l'étudiante que Jean a examiné(e)

Selon, la flexion du participe passé final, la relative doit s'attacher à *cahier* ou à *étudiante* et le verbe *examiner* s'interprétera différemment dans les deux cas (regarder attentivement ou faire passer un examen). Il faudrait donc "attendre" la fin de la phrase pour pouvoir attacher la relative et construire le groupe nominal. Pour notre grammaire, le noeud pied d'attachement de la relative est directement accessible à partir du participe passé (et vice-versa): l'on pourra vérifier les contraintes d'accord et les contraintes sémantiques en même temps, et l'on attachera directement la relative au

bon noeud.

De nombreux systèmes mettent en place des techniques particulières de regard en avant (Cf Parsifal, ANDI). C'est pour nous une composante directement fournie par la lexicalisation de la grammaire. Toutes doivent placer une borne supérieure au regard en avant (3 mots ou structures pour l'analyseur de M. Marcus comme pour celui de G. Sabah et M. Rady 1983). Or l'on peut toujours imaginer une phrase "à cul de sac" (*garden path*) dont l'analyse repose crucialement sur un mot situé au delà de cette limite. Ainsi il n'y a pas de limite à l'empilement des groupes prépositionnels dans la phrase ci-dessus:

J'ai vu les cahiers des étudiantes du lycée de la ville que Jean a examiné(e)(s)

Un exemple formel, que nous devons à une discussion avec M. Marcus, montre que la lexicalisation peut aider là où des techniques de regard en avant seraient inefficaces car la distance où se trouve l'information déterminante n'est pas bornée a priori. L'exemple est le suivant:

S -> A
S -> B
A -> a A
A -> a x
B -> a B
B -> a y

Cette grammaire génère des suites : *aaaax* ou *aaay*, c'est-à-dire un nombre quelconque de *a* suivis de *x* ou de *y*. Seul le dernier mot (*x* ou *y*) permet d'analyser la phrase (un emboîtement de A ou de B). Il faudrait donc un regard en avant illimité, ce qui n'est pas possible.

Avec la grammaire lexicalisée correspondante, le problème ne se pose plus:

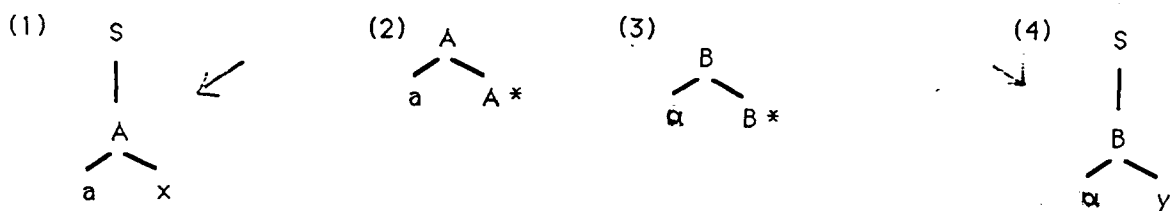


Figure 4

La première passe permet de sélectionner les bons arbres élémentaires: 1, 2 et 3 (si la phrase se termine par *x*), 2 et 3 et 4 (si la phrase se termine par *y*). La deuxième passe construit la bonne analyse et élimine 3 (dans le premier cas) et 2 (dans le second).

Le "regard en avant" fourni par la lexicalisation de la grammaire n'est pas a priori borné en distance, ce qui permet un gain en efficacité et en fiabilité de l'analyse.

2.4. L'analyse des expressions figées

Comme l'ont montré A. Abeillé et Y. Schabes 1989, les expressions figées ne nécessitent aucun mécanisme particulier dans notre système et sont analysées exactement suivant le même système que les expressions libres. Elles fournissent un bon exemple de l'intérêt des arbres lexicalisés à têtes multiples. De nombreuses expressions ont en effet un mot en commun (souvent le verbe); ainsi les 20 000 expressions recensées par M. Gross 1989 correspondent à moins de 4000 verbes (sur les 8 000 verbes du français). La prise en compte dans une grammaire des expressions figées augmente donc considérablement le nombre de structures possibles par verbe, c'est-à-dire le taux d'ambiguïté lexicale. M. Gross 1982 cite l'exemple du verbe *manger* qui n'a pas moins de 50 constructions (libres ou figées) différentes. Il est donc nécessaire d'éliminer au plus vite bon nombre de ces constructions.

Ces expressions posent problème aux techniques classiques de regard en avant, car elles peuvent présenter entre leurs différents éléments figés des discontinuités non bornées (Cf chapitre 4):

*La moutarde semblait vraiment être sur le point de bientôt monter au nez de Paul.
C'est sur le dos de Paul que tout le monde s'accorde à avouer que Marie disait que Jean avait bien cassé du sucre.*

Il faudrait donc un regard en avant de plus de trois éléments (et dans le cas général non borné) pour exclure l'interprétation idiomatique de certaines phrases, par exemple⁵:

*La moutarde semblait vraiment être sur le point de bientôt arriver à Riga.
C'est sur le dos de Paul que tout le monde s'accorde à avouer que Marie disait que Jean avait bien cassé du sucre de canne.*

Le problème de l'analyse des constituants discontinus n'a reçu jusqu'à présent que des solutions partielles et conduit à adopter des stratégies particulières (Cf M. Johnson 1985).

Dans notre représentation, la reconnaissance des expressions figées, continues ou discontinues, est efficace et ne réclame pas de mécanisme particulier.

L'analyse se fait en deux passes comme pour les arbres à tête lexicales uniques. L'analyseur ne sélectionne un arbre idiomatique élémentaire que si:

- toutes les têtes lexicales obligatoires sont présentes dans la phrase
- elles sont présentes dans le même ordre que dans l'arbre élémentaire
- les traits attachés à l'arbre sont compatibles avec ceux de l'expression idiomatique dans le dictionnaire (ex: passif = + /-)

⁵ L'échec ou la réussite d'une analyse déterministe des phrases suivantes avec un regard en avant de trois éléments dépend en fait des constituants définis dans la grammaire.

Les arbres non idiomatiques correspondant aux mêmes têtes lexicales sont également sélectionnés s'il y en a. Prenons par exemple la phrase 0 Jean 1 prend 2 une 3 veste 4, les arbres suivants sont sélectionnés (notamment) :

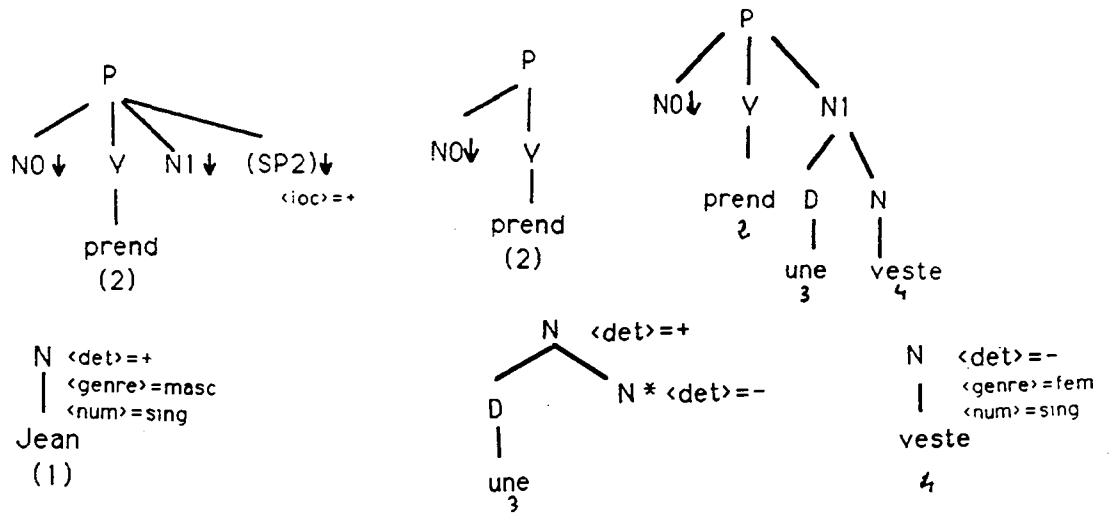


Figure 5

On sélectionne des arbres qui seront ensuite éliminés à l'analyse, par exemple l'arbre de *N₀ prendre* (:= *la mayonnaise prend*), ainsi ceux de *N₀ prendre N₁ à N₂* car le complément prépositionnel y est optionnel. Mais on ne sélectionne pas l'arbre de *N₀ prendre N₁ pour N₂*, car *pour* ne figure pas dans la phrase. On ne sélectionne pas non plus les arbres d'autres expressions avec *prendre* telles : *N₀ prendre N₁ en compte* (car il n'y a pas *en* ni *compte*), *N₀ prendre la fuite* (il n'y a pas *fuite*) ... On élimine donc dès la première passe toutes les expressions qui ne peuvent figurer dans la phrase. En revanche, on sélectionne des expressions dont l'analyse révélera ensuite qu'elles ne s'y trouvaient pas. Par exemple, *N₀ prendre une veste* à partir de la phrase: *Jean prend un peignoir au lieu d'une veste*.

La condition sur l'ordre dans lequel apparaissent dans la phrase les têtes lexicales est importante. Par exemple, les arbres de la famille de *N₀ briser la glace* sont notamment:

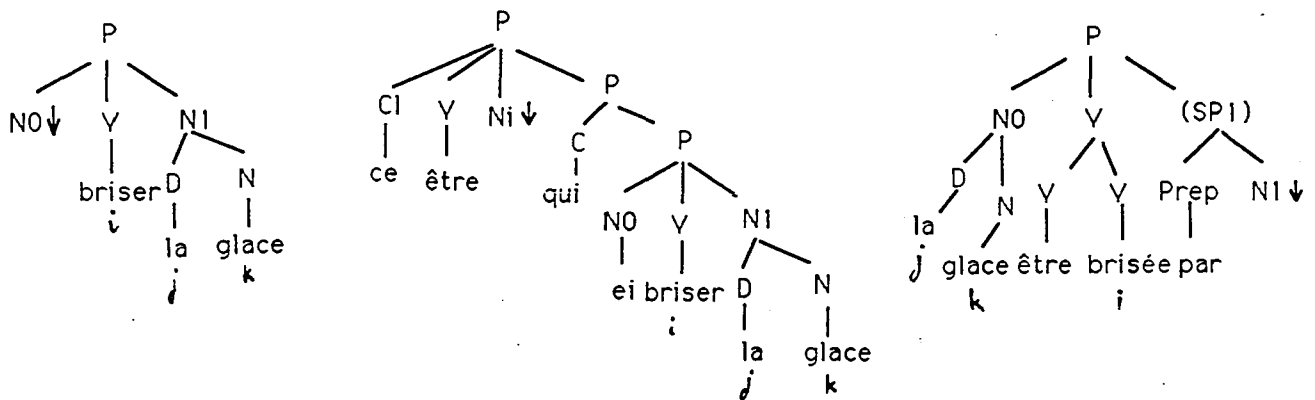


Figure 6

Les arbres idiomatiques imposent ici les ordres suivants entre *briser* (i), *la* (j) et *glace* (k): $i < j < k$ ou $j < k < i$. Les autres ordres, par exemple: $k < i < j$ ou $j < k < i$ sont exclus. Les phrases suivantes permettront donc la sélection des arbres idiomatiques:

- (1) *Jean brise du pain sur la glace* ($i < j < k$)
- (2) *Jean mange de la glace dans une assiette brisée* ($j < k < i$)

En revanche, l'interprétation idiomatique des phrases suivantes sera exclue d'emblée :

- (3) *Jean mange une glace dans la brise* ($k < j < i$)
- (4) *Je sais quelle glace Jean a brisée dans la cuisine* ($k < i < j$)

Pour les phrases (1) et (2), c'est l'analyse proprement dite (deuxième passe) qui éliminera l'arbre de N_0 *briser la glace* et exclura l'interprétation idiomatique.

Dans notre système, l'analyse des constituants discontinus suit exactement les mêmes principes que celle des constituants continus et aucune complication particulière n'est nécessaire. De même pour expressions figées et expressions libres

2.5. Avantage pour l'unification des structures de traits

L'ajout de structures de traits et d'un mécanisme d'unification ne rend pas la capacité générative des grammaires TAG plus forte, à condition, comme pour les grammaires hors contexte, d'interdire les structures de traits de taille non bornée (K. Vijay-Shanker 1987). Il ne rend pas non plus l'analyse des TAG plus complexe. L'analyseur vérifie les contraintes d'unification locales dans la première passe lorsqu'il attache les têtes lexicales fléchies aux préterminaux, et dans la deuxième passe lorsqu'il cherche à combiner entre eux les arbres élémentaires.

La lexicalisation de la grammaire (et ses structures arborescentes étendues) permet de simplifier considérablement le système des traits associés aux catégories dans le lexique et dans les règles syntaxiques. Ainsi nous ne traitons pas par des traits un certain nombre de phénomènes pour lesquels des grammaires de réécriture hors-contexte seraient obligées d'utiliser des contraintes d'unification (ex: sous-catégorisation, expressions figées ...)

De plus, les unités étendues de notre grammaire, et les propriétés particulières de l'adjonction, permettent de minimiser les cas de propagation de traits sur des longueurs non bornées. Ainsi nous évitons ce cas pour les dépendances non bornées (chapitre 5) et pour la reconnaissance des expressions figées (chapitre 4). Dans le cas des expressions figées, nous pouvons vérifier directement des contraintes d'accord entre éléments situés à des niveaux syntagmatiques différents, comme par exemple pour les expressions N_0 *vider Poss_0* *sac*, N_0 *casser Poss_0* *pipe*. Le trait *qu-* en revanche est un exemple de propagation non bornée, comme pour les grammaires hors contexte, puisqu'il se propage de n'importe quel complément enchassé au noeud N dominant:

*quelle fille
la fille de quel voisin
la fille du père de quel voisin*

Mais la lexicalisation de la grammaire apporte un deuxième avantage en ce qu'elle élimine les possibilités de récursion sur les structures de traits.

En particulier, Y. Schabes et A. Joshi 1990 ont montré que les problèmes d'indécidabilité dûs aux structures de traits récursives disparaissaient.

2.6. Techniques d'optimisation

Nous en sommes encore à une phase d'expérimentation. Nous n'avons rien fait pour optimiser l'efficacité de l'analyseur. Néanmoins, différentes techniques sont envisageables, qui prennent appui sur les particularités de la grammaire que se constitue l'analyseur dans la première passe.

Ainsi, il est possible d'optimiser l'utilisation des informations disponibles dès la première passe, afin de soulager encore davantage l'analyse syntaxique proprement dite de la deuxième passe. Diverses techniques de comptage, analogues à celles utilisées en grammaire catégorielle pour comparer le nombre de foncteurs et d'arguments, sont envisageables pour éliminer encore des arbres avant l'analyse proprement dite.

Rappelons certaines particularités de la grammaire chargée dans la première passe, que nous appelons la grammaire du parseur. Toute tête lexicale doit être utilisée une fois et une seule: une fois qu'un arbre est utilisé pour une tête lexicale donnée, les autres arbres possibles pour la même tête deviennent inactifs (ainsi un arbre ne peut se combiner avec lui-même).

Rappelons aussi que les combinaisons entre arbres ne créent pas de catégories supplémentaires. L'on peut donc connaître au départ le nombre et la catégorie des noeuds de l'arbre dérivé final.

L'on définit ainsi quelles sont, parmi les ensembles d'arbres élémentaires correspondant aux structures de la première passe, les forêts bien formées et l'on élimine les autres. L'on peut ainsi exclure certaines phrases incorrectes sans même avoir à les analyser, mais surtout éliminer des structures élémentaires (c'est-à-dire des ambiguïtés potentielles).

a) Une forêt bien formée doit comporter au moins un arbre initial de racine phrastique. L'on n'a donc pas besoin d'analyser la séquence: *Jean un garçon*. L'on sait dès la première passe que ce ne peut être une phrase correcte, puisqu'on ne peut sélectionner d'arbre phrastique à partir de ces trois têtes lexicales.

Reprenons également la phrase: *Jean aime Marie*. L'arbre correspondant à *N0 aimer Pinf* ne peut faire partie d'une forêt bien formée puisque c'est un arbre auxiliaire et que *Jean* et *Marie* ne sélectionnent pas d'arbres phrastiques. C'est donc l'arbre correspondant à *N0 aimer N1* qui doit être sélectionné pour *aimer*.

Dans cet exemple, l'analyseur éliminera aisément l'arbre à complément phrastique lors de l'analyse proprement dite, mais il peut être intéressant d'éliminer ainsi des arbres avant toute analyse dans le cas général.

b) Une forêt bien formée doit également comprendre plus d'arbres élémentaires que de noeuds correspondant à des arguments obligatoires (feuilles marquées pour la substitution). Prenons par exemple la phrase:

Jean travaille pour Marie

Les arbres suivants sont a priori possibles:

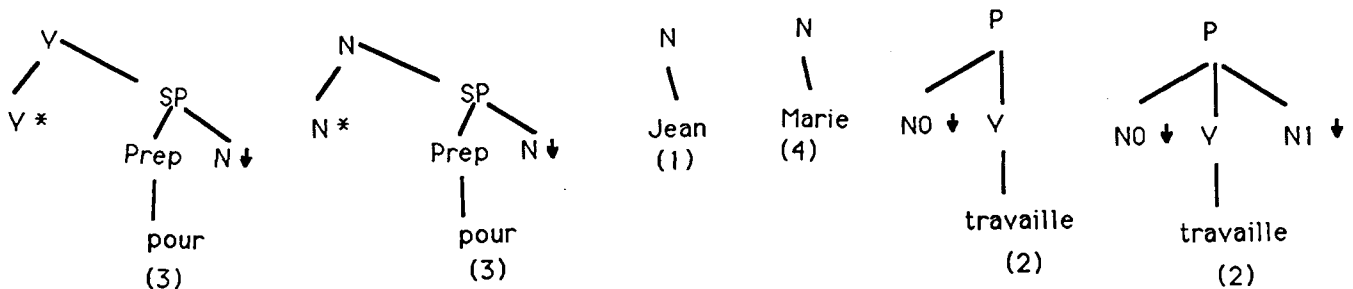


Figure 7

Ce sont forcément les arbres de *travailler* intransitif, non ceux de *travailler* transitif (Cf *Jean travaille le bois*) qui conviennent, sinon il manque un argument N.

Suivant le même principe, une phrase comme *Jean risque pour Marie* sera reconnue comme incorrecte dès la première passe, puisque *risquer* ne sélectionne que des structures transitives.

c) Une forêt bien formée doit avoir autant d'arbres initiaux dont la racine est X que de noeuds de catégorie X correspondant à des arguments obligatoires (feuilles marquées pour la substitution), et ce pour chaque catégorie (sauf P). Ainsi, pour reprendre l'exemple supra de *Jean prend une veste*, l'arbre intransitif de *prendre* peut être éliminé d'emblée, car il y a dans la grammaire deux arbres initiaux de racine N à substituer:

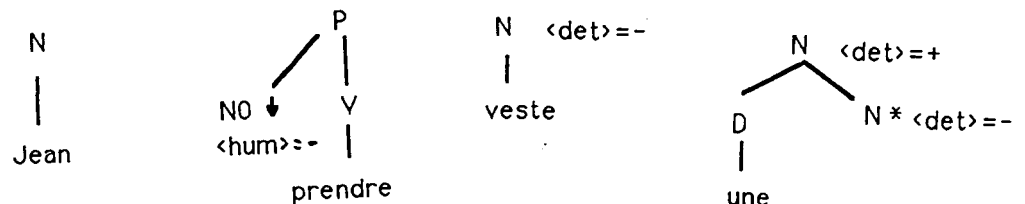


Figure 8

Puisqu'on doit obtenir in fine un arbre initial de racine P, il faut un arbre initial

phrastique de plus qu'il y a de noeuds P à substituer dans la grammaire.

d) Une forêt bien formée comprenant n arbres auxiliaires de racine X doit comprendre au moins $2n + 1$ noeuds de catégorie X, et ce pour chaque catégorie. Ainsi si la grammaire comporte un arbre auxiliaire de catégorie N, elle doit au total comporter au moins 3 noeuds N (c'est-à-dire le noeud racine et le noeud pied de l'arbre auxiliaire plus un noeud où il s'adjoit).

Prenons par exemple la phrase:

Paul cueille une fleur très rouge.

La tête lexicale *très* sélectionne a priori 1 et 2:

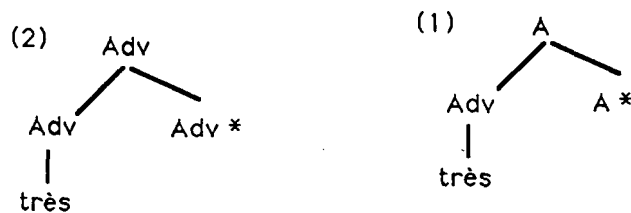


Figure 9

On élimine forcément l'arbre 2, car il n'y a pas de noeud Adv où l'adjoindre ailleurs dans la grammaire de la phrase:

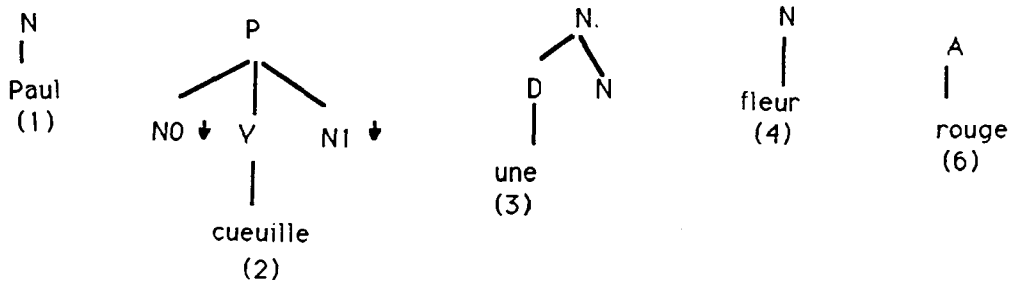


Figure 10

De même pour la phrase: *Paul mange une poire très ferme*. Parmi les arbres élémentaires possibles, on a:

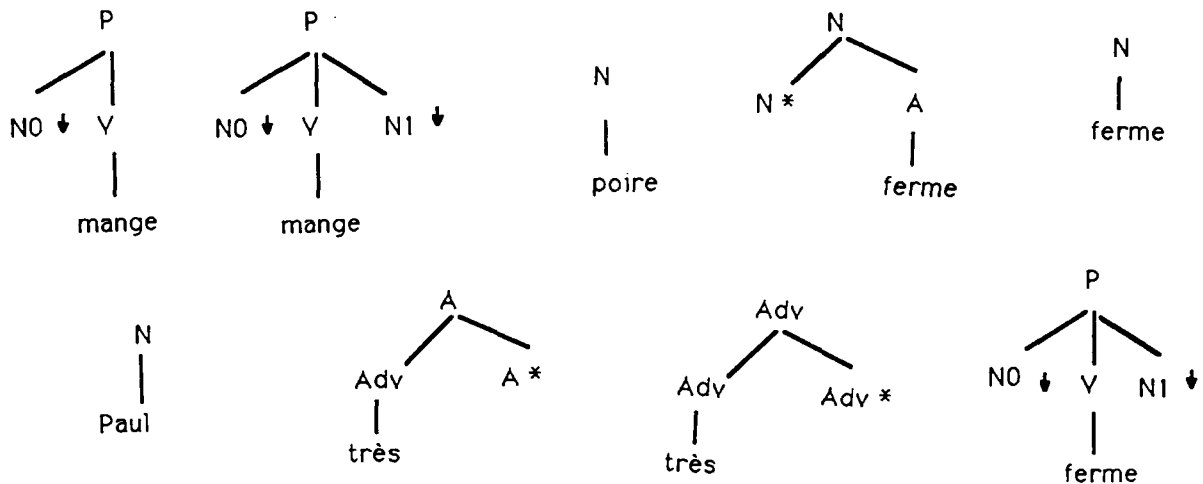


Figure 11

L'adverbe *très* doit être modifieur d'adjectif pour la même raison que supra, *ferme* doit donc être adjectif et non verbe ou nom car sinon il n'y aurait pas de noeud Adj où adjoindre *très*.

Nous avons mentionné seulement quelques exemples de techniques possibles pour réduire encore la taille de la grammaire de l'analyseur dès la première passe et améliorer l'efficacité du traitement dans la deuxième phase. Bien d'autres sont possibles. Mais il faut éviter que ce prétraitement de la première passe ne fasse perdre plus de temps, en anticipant sur l'analyse véritable, que ce qu'il permettra d'économiser dans la deuxième passe. L'essentiel des gains réalisés devraient provenir de la lexicalisation de la grammaire, et non de ces techniques d'optimisation supplémentaires.

On a montré que la lexicalisation de la grammaire permettait de définir une stratégie d'analyse originale, qui garantit que l'analyseur ne travaille qu'avec un sous-ensemble pertinent de la grammaire du français, pour chaque phrase à analyser. Elle permet également de résoudre ou de simplifier des problèmes classiques en analyse de grammaires hors contexte. Le gain est cependant difficile à évaluer de façon générale et dépend fortement du type de grammaire utilisé. Il devrait être important pour une grosse grammaire où les unités à têtes multiples sont importantes.

Le fait de pouvoir réduire le rôle de l'unification et d'utiliser des traits locaux là où une grammaire hors contexte devrait propager des traits sur des longueurs non bornées (Cf chapitres 4 et 5) est également un avantage.

Il convient toutefois de garder à l'esprit que l'adjonction étant plus difficile à reconnaître que la substitution, l'analyse d'une grammaire TAG peut être plus coûteuse que l'analyse d'une grammaire hors contexte (notamment si on travaille avec une petite grammaire).

3 La question des ambiguïtés

3.1. Discussion théorique

L'on peut distinguer deux conceptions de l'ambiguïté syntaxique d'une phrase. Selon la première, minimale, une phrase est autant de fois ambiguë que peuvent lui être associés d'arbres syntagmatiques différents. C'est la conception la plus répandue. Selon la deuxième, qui est plus large, une phrase est autant de fois ambiguë que lui correspondent des décompositions en phrases simples différentes. Cette dernière conception est représentée notamment dans les analyseurs de N. Sager pour l'anglais, de J-H. Jayez pour le français.

Cette deuxième approche de l'ambiguïté est plus ambitieuse, puisqu'elle remet largement en cause la frontière entre analyse syntaxique et analyse sémantique. Elle inclut un certain nombre d'ambiguïtés souvent considérées comme sémantiques, et notamment la plupart des phénomènes de contrôle:

Jean a proposé à Luc de venir
 = 1) *Jean viendra* + *Jean l'a proposé à Luc*
 = 2) *Luc viendra* + *Jean le lui a proposé*

Jean a quitté Luc content
 = 1) *Jean a quitté Luc* + *Luc était content*
 = 2) *Jean a quitté Luc* + *Jean était content*

Elle inclut aussi les ambiguïtés entre interprétation littérale et interprétation idiomatique, qui n'ont pas la même décomposition:

Jean a pris une veste
 = 1) *Jean est quelqu'un* + *Quelqu'un a pris une veste*
 = 2) *Jean est quelqu'un* + *Une veste est quelque chose* + *Quelqu'un a pris quelque chose*

Mais elle n'inclut pas tous les phénomènes de portée. En effet une phrase ambiguë comme:

Tous les hommes aiment une femme

où il peut s'agir d'une seule femme ou de plusieurs femmes différentes, ne donne pas lieu à deux décompositions de ce type.

Notre grammaire TAG peut être utilisée de deux façons: si l'on ne considère que les arbres dérivés construits par l'analyseur, on obtient l'ambiguïté d'une phrase au sens 1. Si l'on considère les arbres de dérivation associés, on est plus proche de l'ambiguïté au sens 2. En effet, la décomposition en arbres élémentaires que fournit un arbre de

dérivation est analogue à une décomposition en phrases simples, si l'on définit au moins un arbre phrastique dans la famille de toute catégorie.

Toutefois, nos arbres de dérivations actuels ne permettent pas toutes les décompositions souhaitées. Notre grammaire actuelle ne rend pas compte de tous les phénomènes de contrôle, ni des réductions par coordination.

3.2. Problèmes spécifiques à l'analyse automatique

Les phénomènes d'ambiguïtés sont parmi les problèmes les plus difficiles pour un analyseur automatique, qui a en général du mal à éliminer toutes les règles possibles non pertinentes pour une phrase donnée. L'on parle alors d'ambiguïtés artificielles, ou de surgénération de l'analyseur. Suivant les stratégies d'analyse et suivant les formalismes grammaticaux utilisés, le problème est plus ou moins important. Aucun système actuel ne le résout parfaitement.

L'on distingue classiquement entre ambiguïtés lexicales, syntaxiques et sémantiques. En TAG, cette distinction perd une partie de son sens. La lexicalisation de la grammaire TAG permet de ramener bon nombre d'ambiguïtés syntaxiques ou sémantiques à des cas d'ambiguïtés lexicales, qui sont plus simples à lever.

L'on considère ainsi l'ambiguïté entre interprétation littérale et interprétation idiomatique comme une ambiguïté lexicale, puisque nous avons des entrées lexicales différentes dans notre dictionnaire. Elle se manifeste par l'obtention de plusieurs arbres de dérivations.

Par exemple *Jean casse sa pipe* sera décomposé soit comme: *Jean* + N_0 *casser sa pipe*, soit comme *Jean* + N_0 *casser* N_1 + *sa* + *pipe*.

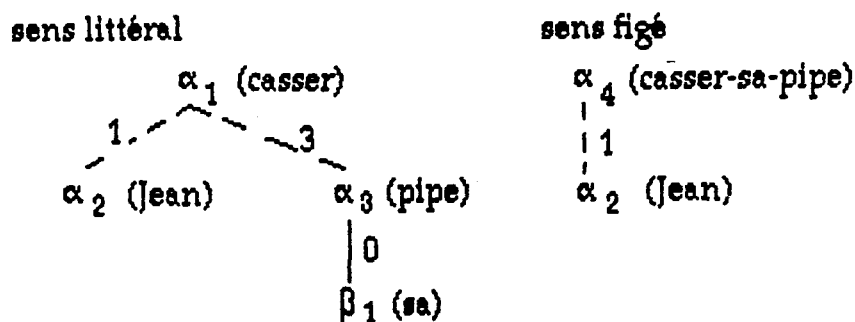


Figure 12

Un autre exemple est celui de l'ambiguïté entre adverbes et compléments. Elle se manifeste par l'obtention de deux arbres dérivés et de deux arbres de dérivation.

Par exemple, la phrase *Jean regarde la nuit* sera analysée de deux façons selon que *la nuit* est le complément de *regarder* ou un adverbial. Ces deux analyses correspondent à deux catégories différentes du mot *nuit*: dans la première, il est tête d'un arbre initial N, dans la seconde d'un arbre auxiliaire (adverbe figé : tête = *nuit* et déterminant = *cette/une/la*):

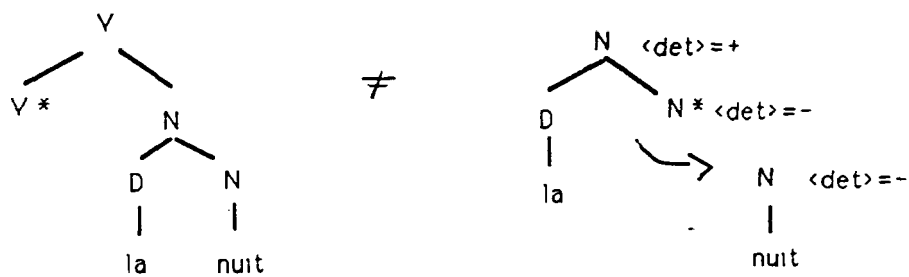
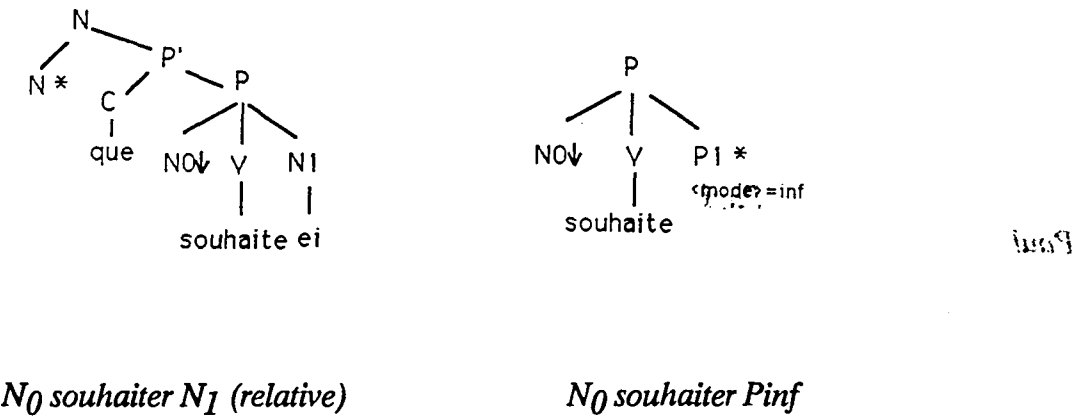


Figure 13

Les ambiguïtés entre adverbes et compléments se voient pour nous dans l'arbre dérivé et dans l'arbre de dérivation. Elles sont souvent surgénérées par les analyseurs qui les considèrent comme purement "syntaxiques", alors que les restrictions lexicales sont ici essentielles (peu de noms sont des adverbiaux de type Det N).

Un autre exemple d'ambiguïté qui est pour nous lexicale est celle entre relatives et complétives. Une phrase comme *Il faut malgré ces documents que l'on souhaite faire une enquête* est ambiguë selon que la tête de la complétive est *souhaite* (*il faut ... que l'on souhaite ...*) ou *faire* (*il faut ... faire ...*). Elle est pour nous lexicale ambiguë car le verbe *souhaiter* n'a pas la même structure dans les deux interprétations. Dans la première, il prend une complétive à l'infinitif, dans la seconde il a un complément nominal et est tête d'une relative :



N₀ souhaiter N₁ (relative)

N₀ souhaiter P_{inf}

Figure 14

Certains attachements prépositionnels sont également analysés comme des ambiguïtés lexicales. Par exemple: *Jean voit la fille avec un télescope* (selon que *avec* est tête d'un arbre auxiliaire verbal ou nominal). De même, *Jean a répondu intelligemment* (selon que l'adverbe est tête d'un arbre auxiliaire de racine verbale ou phrastique: *Jean a répondu de manière intelligente* ou *Jean a eu l'intelligence de répondre quelque chose*).

Il est relativement aisé d'éliminer sur une base lexicale certains attachements prépositionnels. Ainsi *avec* tête d'un N abstrait sans déterminant ne peut modifier qu'un verbe (ou une phrase) mais pas un nom. Sa structure sera donc celle d'un arbre

auxiliaire de racine V et l'hypothèse d'un attachement au nom ne sera jamais faite.

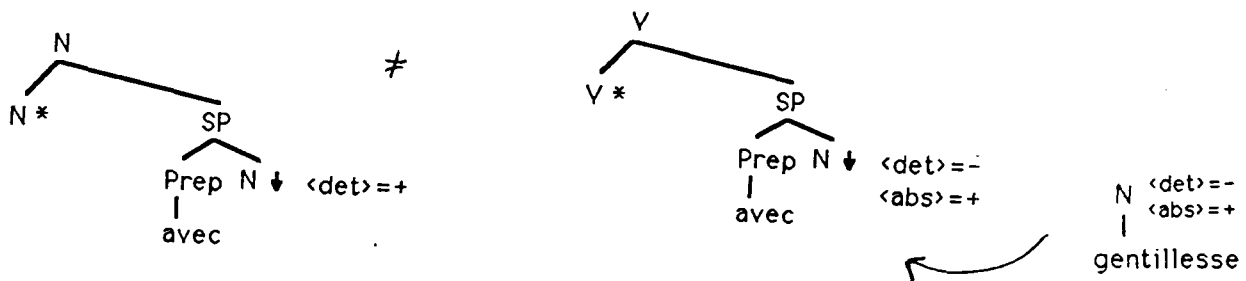


Figure 15

- N₀ répondre avec (mépris + arrogance + gentillesse..)*
** Une réponse avec (mépris + arrogance + gentillesse..)*

En revanche, les phrases suivantes sont des cas d'ambiguïtés non lexicales:

- (1) *Jean a dit que Marie était partie vraisemblablement*
 (2) *Tu vois les cahiers du cartable que je prends*

En (1), *vraisemblablement* peut s'attacher au noeud P de l'arbre de *dire* ou à celui de l'arbre de *partir*. En (2), la relative peut s'attacher à *cahier* ou à *cartables*. Dans les deux cas l'arbre auxiliaire est le même.

Pour les ambiguïtés lexicales comme non lexicales, notre grammaire permet un certain nombre d'ambiguïtés artificielles. Elles sont dues notamment au fait que nos restrictions de sélection ne sont pas assez fines. Par exemple, la phrase suivante sera analysée comme ambiguë : *Jean pense malgré ces éléments que Paul saura faire le travail*, alors que l'analyse à complétive (*Jean pense que P*) est beaucoup plus naturelle que celle à relative (*Jean pense faire W, malgré les éléments que Paul saura*). Il faudrait donc soit ajouter des sous-traités supplémentaires, pour exclure *élément* des objets possibles de *savoir* (ce qui est en fait une tâche infinie; Cf A. Guillet 1990), soit utiliser des heuristiques pour privilégier la construction la plus probable: ici *N₀ savoir V_{inf} W*.

Un intérêt de notre grammaire lexicalisée est, dans cette optique, de pouvoir s'appuyer sur des probabilités "lexicales" plus fines et plus précises que les habituelles probabilités catégorielles ou d'attachement. Ainsi, pour le verbe *savoir*, les structures *N₀ V N₁* seront moins fréquentes que les structures *N₀ V que P* ou *N₀ V V_{inf} W*; pour un verbe comme *aimer*, ce sera le contraire. De même, il semble y avoir de grandes variations lexicales sur les préférences d'attachement des groupes prépositionnels (en ne considérant que les adverbiaux, et en excluant les compléments et les groupes figés). Le modifieur *selon* *N* semble s'attacher préférentiellement à un verbe, et le modifieur *de* *Det N* préférentiellement à un nom. L'on peut ainsi affecter des probabilités relatives à nos arbres élémentaires (pour une tête lexicale donnée) et faire l'analyse d'abord avec les plus probables (reprenre les moins probables si la première tentative ne donne rien).

C'est une direction de recherche que nous souhaitons explorer. Il reste à calculer de façon fiable de telles probabilités pour un domaine d'application donné. Ainsi, la probabilité d'occurrence de certains mots en structures libres peut être très inférieure (dans un contexte donné) à leur probabilité d'occurrence au sein d'une expression figée. Un mot comme *vessie* aura, en tant que nom simple, une probabilité quasiment nulle dans les pages politiques ou culturelles du *Monde*, par exemple, mais elle y apparaît régulièrement dans l'expression *NO prendre des vessies pour des lanternes*. L'on pourra donc associer aux arbres élémentaires respectifs une probabilité fort différente. La difficulté est l'énorme masse de calculs à effectuer sur des textes préanalysés pour établir au préalable ces probabilités.

4 Implémentation actuelle

Rappelons que nous nous plaçons dans le cadre de l'analyse de phrases isolées, et non de l'analyse de textes. Nous omettons également des problèmes fréquemment rencontrés sur phrases réelles: séparateurs, mots inconnus, nombres, noms propres. Une préanalyse lexicale a été définie par M. Silberstein 1989; pour les mots inconnus voir également J. Véronis 1988⁶. Nous utilisons un dictionnaire de formes fléchies, mais l'on pourrait aussi, pour une application donnée, préférer une première étape d'analyse morphologique proprement dite.

La grammaire décrite dans les chapitres précédents a été implémentée pour un lexique correspondant aux 2500 mots les plus fréquents du français. 4000 entrées lexicales ont été définies (qui correspondent à des arbres simples ou à des familles d'arbres).

Le dictionnaire de formes fléchies associé a été généré automatiquement (programme Lisp indépendant) et comprend plus de 10 000 formes.

Il existe une interface élaborée pour l'anglais qui permet une extension aisée du dictionnaire syntaxique (K. Bishop 1989; version Symbolics). Une interface du même type est en cours d'écriture pour le français (version Sun).

Les fichiers d'arbres comprennent actuellement 80 "familles" différentes, avec une moyenne d'une quinzaine d'arbres élémentaires par famille, plus quelques centaines d'arbres simples (prépositions, déterminants, clitics...). Ils comportent au total environ 1500 arbres élémentaires différents.

Cette grammaire a été testée sur des phrases simples avec un analyseur réalisé à l'université de Pennsylvanie.

L'algorithme d'analyse utilisé dans la seconde passe est relativement indépendant de la grammaire: il peut être montant, descendant, avec retour arrière, parallèle, déterministe etc.... Nous utilisons un analyseur de type Earley qui est un analyseur mixte (montant avec une composante prédictive), et qui procède de gauche à droite. Au lieu de règles à points (*dotted rules*), il utilise des arbres à points qui séparent à chaque noeud l'arbre en une partie gauche (reconnue) et une partie droite (à reconnaître). Il est décrit en détail

⁶ Le seul prétraitement que nous utilisons est pour les enclises: *au* = *à le...*

dans Y. Schabes et A. Joshi 1988, et Y. Schabes 1990. Il utilise la stratégie en deux temps décrite ci-dessus et unifie les structures de traits. Il est le seul, parmi les algorithmes d'analyse définis pour les TAG⁷, à incorporer toutes les extensions du formalisme que nous utilisons (substitution, lexicalisation, unification; Y. Schabes et A. Joshi 1990). Il tire pleinement avantage de la lexicalisation de la grammaire.

La complexité de cet algorithme est la suivante, G étant la taille de la grammaire et n la longueur de la phrase à analyser (Y. Schabes 1990) :

	temps	espace
Grammaires ambiguës:		
pire des cas	G^2n^6	Gn^4
meilleur cas	Gn	Gn
Grammaires non ambiguës:		
pire des cas	G^2n^4	Gn^4
meilleur cas	Gn	Gn

Le programme actuel est écrit en Common Lisp. Une version existe sur machine LISP (Symbolics), une autre est en cours d'écriture sur SUN.

Les fichiers d'arbres et les dictionnaires sont indépendants du langage et de l'algorithme d'analyse. Les fichiers d'arbres sont lus et édités par un éditeur d'arbres spécialement conçu pour le formalisme. Cet éditeur existe en deux versions, l'une sur machine LISP Symbolics (A. Chalik), l'autre sur Sun (P. Paroubek).

Les fichiers d'arbres sont séparés des fichiers dictionnaires au départ, mais pourraient être fusionnés avant analyse si la grammaire était totalement précompilée. Dans la version actuelle, la compilation des arbres et celle du dictionnaire se fait indépendamment (avant analyse).

Le temps moyen d'analyse d'une phrase, testé sur des phrases de 10 à 30 mots (sur machine Lisp), est inférieur à 30 secondes.

L'analyse des TAG synchrones n'est pas encore implémentée, mais est en cours d'implémentation. L'analyseur actuel utilise donc les anciennes familles (par mêmes noms de noeuds et non par des liens; Cf chapitre 3).

Voir annexe 1 pour les symboles et traits employés, annexe 2 pour les différents arbres élémentaires, annexe 3 pour des extraits des dictionnaires.

⁷ D'autres algorithmes ont été définis pour les grammaires d'arbres adjoints, souvent inspirés des analyseurs existant pour les grammaires hors contexte: un analyseur montant de type CKY (K. Vijay-Shanker 1987), un analyseur parallèle (M. Palis et S. Shende 1990) et quelques autres (B. Lang 1990, K. Harbush 1990). Aucun n'a été testé avec notre grammaire.

5. Autres applications

Elles sont bien entendu multiples, puisque la plupart des applications de traitement automatique du langage ont besoin d'un module d'analyse syntaxique.

Nous n'en mentionnons que quelques unes, généralement en cours de réalisation, afin de mettre l'accent sur ce que notre grammaire, pas seulement par sa couverture mais aussi par le type de représentations qu'elle met en oeuvre, peut apporter de spécifique.

5.1. Génération

L'idée d'utiliser les mêmes grammaires pour l'analyse et la génération est assez récente (Cf M. Dymetman, P. Isabelle 1988; S. Shieber 1988), même si le point de vue théorique et psycholinguistique qui la sous-tend est bien plus ancien.

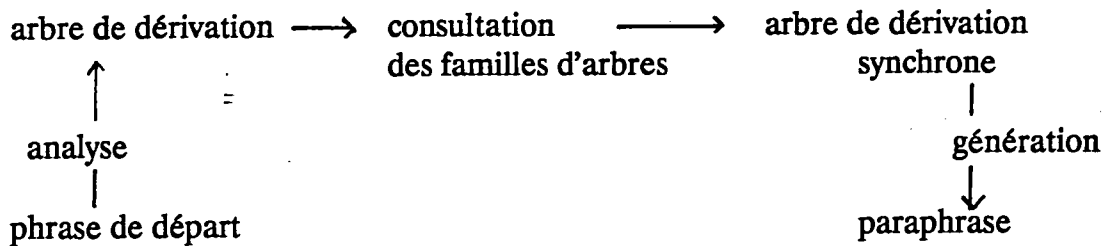
Or la construction de grammaires "réversibles" se heurte en général à la non correspondance entre unités syntaxiques (point d'arrivée de l'analyse) et unités sémantiques (point de départ de la génération). Ainsi, les expressions figées correspondent à une unité (à un concept) en génération, alors que la plupart des analyses décomposeront celles qui sont syntaxiquement régulières en autant de syntagmes indépendants. De même, les mots vides sémantiquement (complémenteurs, prépositions marqueurs de complément) font partie du même concept que le verbe pour la génération, alors que pour l'analyse il faut en faire des catégories distinctes.

Notre grammaire semble particulièrement bien placée pour éviter ces problèmes puisque nos unités sont naturelles sémantiquement, tout en ayant une structure syntagmatique interne régulière. L'on peut ainsi générer d'un bloc les expressions figées et phrases à verbe support, qui correspondent à des arbres élémentaires, de même pour *penser + que* ou *donner + à*, sans pour autant considérer ces suites de mots comme des syntagmes (ce qu'ils ne sont pas).

La génération a bien entendu ses contraintes propres (Cf L. Danlos 1989), et notre grammaire ne peut être qu'un point de départ, auquel il est nécessaire d'ajouter d'autres contraintes.

5.2. Paraphrase automatique

Les TAG synchrones utilisées pour la constitution de familles d'arbres permettent d'exprimer des équivalences entre arbres de dérivation. Associées à un module de génération, elles permettent d'envisager aisément une application de paraphrase automatique, sur le principe suivant:



Ainsi, à partir de la phrase: *Quelle nouvelle femme Jean aime-t-il ?*, on génèrera par exemple:

Jean aime quelle nouvelle femme ?
Quelle nouvelle femme est aimée par Jean ?
C'est quelle nouvelle femme que Jean aime ?

En effet, les arbres élémentaires correspondants sont tous reliés dans la famille du verbe *aimer* et leurs arguments sont les mêmes. De même pour la phrase plus complexe *Que Jean ait pris une veste surprend Marie* qui, dans son interprétation idiomatique, pourra être paraphrasée par:

= *La veste que Jean a prise surprend Marie.*
 = *Marie est surprise par la veste que Jean a prise*

Le même arbre élémentaire est utilisé pour (*N0 + Que P0*) *surprendre N1*, et les deux arbres idiomatiques sont liés dans la famille de *N0 prendre une veste* idiomatique.

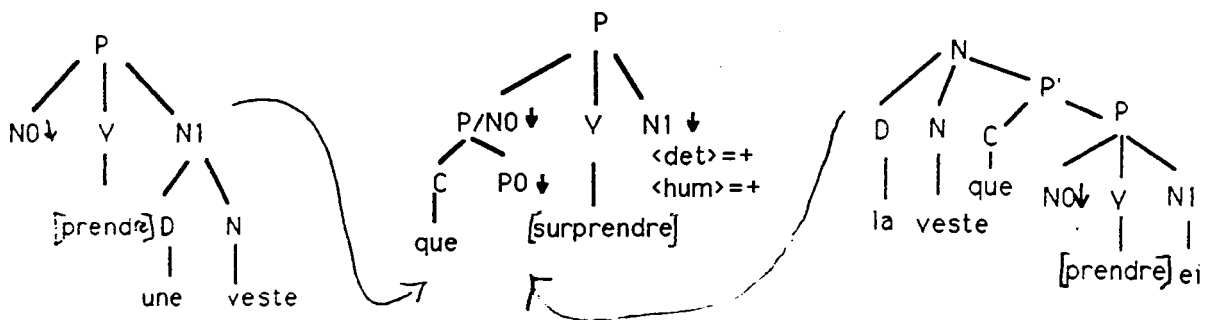


Figure 16

Il est possible d'étendre cette application par le couplage systématique des arbres élémentaires potentiellement synonymes de notre grammaire, par exemple :

N0 est en colère = *La moutarde monte au nez de N0*

Un dictionnaire TAG "sémantique" est en cours qui est constitué de deux dictionnaires TAG syntaxiques reliés entre eux. Tous les arbres identiques (et toutes les entrées identiques) sont reliés entre eux, les noeuds identiques sont liés entre eux, mais des noeuds différents le sont également dans le cas des expressions figées et des phrases à verbe support (Cf chapitre 4). Nous avons également commencé à relier certaines entrées morphologiquement apparentées, par exemple:

N₀ commettre une agression contre N₁ = N₀ agresser N₁

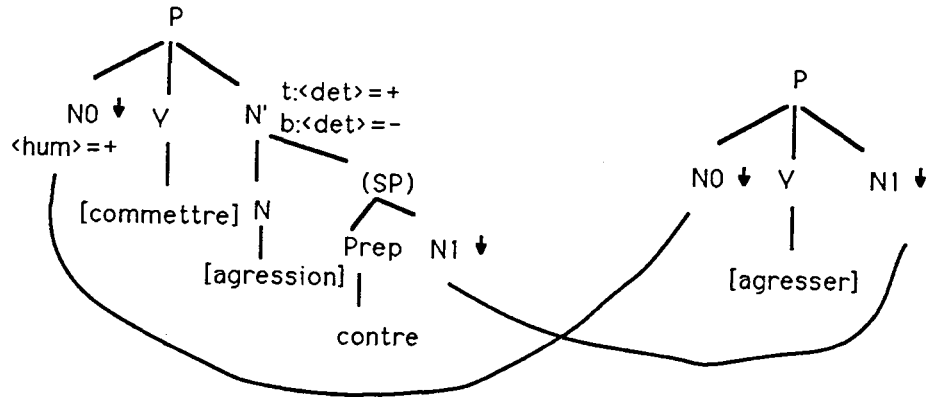
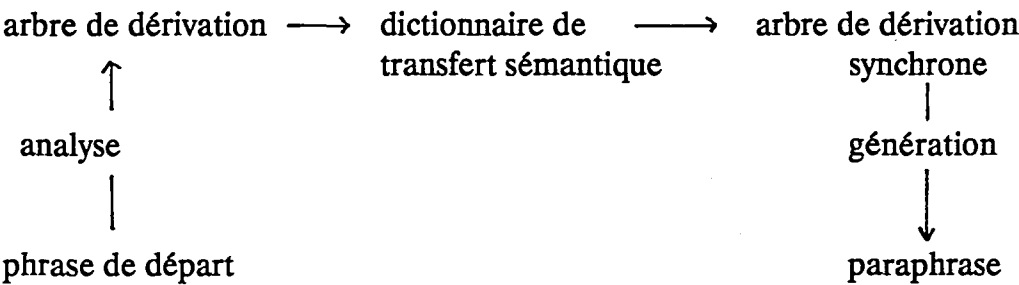


Figure 17

Le système devient donc:

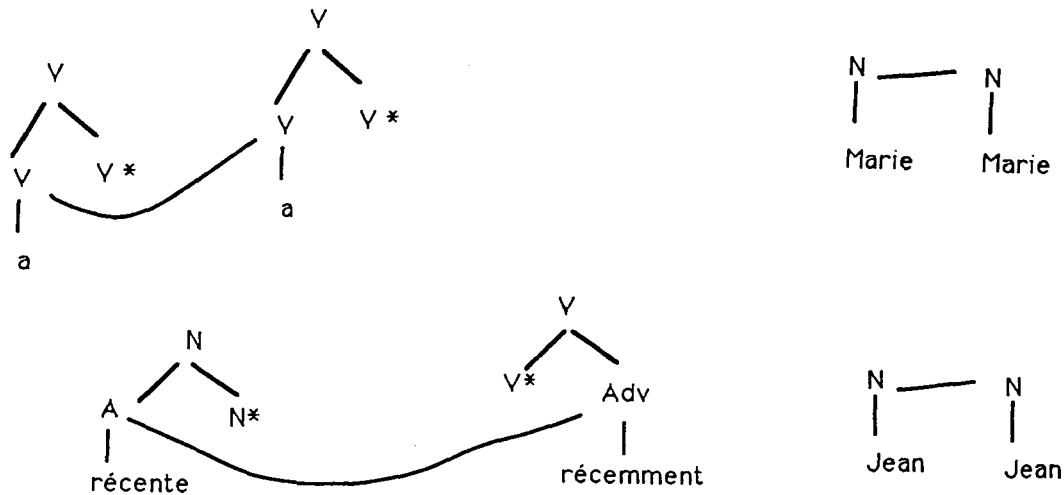


Un tel système présente plusieurs avantages: il est fondé uniquement sur les unités linguistiques mêmes, ce qui évite le recours à des concepts plus arbitraires et souvent liés à une application particulière. Notre grammaire permet également, grâce à ses unités étendues, de manipuler des unités sémantiquement plus naturelles que les grammaires hors contexte habituelles. Il s'agit donc d'un "dictionnaire de synonymes" qui exprime des équivalences non entre mots mais entre phrases simples.

Ce système permet aussi d'exprimer facilement des correspondances entre unités lexicales (et donc entre arbres élémentaires) de structures syntaxiques assez différentes. Contrairement aux habituels dictionnaires de synonymes, un lien de synonymie peut ainsi être défini entre unités de catégories différentes, par exemple entre adjectifs et adverbes. Une phrase comme *Jean a récemment agressé Marie* pourra donc être

paraphrasée par *Jean a commis une récente agression contre Marie*:

arbres élémentaires (autre ceux supra)



arbres de dérivation

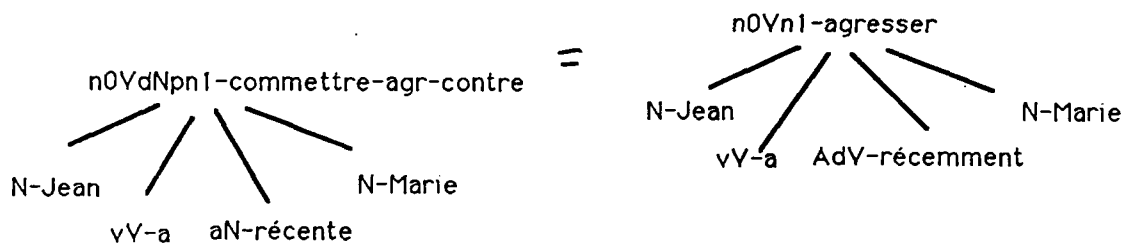


Figure 18

Deux limitations sont pour l'instant évidentes. D'une part, il est des entrées lexicales qui n'ont pas forcément de correspondant lexical, par exemple *on*, *quelqu'un*, *quelque chose*; il faudrait donc définir un arbre vide auquel ils puissent être liés (puisque deux dérivations sont synchrones seulement si elles ont même nombre d'arbres élémentaires). Ainsi, une phrase comme *Marie est aimée* pourrait être paraphrasée par: (*On + Quelqu'un*) aime Marie, et *Jean mange* par *Jean mange quelque chose*.

D'autre part, il est bien clair que les relations mettent en jeu des groupements de mots libres aussi bien que des groupement figés. Par exemple N_0 déplaire à $N_1 = N_0$ ne pas plaire à N_1 , c'est-à-dire la combinaison de *ne pas* et de N_0 plaire à N_1 , ou N_0 chuchoter à N_1 que $P = N_0$ dire à N_1 que P + *en chuchotant*.

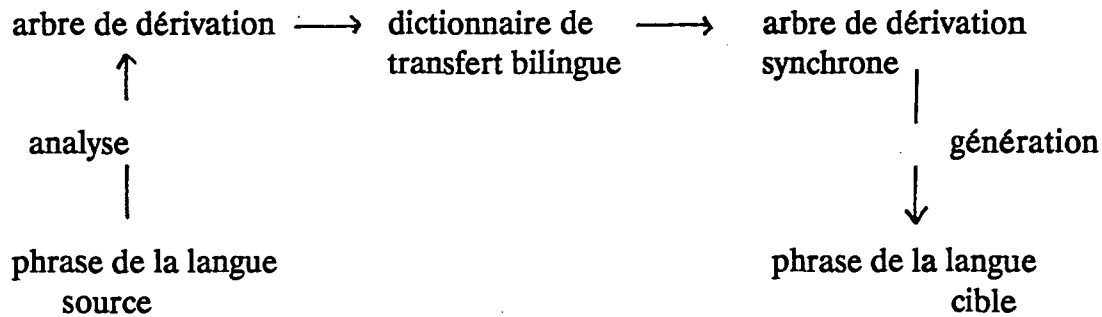
Il faudrait donc, pour une application vraiment fonctionnelle, ajouter des entrées lexicales formées de la combinaison de plusieurs arbres élémentaires, et concevoir le dictionnaire de transfert sémantique avec d'un côté le dictionnaire TAG classique, de l'autre un dictionnaire TAG complété par autant de nouvelles entrées lexicales⁸.

⁸ On pourrait penser à compléter le dictionnaire de façon semi automatique en se basant sur des régularités syntaxiques ou morphologiques. Par exemple, la relation N_0 V que P à $N_2 = N_0$ dire que P à N_2 en V-ant est générale, de même que N_0 V (Loc N_1) (Vinf) = N_0 aller en V-ant (Loc N_1) (Vinf) (M. Gross 1975). En revanche, N_0 dé-V W = N_0 ne pas V W l'est moins (Cf *poser/déposer*, *river/dériver*...), de même que N_0 être

5.3. Traduction automatique

Une grammaire d'arbres adjoints pour l'anglais est en cours d'élaboration, sur le même modèle que celle du français (A. Abeillé et al. 1990a).

Un module de transfert qui relie ces deux grammaires, c'est-à-dire les entrées des deux dictionnaires TAG correspondants, est également en cours d'élaboration. Il est basé sur le principe des TAG synchrones. La stratégie est donc (A. Abeillé et al. 1990b) :



La non correspondance entre les unités lexicales des deux langues peut amener, comme pour la paraphrase, à définir de nouvelles entrées. Il est à noter que contrairement à d'autres formalismes, les grammaires d'arbres adjoints permettent de minimiser ces discordances. Ainsi la correspondance entre un verbe simple dans une langue et une expression figée (ou une phrase à verbe support) dans l'autre s'exprime de façon naturelle et sans mécanisme particulier puisque dans chaque langue on les considère comme des unités lexicales à part entière. Par exemple: *N₀ commute* = *N₀ faire la navette* ou:

John alluded to the crime.
 = *John a fait allusion au crime.* (exemple de M. Salkoff 1990)

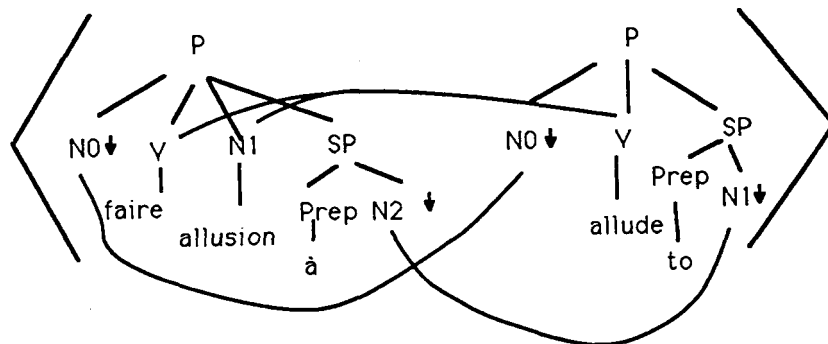


Figure 19

V-able (par X) = X pouvoir V N₀ (D. Leeman 1990). D'autres études sont en cours au LADL, mais, d'une manière générale, il ne semble pas que préfixes et suffixes puissent fournir des informations syntaxiques et sémantiques suffisamment fiables.

En revanche, lorsque la traduction est une périphrase non ressentie comme figée dans une perspective monolingue, il faut définir une nouvelle entrée qui combine plusieurs structures élémentaires:

John limped to the store
 ?? *Jean a boîté jusqu'au magasin*
 = *John est allé jusqu'au magasin en boîtant*

Il faut donc considérer que, dans une perspective de traduction anglais-français, *N₀ aller Loc N₁ en boîtant* devient une expression figée, et la rajouter au dictionnaire de transfert. Celui-ci relie donc le dictionnaire TAG français et le dictionnaire TAG anglais ainsi "complétés".

Lorsque ces nouvelles combinaisons sont systématiques (ex: Verbe anglais = faire + verbe français), on peut utiliser un trait qui force l'adjonction d'un arbre élémentaire supplémentaire dans la langue cible, par exemple celle du causatif:

John walked the dog to the store
 = * *John a marché le chien jusqu'au magasin*
 = *John a fait marcher le chien jusqu'au magasin*

The sun melts the snow
 = * *Le soleil fond la neige*
 = *Le soleil fait fondre la neige*

Notre formalisme permet également aisément d'exprimer des relations entre unités de structures ou de catégories lexicales différentes.

Conclusion

Nous avons présenté dans cette thèse une grammaire du français écrite dans le formalisme des grammaires d'arbres adjoints. Nous avons d'abord présenté le formalisme original et les modifications que nous y avons apportées (substitution, lexicalisation). Nous avons justifié ensuite notre usage des représentations arborescentes, ainsi que les différentes catégories et syntagmes que nous employons. En particulier, nous avons remis en cause la notion de syntagme verbal pour une langue comme le français (chapitre 2). Nous avons exposé ensuite les différents types d'arbres élémentaires, définis par rapport à la catégorie de leurs têtes lexicales, ainsi que les principes de leur regroupement en familles d'arbres pour la prise en compte des différentes transformations (au sens linguistique du terme; chapitre 3). Puis nous avons montré l'intérêt des arbres lexicalisés pour la représentation des phénomènes non compositionnels rencontrés dans les phrases à verbe support et les expressions figées. Ce qui nous a permis de donner les bases de l'analyse sémantique qui devrait compléter l'analyse syntaxique dans notre système (TAG autosynchrones; chapitre 4). Nous avons ensuite expliqué les principes gouvernant les phénomènes d'extraction qui découlent naturellement des arbres élémentaires définis dans notre grammaire (chapitre 5), puis exposé quelques développements en cours (chapitre 6) visant à étendre encore davantage la couverture de notre grammaire. Enfin, nous avons exposé les principes de l'analyse automatique découlant de la lexicalisation de notre grammaire et présenté l'état actuel de l'implémentation de notre système d'analyse syntaxique du français (chapitre 7).

Notre grammaire est loin d'être achevée, mais nous espérons que sa couverture est déjà suffisante pour permettre de bons résultats. Nous l'avons testée sur des phrases préfabriquées pour l'instant et il nous reste à tester notre système sur des phrases tirées de textes réels. Nous l'avons testée avec un analyseur syntaxique prédéfini (de type Earley); il serait intéressant d'en explorer d'autres.

Nous voudrions insister sur trois points en conclusion :

1. Tous les systèmes de traitement automatiques sont basés sur la distinction entre composant syntaxique et composant lexical. Dans les équipes suffisamment importantes, ce ne sont en général pas les mêmes personnes qui écrivent des grammaires formalisées et qui écrivent des dictionnaires électroniques (la dernière activité pouvant être reléguée à de simples opérateurs de saisie). La coupure est moindre que celle qui existe traditionnellement entre livres de grammaires et dictionnaires papiers puisque le traitement automatique suppose une articulation explicite entre les deux composants, articulation qui dans le cas humain est laissée à l'intuition ou au génie du lecteur. Elle existe néanmoins et ne se justifie guère.

Nous avons remis en cause une telle distinction. Nous basant sur une théorie linguistique qui articule plus étroitement lexique et syntaxe (les lexiques-grammaires du LADL), nous avons proposé un modèle informatique où lexique et grammaire forment un seul et même composant. Ce modèle est basé sur le formalisme des grammaires

d'arbre adjoints auquel nous avons (en collaboration avec A. Joshi et son équipe) apporté des modifications substantielles. Il en résulte des gains appréciables quant à la couverture de la grammaire, à la finesse des descriptions linguistiques et à l'efficacité du traitement automatique.

2. Notre grammaire n'est pas basée sur des règles de réécriture hors contexte mais sur des arbres élémentaires, c'est-à-dire sur des unités syntaxiques "étendues". La plupart des systèmes de traitement automatiques basés sur des grammaires de réécriture hors contexte, c'est-à-dire sur des modèles opérant par concaténation locale de symboles, considèrent ces symboles comme des systèmes de traits complexes combinables par unification pour tourner en partie les difficultés liées à cette stricte localité. Notre modèle permet de définir des domaines élémentaires d'étendue variable et donc d'échapper à cette stricte localité. Ce faisant, nous avons pu établir une distinction linguistiquement motivée entre ce qu'il est préférable de représenter par un système de traits et ce qu'il est préférable de considérer comme des structures ou des symboles distincts et indécomposables.

3. Notre modèle nous permet également d'intégrer naturellement à la grammaire générale du français les nombreuses expressions figées ou semi-figées. Il n'y a pas en général prise en compte des expressions figées dans les systèmes de traitement automatique. S'il y en a, c'est sur le mode d'une bipartition stricte entre expressions libres et expressions figées, ces dernières étant considérées comme non susceptibles de variations (et souvent agglutinées lors d'un prétraitement). Notre modèle rend compte de la grande variabilité des expressions dites figées, ainsi que des nombreux phénomènes de figement (notamment prépositionnels) affectant les expressions dites libres. Nous avons donc rendu compte d'un continuum entre expressions figées et expressions libres.

Notre mode de représentation des expressions figées est particulièrement simple et leur grammaire est totalement intégrée à celle des phrases "libres".

BIBLIOGRAPHIE

- A. Abeillé, 1988a. "Light verb constructions and extraction out of NP in Tree Adjoining Grammar", *Papers from the 24th Regional Meeting of the Chicago Linguistic Society*, Chicago.
- A. Abeillé, 1988b. "Parsing French with a Tree Adjoining Grammar", *Actes 12^e COLING*, Budapest.
- A. Abeillé, 1988c. "A Lexicalized Tree Adjoining Grammar for French: the general framework", *Rapport technique*, CIS Dept, Université de Pennsylvanie, Philadelphie.
- A. Abeillé, Y. Schabes, 1989. "Parsing Idioms with a lexicalized Tree Adjoining Grammar", *Actes 4th European Conference of the Association of Computational Linguistics*, Manchester.
- A. Abeillé, K. Bishop, S. Cote, A. Joshi, Y. Schabes, 1989. "Lexicalized TAGs, Parsing and Lexicon", *Actes DARPA Speech and Natural Language Workshop*, Philadelphie.
- A. Abeillé, K. Bishop, S. Cote, Y. Schabes, 1990. "A Lexicalized Tree Adjoining Grammar for English", *Rapport technique*, CIS Dept, Université de Pennsylvanie, Philadelphie.
- A. Abeillé, Y. Schabes, 1990. "Non compositional discontinuous constituents in Tree Adjoining grammar", *Actes International Symposium on discontinuous constituency*, Tilburg, Pays-Bas.
- A. Abeillé, Y. Schabes, A. Joshi, 1990. "Using Lexicalized TAGs for machine translation", *Actes 13^e COLING*, Helsinki.
- A. Abeillé, 1990a. "Lexical and Syntactic rules in a Tree Adjoining Grammar", *Actes 28^e annual meeting of ACL*, Pittsburgh.
- A. Abeillé, 1990b. "L'unification dans une grammaire d'arbres adjoints: quelques exemples en syntaxe française", *T.A. Informations*, bulletin de l'ATALA, Paris.
- S. Abney, 1986. "Functional elements and licensing", *actes GLOW*, Barcelone.
- A. Aho, J. Ullman, 1973. *Theory of parsing, translation and compiling*. vol. 1 Parsing, Prentice-Hall.
- K. Bashung, G. Bès, A. Corluy, T. Guillotin, 1987. "Auxiliaires and clitics in French UCG grammar", *Actes 3rd European Conference of ACL*, Copenhague.
- T. Becker, 1990. "Metarules in Tree Adjoining grammars", *Actes First International Workshop on Tree Adjoining Grammars*, Sarrebruck.
- A. Bérard-Dugourd, 1990. *Rôle d'une analyse automatique dans un système de compréhension du langage naturel*, Thèse de Doctorat, Université Paris 7.
- G. Bès, C. Gardent, 1989. "French Order without order", *Actes 4th European Conference of ACL*, Manchester.
- J.P. Boons, A. Guillet, C. Leclère, 1976a. *La Structure des Phrases Simples en français: Constructions Intransitives*, Droz, Genève.

J.P. Boons, A. Guillet, C. Leclère, 1976b. *La Structure des Phrases Simples en français: Classes de constructions transitives*, Rapport de recherches n°6, LADL, Université Paris 7.

A. Borillo, 1971. "Remarques sur les verbes symétriques français", *Langue Française*, n°11.

J. Bresnan, 1978. "A realistic transformational grammar" in M. Halle, J. Bresnan, G. Miller (eds), *Linguistic Theory and Psychological reality*, MIT Press.

J. Bresnan, 1982. "Passive in lexical theory", in Bresnan (ed), *The Mental Representation of Grammatical Relations*, MIT Press.

J. Bresnan, R. Kaplan, 1982. "Lexical Functional Grammar: a formalism for grammatical representations", in Bresnan (ed), *The Mental Representation of Grammatical Relations*, MIT Press.

L. Burzio, 1981. *Intransitive verbs and Italian auxiliaries*, Thèse de PhD, MIT.

J. Calder, E. Klein, H. Zeevat, 1988. "Unification categorial grammar: a concise, extendable grammar for natural language development", *Actes 12° COLING*, Budapest.

N. Catach, *Les listes orthographiques de base du Français*, Nathan, 1984.

R. Catell, 1984. *Complex Predicates in English*, Syntax and Semantics vol. 17, Academic Press.

N. Chomsky, 1957. *Syntactic Structures*, Mouton, La Haye. traduction française, *Structures syntaxiques*, Seuil, 1965.

N. Chomsky 1965. *Aspects of the theory of syntax*, MIT Press. traduction française, *Aspects de la théorie syntaxique*, Seuil, 1970.

N. Chomsky 1971. "Remarks on nominalization", in R. Jacobs et P. Rosenbaum, *Readings in English transformational grammar*, Ginn, Blaisdell.

N. Chomsky, 1973. "Conditions on transformations", in S. Anderson, P. Kirsparsky (eds), *A Festschrift for M. Halle*, Holt, Reinhart, Winston, New York.

N. Chomsky, 1977. "On wh-movement", in P. Culicover et al. (eds), *Formal Syntax*, Reidel.

N. Chomsky, 1981. *Lectures on government and Binding*, Foris Dordrecht. (traduction française, *La Nouvelle Syntaxe*, Seuil, 1988).

N. Chomsky, 1986. *Barriers*, MIT Press.

K. Church, 1988. "A stochastic parts program and noun phrase parser for unrestricted text", *Actes 2° International Conference on applied Natural Language Processing*, Austin.

A. Daladier, 1978. *Problèmes d'analyse d'un type de nominalisation en français et de certains groupes nominaux complexes*, Thèse de 3ème cycle, Université Paris 7.

L. Danlos, 1980. *Représentations d'informations linguistiques: constructions N être Prep X*, Thèse de 3ème cycle, Université Paris 7.

L. Danlos, 1988. "Les expressions figées construites avec le verbe Etre Prep", *Langages*, n°90.

- L. Danlos, 1989. "Some pronominalization issues in generation of texts in Romance languages", in M. Gross et D. Perrin (eds), *Electronic dictionaries and automata in computational linguistics*, Springer Verlag.
- L. Danlos, A. Bech, F. Caroli, B. Daille, N. Mesli, F. Namer, S. Mohr, "Support verbs and predicative nouns" in J. Durand (ed), *Eurotra Reference manual 6.1*, Eurotra, Luxembourg.
- L. Danlos, à paraître. "Translation of a semi-compositional phenomenon: predicative nouns and support verb constructions", *Journal of French Linguistics*, Cambridge University Press.
- F. Debili, 1977. *Traitements syntaxiques utilisant des matrices fréquentielles construites automatiquement par apprentissage*, Thèse de Docteur Ingénieur, Université Paris 7.
- A-M. Derouault, 1985. *Modélisation d'une langue naturelle pour la désambiguation des chaînes phonétiques*, Thèse de doctorat d'Etat, Université Paris 7.
- O. Ducrot, 1968. "Sur quelques illogismes des langues naturelles", *Langages*, 2.
- J. Earley, 1970. "An efficient context-free parsing algorithm", *Communications of the ACM*, 13-2.
- J. Emonds, 1970. *Root and Structure Preserving Transformations*, Thèse de PhD, MIT, Cambridge.
- J. Emonds, 1978. "The verbal complex V' V in French", *Linguistic Inquiry*, 9:2.
- N. Erteschik, 1977. *On the nature of island constraints*, Thèse de PhD, MIT.
- B. Fraser, 1970. "Idioms within a transformational grammar", *Foundations of language*, 6.
- N. Fukui, M. Speas, 1985. "Specifiers and projections", *MIT Working papers in Linguistics*, n°8, MIT.
- A. Gal, J-H. Jayez, P. Levasseur, M. Liscouet, J-M. Walle, 1986. *Compréhension automatique du langage naturel, le cas de l'interrogation simple en français*, Masson.
- C. Gardent, G. Bès, F. Jurie, K. Bashung, 1989. "Efficient parsing for French", *Actes 27th annual meeting of ACL*, Vancouver.
- G. Gazdar, 1981. "Unbounded Dependencies and Coordinate Structures", *Linguistic Inquiry*, 12, pp. 155-184.
- G. Gazdar, E. Klein, G. Pullum, I. Sag, 1985. *Generalized Phrase Structure Grammar*, Harvard University Press.
- G. Gazdar, 1988. "Applicability of indexed grammars to natural language", in U. Reyle, C. Rohrer (eds), *Natural language parsing and linguistic theories*, Dordrecht Reidel.
- J. Giry-Schneider, 1976. "Constructions passives à sujet à déterminant zéro", *Recherches Linguistiques*, n°10, Vincennes.
- J. Giry-Schneider, 1978. *Les nominalisations en français. L'opérateur Faire dans le lexique*, Droz, Genève.
- J. Giry-Schneider, 1987. *Les prédicats nominaux en français. Les phrases simples à verbe support*, Droz, Genève.

- D. Godard, 1980a. "La contrainte du syntagme nominal en français et les verbes ponts", *LINX*, n°2.
- D. Godard, 1980b. *Les relatives parenthétiques du français et les conditions sur les règles de mouvement*, Thèse de 3ème cycle, Université Paris 7.
- D. Godard, 1988. "Sujets et compléments génitifs dans le groupe nominal", in Milner (ed), *Recherches nouvelles sur le langage*, Publications de l'ERA 642, Université Paris 7.
- D. Godard, à paraître. "Extraction from NP in French", ms soumis à *Natural Language and Linguistic Theories*.
- G. Gougenheim, 1959. "Y a-t-il des prépositions vides en français ?", *Le Français Moderne*, n°1, pp. 1-23.
- S. Greibach, 1974. "Some restrictions on W-grammars", *International Journal of Computer and Information Science*, 3:415-462.
- J. Grimshaw, 1982. "Reflexivization and Romance clitics", in J. Bresnan et R. Kaplan (eds), 1982.
- J. Grimshaw, 1988. *Argument-adjuncts*, Rapport technique, Lexicon Project, MIT.
- G. Gross, 1981. "Les prépositions composées", *Actes du colloque franco-allemand de grammaire transformationnelle*, vol. 2, Niemeyer.
- G. Gross, 1990. *Etude de constructions converses en français*, Droz, Genève.
- M. Gross, 1969. "Remarques sur la notion d'objet direct en Français", *Langue française*, n°3, Paris.
- M. Gross, 1968. *Grammaire transformationnelle du français: syntaxe du verbe*, Larousse (réed Cantilène, 1986).
- M. Gross, 1975. *Méthodes en syntaxe*, Hermann, Paris.
- M. Gross, 1976. "Sur quelques groupes nominaux complexes", in J.C. Chevalier et M. Gross (eds), *Méthodes en grammaire française*, Klincksieck.
- M. Gross, 1977a. "Une analyse non présuppositionnelle de l'effet contrastif: l'extraction dans *c'est ... qu...* et la négation", *Linguisticae Investigationes*, 1:2, J. Benjamins.
- M. Gross 1977b. *Grammaire transformationnelle du français: syntaxe du nom*, Larousse (réed Cantilène, 1987).
- M. Gross, 1981. "Les bases empiriques de la notion de prédicat sémantique", *Langages*, n°63, Paris.
- M. Gross, 1982. "Classification des phrases figées en français", *Revue Québécoise de Linguistique*, 11:2, Montréal.
- M. Gross, 1986. "Lexicon grammar: the representation of compound words", *Actes 11° COLING*, Bonn.
- M. Gross, 1988. "Limites entre phrases figées, phrases à verbe support et phrases libres", *Langages*, n° 90, Paris.

- M. Gross, 1989a. "La construction de dictionnaires électroniques", *Annales des Télécommunications*, 44:1-2, Paris.
- M. Gross, 1989b. "Les expressions figées en Français", Rapport technique, LADL, Université Paris 7 (à paraître Asstril, 1991).
- M. Gross 1989c. "The use of finite state automata in the lexical representation of natural language", in M. Gross et D. Perrin (eds), *Electronic dictionaries and automata in computational linguistics*, Springer Verlag.
- M. Gross, 1990. *Grammaire transformationnelle du français: syntaxe de l'adverbe*, Asstril, Université Paris 7.
- G. Gougenheim, 1959. "Y a-t-il des prépositions vides en français ?", *Le Français Moderne*, n°1, pp. 1-25.
- A. Guillet, 1990. *Une classification des constructions transitives à argument locatif*, Thèse d'Etat, Université Paris 7.
- Z. Harris, 1951. *Methods in structural Linguistics*, University of Chicago Press.
- Z. Harris, 1962. *String analysis of sentence structure*, Mouton, La Haye.
- Z. Harris 1976. *Notes du cours de syntaxe*, Seuil, Paris.
- H. Huot, 1981. *Constructions infinitives du français: le subordonnant de*, Droz, Genève.
- H. Huot, 1986. "Le subjonctif dans les complétives: subjectivité et modalisation" in M. Ronat et D. Couquaux (eds), *La grammaire modulaire*, Minuit.
- R. Jackendoff, 1977. *X-bar syntax: a study of phrase structure*, MIT Press.
- J-H. Jayez, 1982. *Compréhension automatique du langage naturel, le cas du groupe nominal en français*, Masson.
- M. Johnson, 1985. "Parsing with discontinuous elements", *Actes 23° annual meeting of ACL*, Stanford.
- A. Joshi, 1969. "Properties of formal grammars with mixed type of rules and their linguistic relevance", *Actes Coling'69*, Stockholm.
- A. Joshi, S. Kosaru, H. Yamada, 1972. "String Adjuncts Grammars", *Information and Control*, vol. 21, n°2-3.
- A. Joshi, L. Levy, M. Takahashi, 1975. "Tree Adjunct Grammars", *Journal of the Computer and System Sciences*, 10:1.
- A. Joshi, 1985. "How much context-sensitivity is necessary for characterizing structural descriptions", in *Natural Language Processing : Psycholinguistic, Computational and Theoretical Perspectives*, Cambridge University Press.
- A. Joshi, 1987a. "Introduction to Tree Adjoining Grammar", in A. Manaster Ramer (ed), *The Mathematics of Language*, J. Benjamins.
- A. Joshi, 1987b. "The relevance of Tree Adjoining Grammar to generation", in G. Kempen (ed), *Natural Language Generation*, Dordrecht.

- A. Joshi, K. Vijay-Shanker, D. Weir, 1989. "The Convergence of mildly context-sensitive formalisms", *Rapport technique*, Université de Pennsylvanie, Philadelphie.
- A. Joshi, K. Vijay-Shanker, 1989. "Treatment of long distance dependencies in LFG and TAG: Functional Uncertainty is a corollary in TAG", *Actes 27^e annual meeting of ACL*, Vancouver.
- A. Joshi, 1990. "Coordination in TAG in a CCG manner: fixed and flexible phrase-structure", *Actes 1st International Workshop on Tree Adjoining Grammars*, Sarrebruck.
- A. Juilland, D. Brodin, C. Davidovitch, 1970. *Frequency dictionary of French words: 5000 words*, Mouton, La Haye.
- H. Kamp, 1984. "A theory of truth and semantic representation" in Groenendijk et al. (eds), *Truth, interpretation and information*, Foris.
- R. Kaplan, A. Zaenen, 1989. "Long distance dependencies, Constituent structure and Functional uncertainty", in M. Baltin & A. Kroch (eds), *Alternative Conceptions of Phrase Structures*, Chicago University Press.
- L. Karttunen, 1989. "Radical lexicalism", in M. Baltin & A. Kroch (eds), *Alternative Conceptions of Phrase-structure*, Chicago University Press.
- R. Kasper et W. Rounds, 1986. "A logical semantics for feature structures", *Actes 24th annual meeting of ACL*, New York. (traduit dans P. Miller, T. Torris (eds), 1990)
- M. Kay, 1983. "Unification grammar", *Actes 10th annual meeting of the Berkeley Linguistics Society*, Berkeley.
- M. Kay, 1984. "Functional unification grammar: a formalism for machine translation", *Actes 10^e COLING*, Stanford.
- R. Kayne, 1975. *French syntax*, MIT Press (traduction française *Syntaxe du français : le cycle transformationnel*, Seuil, 1977).
- S. Keyser, 1968. Review of S. Jacobson, *Adverbial positions in English*, *Language*, n°44, pp. 357-374.
- S. Keyser, P. Postal, 1976. *Beginning English Grammar*, Harper and Row, New York.
- A. Kornai, G. Pullum, 1990. "The X-bar theory of phrase structure", *Language*, 66:1.
- A. Kroch, A. Joshi, 1985. "Some Aspects of the linguistic relevance of Tree Adjoining Grammar", *Rapport technique*, CIS Dept, Université de Pennsylvanie, Philadelphie.
- A. Kroch, 1987. "Unbounded Dependencies and Subjacency in a Tree Adjoining Grammar", in A. Manaster Ramer (ed) : *The Mathematics of Language*, J. Benjamins.
- A. Kroch, 1989. "Assymetries in long distance extraction in a TAG Grammar", in M. Baltin and A. Kroch (eds) : *Alternative Conceptions of Phrase Structure*, Chicago Press.
- A. Kroch, A. Joshi 1986. "Analysing Extraposition in a TAG", in G. Huck and A. Ojeda (eds): *Discontinuous Constituents*, Syntax and Semantics vol. 20, Academic Press, New York.
- S. Kuno, 1965. "A system for transformational analysis", in Mathematical Linguistics and Automatic Translation, NSF Report n°15, The Computation Laboratory, Harvard University.

- S. Y. Kuroda, 1968. "English relativization and certain related problems", *Language*, 44:1.
- J. Labelle, 1974. *Etude de constructions avec opérateur AVOIR (nominalisations et extensions)*, Thèse de 3ème cycle, Université Paris 7.
- B. Lafaye, 1841. *Dictionnaire des synonymes français*, Hachette, Paris.
- G. Lakoff, J. Ross, 1966. "Criterion for Verb Phrase constituency", in *Mathematical Linguistics and Automatic Translation*, NSF Report 17:2, The Computation Laboratory, Harvard University.
- B. Lang, 1990. "The systematic construction of Earley parsers: application to the production of $O(n^6)$ Earley parsers for TAGs", *Actes 1st International Workshop on Tree Adjoining Grammars*, Sarrebruck.
- E. Laporte, "La reconnaissance des expressions figées lors de l'analyse automatique", *Langages*, n°88, 1990.
- R. Larson, 1988. "On double object constructions", *Linguistic Inquiry*, 19:3.
- C. Leclère, 1971. "Remarques sur les substantifs opérateurs", *Langue Française*, n°11.
- J. McCawley, 1982. "Parenthetical and discontinuous constituent structures", *Linguistic Inquiry*, n° 13.
- A. Marantz, 1981. *On the nature of grammatical relations*, Thèse de PhD, MIT, Cambridge.
- M. Marcus, 1980. *A theory of syntactic recognition for natural language*, MIT Press, Cambridge.
- R. Martin, 1974. "La notion d'adverbe de phrase: essai d'interprétation en grammaire générative", *Actes du colloque franco-allemand de grammaire transformationnelle*, vol. 2, Niemeyer.
- D. Maurel, 1989. *Reconnaissance de séquences de mots par automate: adverbes de date du français*, Thèse de Doctorat, Université Paris 7.
- A. Meunier, 1977. "Sur les bases syntaxiques de la morphologie dérivationnelle", *Linguisticae Investigationes*, 1:2, J. Benjamins, Amsterdam.
- D. Michelson, 1969. "An examination of Lakoff and Ross's criterion for VP constituency", *Glossa*, 3:2.
- P. Miller, T. Torris (eds), 1990. *Formalismes syntaxiques pour le traitement automatique du langage naturel*, Hermès, Paris.
- J.C. Milner, 1978. *De la syntaxe à l'interprétation*, Seuil.
- J.C. Milner, 1982. *Ordre et raisons de langue*, Seuil.
- J.C. Milner, 1984. "De l'inutilité des arbres en linguistique", *TA Informations*, bulletin de l'ATALA. (repris dans Publications de l'ERA 942, DRL, Université Paris 7, 1985).
- J.C. Milner, 1986. "Introduction à un traitement du passif", Publications de l'ERA 942, DRL, Université Paris 7.

- J.C. Milner, 1989. *Introduction à une science du langage*, Seuil.
- C. Molinier, 1984. *Syntaxe et sémantique des adverbes en -ment*, Thèse de doctorat, Université Paris 7-Jussieu.
- F. Namer, B. Daille, T. Kyriacopoulou, O. Laurens, A. Nehmé, G. Recourcé (et al.), 1990. *Rapport d'implémentation du français 4.0*, Eurotra France, LISH, Paris.
- D. de Negroni-Peyre, 1978. "Nominalisations par être en et réflexivation", *Linguisticae Investigationes*, 2:1, J. Benjamins, Amsterdam.
- F. Newmeyer, 1975. "The regularity of idiom behaviour", *Lingua* n°34.
- M. Palis, S. Shende, 1990. "Fast parallel recognition of Tree Adjoining Languages", *Theoretical computer science*.
- S. Peters, R. Ritchie, 1968. "On the generative power of transformational grammars", Rapport Technique, Université de Washington. Publié dans *Information Science* 6: 49-83, 1973.
- L. Perrault, 1985. "On the mathematical properties of linguistic theories", *Actes 23^e annual meeting of ACL*, New York. (publié dans B. Grosz, K. Sparck-Jones, B. Weber, *Readings in natural language processing*, Morgan Kaufmann, 1987).
- D. Perlmutter, P. Postal, 1977. "Toward a universal characterization of passivization", *actes 3^e annual meeting of Berkeley Linguistic Society*. (publié dans Perlmutter (ed), *Studies in relationnal grammar*, Chicago Press, 1983).
- L. Picabia, 1978. *Etude transformationnelle des adjectifs du français*, Droz, Genève.
- M. Piot, 1978. *Etude de quelques classes de subordination en français*, Thèse de 3^eme cycle, Université Paris 8.
- C. Pollard, 1985. *Generalized Phrase Structure Grammar, Head Grammars and natural language*, Thèse de PhD, Stanford University.
- M. Rady, 1983. *L'ambiguïté du langage naturel est-elle la source du non déterminisme des procédures de traitement ?*, Thèse de doctorat ès sciences, Université Paris 6.
- E. Ristad, 1986. "Computational complexity of current GPSG theory", *Actes 24^e annual meeting of ACL*, Morristown.
- J. Ross, 1967. *Constraints on variables in syntax*, Thèse de PhD, MIT (publié sous le titre: *Infinite Syntax!*, Reidel Dordrecht, 1985).
- N. Ruwet, 1967. *Introduction à la grammaire générative*, Plon.
- N. Ruwet, 1982. "Les constructions pronominales et moyennes" in *Théorie syntaxique et syntaxe du français*, Seuil.
- N. Ruwet, 1984. "Du bon usage des expressions idiomatiques en grammaire générative", *Recherches Linguistiques*, n°11.
- G. Sabah, 1988. *L'intelligence artificielle et le langage; Représentation des connaissances: vol.1*, Hermès, Paris.

- G. Sabah, 1990. *L'intelligence artificielle et le langage; Processus de compréhension*: vol.2, Hermès, Paris.
- G. Sabah, M. Rady, 1983. "A deterministic syntactic-semantic parser applied to French", *Actes 8° IJCAI*, Karlsruhe.
- I. Sag, 1976. "A note on verb-phrase deletion", *Linguistic Inquiry*, 7:4, p.664.
- N. Sager, 1981. *Natural Language Information Processing: a computer grammar of English and its applications*, Addison-Wesley, Reading.
- M. Salkoff, 1973. *Une grammaire en chaîne du Français: analyse distributionnelle*, Dunod.
- M. Salkoff, 1979. *Analyse syntaxique du Français: Grammaire en chaîne*, J. Benjamins.
- M. Salkoff, 1990. "Translation of support-verb constructions", *Actes 13° Coling*, Helsinki.
- Y. Schabes, A. Abeillé, A. Joshi, 1988. "Parsing strategies with Lexicalized Grammars: an application to Tree Adjoining Grammars", *Actes 12° Coling*, Budapest, 1988.
- Y. Schabes, A. Joshi, 1988. "An Earley-type Algorithm for Tree Adjoining Grammars", *Actes 26° annual meeting of ACL*, Buffalo.
- Y. Schabes, A. Joshi, 1990. "Parsing Lexicalized Grammars", in M. Tomita (ed), *Current Issues in Parsing Technologies*, Kluwer.
- Y. Schabes, K. Vijay Shanker, 1990. "A deterministic left to right parser for Tree Adjoining Grammar", *Actes 28° annual meeting of ACL*, Pittsburg.
- Y. Schabes, 1990. *Mathematical and computational aspects of lexicalized grammars*, Thèse de PhD, Université de Pennsylvanie, Philadelphie.
- S. Schlyter, 1974. "Une hiérarchie d'adverbes en français et leurs distributions", *Actes du colloque franco-allemand de grammaire transformationnelle*, vol. 2, Niemeyer.
- S. Shieber, S. Stucky, H. Uszkoreit, J. Robinson, 1983. "Formal constraints on metarules", *Actes 21° annual meeting of ACL*, Cambridge.
- S. Shieber, 1986. *An introduction to unification-based approaches to grammars*, CSLI series, Chicago Press. (traduit dans P. Miller, T. Torris (eds), 1990)
- S. Shieber, Y. Schabes, 1990. "Synchronous Tree Adjoining Grammars", *actes 13° COLING*, Helsinki.
- M. Silberstein, 1989. *Dictionnaires électroniques et reconnaissance lexicale automatique*, Thèse de doctorat, Université Paris 7.
- I. Simatos, 1986. *Eléments pour une théorie des expressions idiomatiques*, Thèse de doctorat, Université Paris 7.
- K. De Smedt, G. Kempen, 1990. "Segment grammar: a formalism for incremental sentence generation" in C. Paris et al. (ed) *Natural language generation and computational linguistics*, Dordrecht, Kluwer.
- E. Spang Hanssen, 1963. *Les prépositions incolores du français moderne*, GEC Gads Forlag, Copenhagen.

M. Steedman, 1985. "Dependency and coordination in the grammar of Dutch and English", *Language*, n°61.

M. Steedman, 1988. "Combinators and grammars", in T. Oehrle, E. Bach, D. Wheeler (eds), *Categorial grammars and natural language structures*, Dordrecht Reidel.

L. Tesnière, 1956. *Eléments de Syntaxe Structurale*, Klincksieck.

J. Vergne, 1990. "A parser without a dictionary as a tool for research into French syntax", *Actes 13° COLING*, Helsinki.

J. Véronis, 1988. *Le traitement de l'erreur dans un dialogue homme-machine*, Thèse de l'université d'Aix-Marseille.

K. Vijay-Shanker, 1987. *A Study of Tree Adjoining Grammar*, Thèse de PhD, Université de Pennsylvanie, Philadelphie.

K. Vijay-Shanker, A. Joshi, 1988. "A Feature-based Tree Adjoining Grammar", *Actes 12° COLING*, Budapest. (à paraître in *Journal of Computational Linguistics*, 1991).

R. Vivès, 1983. *Avoir, prendre et perdre: constructions à verbe support et extensions aspectuelles*, Thèse de 3ème cycle, Université Paris 8.

T. Wasow, 1977. "Transformations and the lexicon", in P. Culicover et al. (eds), *Formal syntax*, Academic press, New York.

T. Wasow, Sag I., Nunberg G., 1982. "Idioms: an interim report", *Actes 13° International Congress of Linguists*, Tokyo.

E. Wehrli, 1984. "A Gouvernement-Binding parser for French", Working paper n° 48, Institut pour les Etudes Sémantiques et Cognitives, Université de Genève.

E. Wehrli, 1988. "Parsing with a GB grammar", in U. Reyle, C. Rohrer, *Natural Language parsing and linguistic theories*, Dordrecht.

D. Weir, 1988. *Characterizing mildly context-sensitive grammar formalisms*, thèse de PhD, Université de Pennsylvanie, Philadelphie.

R.S. Wells, 1947. "Immediate constituents", *Language*, n°23. (traduit dans *Langages*, n°20, 1970).

A. Wierzbicka, 1982. "Why you can have a drink while you can't * have an eat", *Language*, 58:4.

E. Williams, 1980. "Predication", *Linguistic Inquiry*, n°11.

D. Younger, 1967. "Recognition and parsing of context-free languages in time n^3 ", *Information and Control*, 10-2.

A. Zribi-Hertz, 1982. "La construction se-moyen et son statut dans le triangle: passif-moyen-réfléchi", *Linguisticae Investigationes*, VI-2, J.Benamins.

A. Zribi-Hertz, 1986. *Relations anaphoriques en français: esquisse d'une grammaire générative raisonnée de la réflexivité et de l'ellipse structurale*, Thèse de Doctorat d'Etat, Université Paris 8.

P. Zweigenbaum, M. Cavazza, 1990. "Deep sentence understanding in a restricted domain", *Actes 13° COLING*, Helsinki.

ANNEXE 1

Symboles des traits et catégories utilisés

La distinction entre traits lexicaux (portant sur des préterminaux) et traits syntaxiques (portant sur d'autres catégories) n'est pas valable en TAG. On distingue les traits associés à un noeud et ceux associés à un arbre élémentaire considéré comme un tout. Les deux types de traits figurent dans les arbres (indépendamment de leur tête lexicale : contraintes structurales) et dans le dictionnaire (contraintes lexicales).

Catégories de noeuds

Prep : préposition (ou conjonction de subordination)
 D : déterminant
 Cl : clitique
 V : verbe
 P : phrase
 N : nom ou groupe nominal
 SP : syntagme prépositionnel
 A : adjectif ou groupe adjectival
 Adv : adverbe ou groupe adverbial
 C : complémenteur (que, de, si, à)
 Conj : conjonction de coordination
 N₀ : premier argument (sujet)
 N_i : argument nominal
 P_i : argument phrastique

Traits associés à un noeud

	valeur	noeuds
qu	+, -	N, P
inv	+, -	V, P
princ	+, -	P
num	+, -	N, D, A, V
genre	+, -	N, D, A, V
pers	+, -	N, V
hum	+, -	N
abs	+, -	N
conc	+, -	N
mass	+, -	N
det	+, -	N
def	+, -	N
indef	+, -	N
part	+, -	N

poss	+, -	N
quant	+, -	N
modif	+, -	N, V, SP, A
loc	+, -	SP, Adv, N
man	+, -	SP, Adv
nom	+, -	V
acc	+, -	V
refl	+, -	V
dat	+, -	V
dat-h	+, -	V
gen	+, -	V
aux	être, avoir	V
mode	ind, imp, subj, inf, ppart	V, P

Traits associés à un arbre

	valeur	familles libres	familles figées
<i>passif</i>	+, -	n0Vn1, n0Vn1ps2, n0Vn1pn2	n0Vn1PN2, n0VDN1, n0VDN1PN2, n0VdN1, n0VNpn1, n0VDNpn1
<i>pass-aux</i>	+, -	n0Vs1	
<i>extrap</i>	+, -	n0V, n0Vpn1, n0Vps1	n0VDN, DN0V, DN0Vpn1 (et les arbres passifs des familles ci-dessus)
<i>extrap-pass</i>	+, -	n0Vpn1, n0Vps1	
<i>clivée₁</i>	+, -		n0VDNpn1
<i>sym</i> n0VdNpn1	+	n0Vpn1, n0ClVpn1, s0Vn1pn2, n0Vs1pn2	n0VPNpn1,
<i>sym</i>	0, 1	n0Vn1pn2	
<i>neut</i>	+, -	n0Vn1, n0Vn1pn2	n0VDN1

ANNEXE 2

FAMILLES D'ARBRES

LISTE DES FAMILLES UTILISEES

1. Phrases à verbe simple (25 familles)

1.1 avec arguments nominaux

n0V:	Jean dort
n0Vn1:	Jean regarde Marie
n0CIV:	Marie s'est évanouie
n0CIVn1:	Cet enfant s'appelle Marie
n0Vpn1:	Jean compte sur Marie
n0CIVpn1:	Jean se bat contre Paul
n0Vn1pn2:	Jean donne un livre à Marc
n0Vpn1pn2:	Jean parle de ce livre avec Marie
n0Vn1n2:	On a élu Jean président

1.2 avec arguments phrastiques

s0V:	(De la neige en août + Qu'il neige en août) arrive quelque fois
s0Vn1:	(Le départ de Jean + Que Jean soit parti) désole Marie
s0Vpn1:	(Le départ de Jean + Que Jean soit parti) déplaît à Marie
n0Vs1:	Jean trouve qu'il fait froid
s0Vs1:	(Cet épais tapis de neige + Qu'il neige) prouve qu'il fait froid
n0CIVps1:	Jean se défend d'avoir dit cela
n0Vps1:	Jean rêve d'aller à Bilbao
n0Vpn1s2:	Jean dit à Marie qu'il fait beau
n0Vn1ps2:	Jean a convaincu Marie de partir
n0Vs1ps2:	Jean tire qu'il fait froid de ce qu'il neige
s0Vn1ps2:	(La neige + Qu'il neige) habitue Jean à ce qu'il fasse froid
n0Vpn1ps2:	Jean convient avec Marie d'aller à la piscine
n0Vps1ps2:	Jean argue de ce qu'il neige pour partir plus tôt

1.3 avec autres arguments

n0Vpp1 :	Jean va (à Paris + chez lui)
n0Vn1pp2:	Jean met la boîte (sur la table + dans sa poche)
n0Vn1a:	Jean trouve Marie fatiguée
n0Vas2:	Jean trouve heureux que Paul soit parti
n0Vad:	Jean va bien
s0CIVad:	(Le départ de Marie + Que Marie soit partie) s'est bien passé.

2. Phrases copulatives (6 familles)

n0Va:	Jean est joyeux
s0Va:	(Le départ de Marie + Que Marie parte) était imprévu

n0Vapn1: Jean est content de son nouveau chapeau
 n0Vaps1: Jean est impatient de partir
 s0Vapn1: (Le départ de Marie + Que Marie parte) est dépendant de Jean
 s0Vaps1: (La neige + Qu'il neige) est susceptible de faire des heureux

3. Phrases à verbe support (15 familles)

3.1 avec arguments nominaux

n0VDN: Jean a fait une gaffe
 n0VdNpn1: Jean a commis (une + cette) agression contre Marie hier soir
 n0VDNpn1: Paul a fait la rencontre de Jean hier soir
 n0VDN1pn2: Paul fait la cour à Marie
 n0VPNpn1: Paul est en colère contre Marie
 n0VPDNpn1: Jean est dans l'ignorance de cet événement
 n0VDNpn1pn2: Jean fait l'analyse de la situation à Marie

3.2 avec arguments phrastiques

s0VN: (Le départ de Marie + Que Marie soit partie) a fait scandale
 n0VDNps1: Jean a l'espoir de partir demain
 n0VNps1: Jean a envie de partir demain
 n0VDNps1pn2: Jean a donné l'ordre à Luc de partir demain
 s0VPDNpn1: (Le départ de Marie + Que Marie soit partie) est à l'avantage de Paul

4. Phrases figées (22 familles)

4.1. A sujet figé

N0Vs1: Il faut que Marie parte
 N0V: Il pleut
 N0CIVs1: Il s'agit de faire vite
 DN0Vpn1: Des ailes poussent à Jean
 DN0Vn1: La grâce touche Jean
 DN0VPDNpn1: La moutarde monte au nez de Paul

4.2. Autres compléments figés

n0VDN1: Jean a cassé sa pipe
 n0VPDN1: L'affaire est tombée à l'eau
 n0VPN: Jean a battu en retraite
 n0VdN1: Marie attend un enfant
 n0VDN1PDN2: Jean prend le taureau par les cornes
 n0VDN1PN2: Jean bâtit des châteaux en Espagne
 n0VPN1ps2: Jean est de taille à faire cela
 n0Vn1PN2: Jean prend la situation en mains

n0Vn1PDN2: Ils ont passé Luc par les armes
n0VPDAN1: Marie était dans ses petits souliers
n0VPDNpn1: Paul marche sur les brisées de Max
n0VDN1PDNpn2: Jean apporte de l'eau au moulin de Max
n0VDAN1: Jean boit du petit lait
n0VN1pn2: Jean a eu vent de cette affaire
n0VDN1pn2: Jean donne le ton aux soirées qu'il fréquente
n0Va: Jean devient chèvre

EXEMPLES DE FAMILLES

FAMILLE N0V

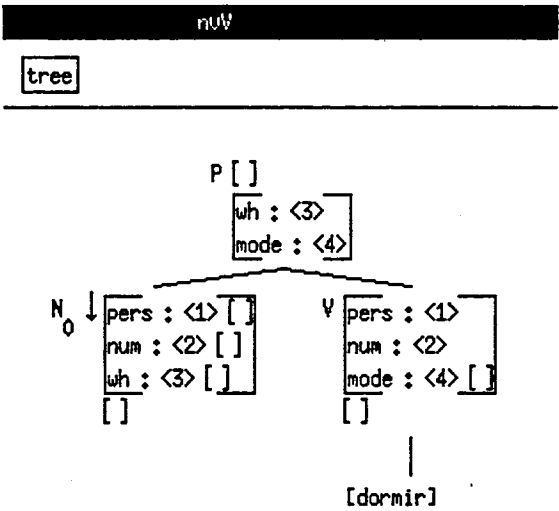
entrée lexicale: [dormir],V: n0V{N0:<hum> = +,extrap=-}.

Arbres élémentaires	Traits	Exemples de phrases
n0V		<i>Jean dort depuis midi</i>
R0n0V	relative	<i>Je connais la personne qui dort ici</i>
C0n0V	clivée	<i>C'est Jean qui dort ici</i>

ARBRES ELEMENTAIRES DE LA FAMILLE N0V

entrée lexicale: [dormir],V: n0V{N0:<hum> = +,extrap=-}.

n0V



FAMILLE n0Vn1

entrée lexicale: [fabriquer], V:

n0Vn1{N0: <det> = +, N1: <conc> = +, N1: <det> = +, neut = -}.

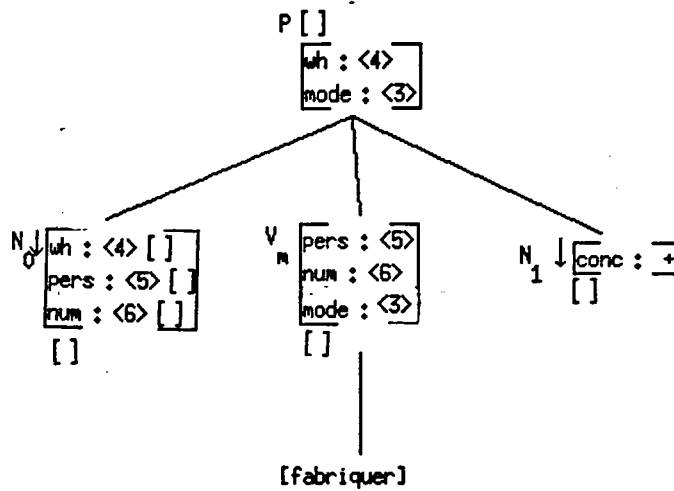
Arbres élémentaires	Traits	Exemples de phrases
n0Vn1		<i>Jean fabrique des jouets</i>
Wn0Vn1	question	<i>Qu'est-ce que Jean fabrique?</i>
R0n0Vn1	relative	<i>Je connais la personne qui fabrique ces jouets</i>
R1n0Vn1	relative	<i>J'aime les jouets que Jean fabrique</i>
C0n0Vn1	clivée	<i>C'est Jean qui fabrique des jouets</i>
C1n0Vn1	clivée	<i>Ce sont des jouets que Jean fabrique</i>
n0Vn1(pass)	passif	<i>Des jouets sont fabriqués par Jean</i>
n0Vn1(extr) dernier	extraposition + passif	<i>Il a été fabriqué trois tonnes de jouets le mois dernier</i>
Wn0Vn1(pass)	question + passif	<i>Par qui des jouets sont-ils fabriqués?</i>
R0n0Vn1(pass)	relative + passif	<i>J'aime les jouets qui sont fabriqués par Jean</i>
R1n0Vn1(pass) fabriqués	relative + passif	<i>Je connais l'homme par qui ces jouets sont fabriqués</i>
C0n0Vn1(pass)	clivée + passif	<i>Ce sont des jouets qui sont fabriqués par Jean</i>
Cp1n0Vn1(pass)	clivée + passif	<i>C'est par Jean que des jouets sont fabriqués</i>

ARBRES ELEMENTAIRES DE LA FAMILLE N0VN1

entrée lexicale: [fabriquer], V:

n0Vn1{N0:<det> = +, N1:<det> = +, N1:<conc> = +, neut = -}.

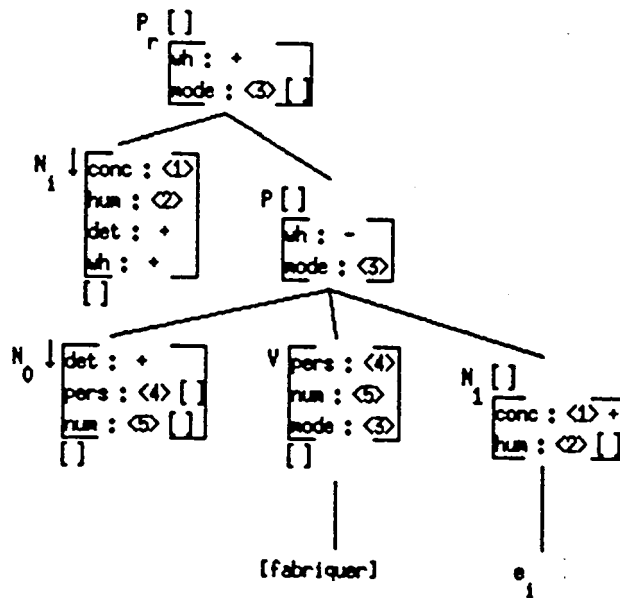
n0Vn1



Jean fabrique des jouets

Question

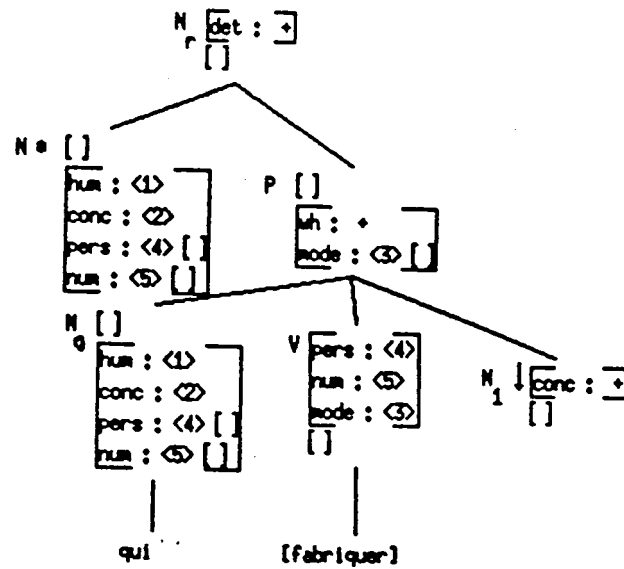
Wn0Vn1



Qu'est-ce que Jean fabrique?

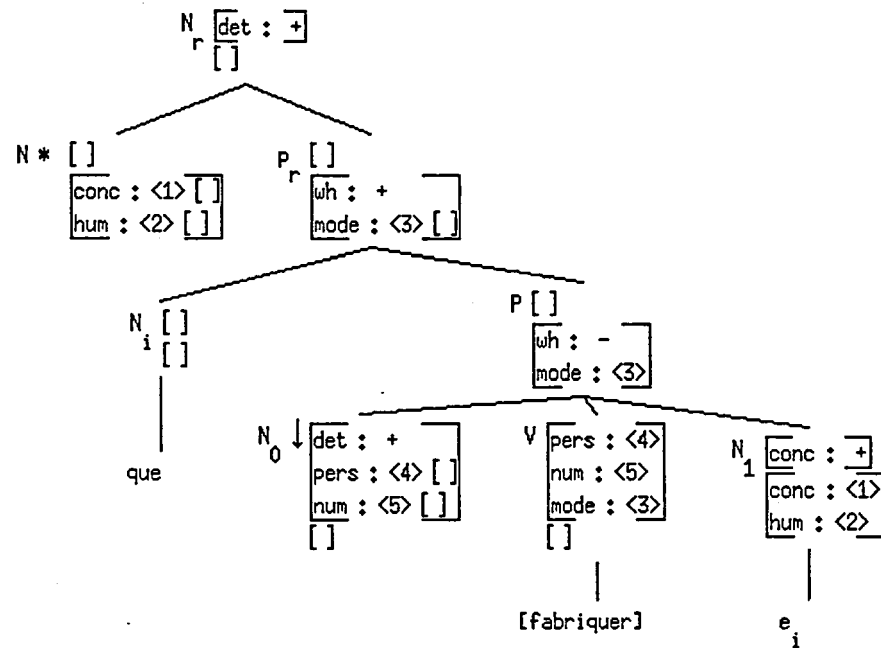
Relatives

R0n0Vn1



Je connais la personne qui fabrique ces jouets

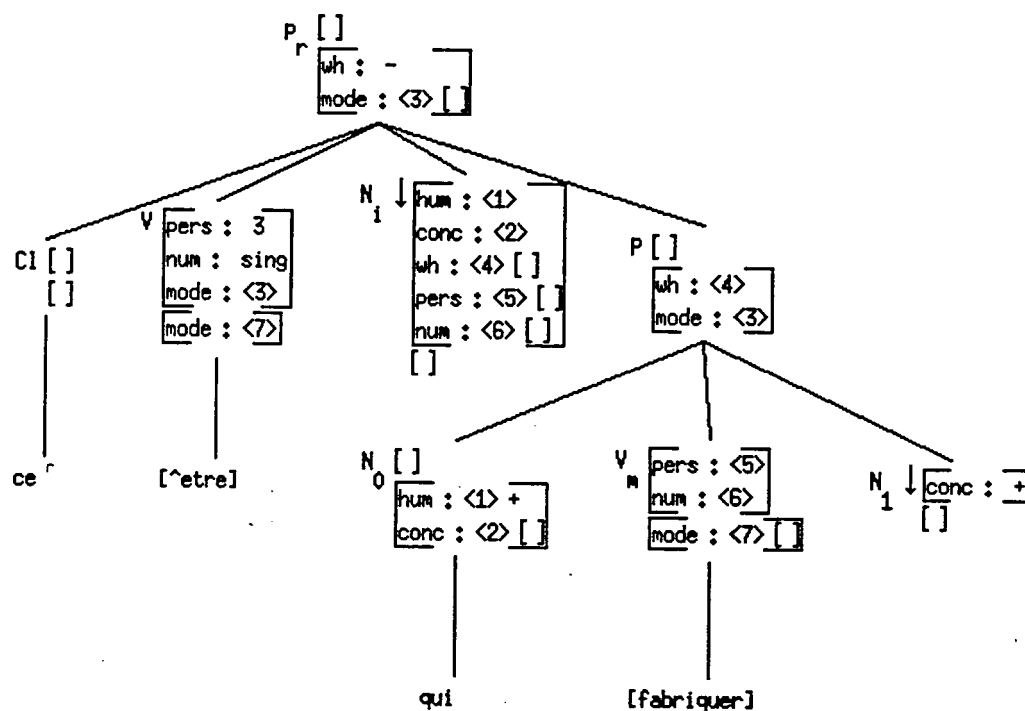
R1n0Vn1



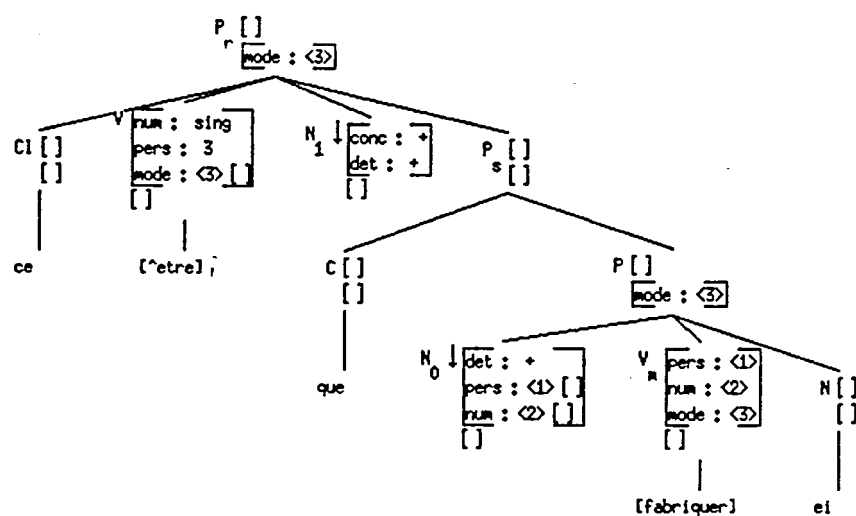
J'aime les jouets que Jean fabrique

Clivées

C0n0Vn1

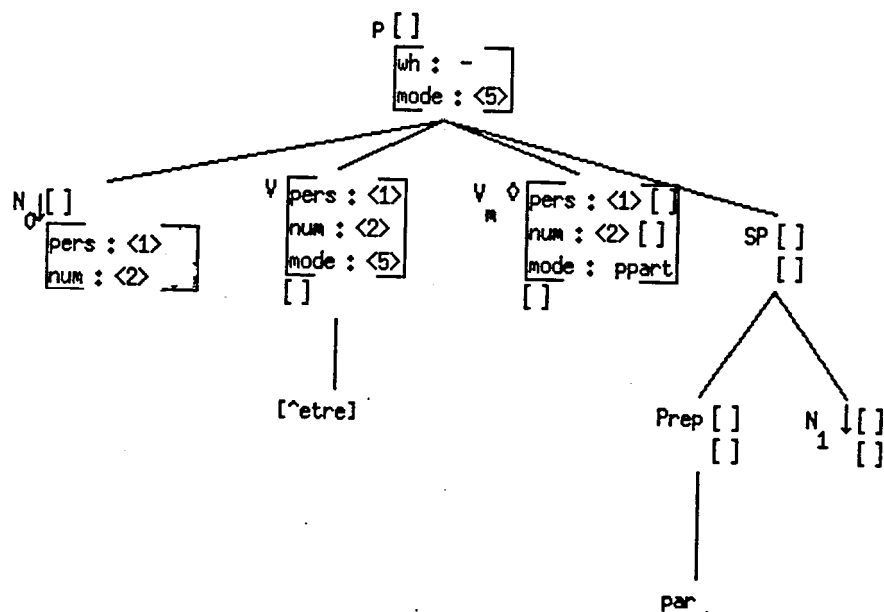
*C'est Jean qui fabrique des jouets*

C1n0Vn1

*Ce sont des jouets que Jean fabrique*

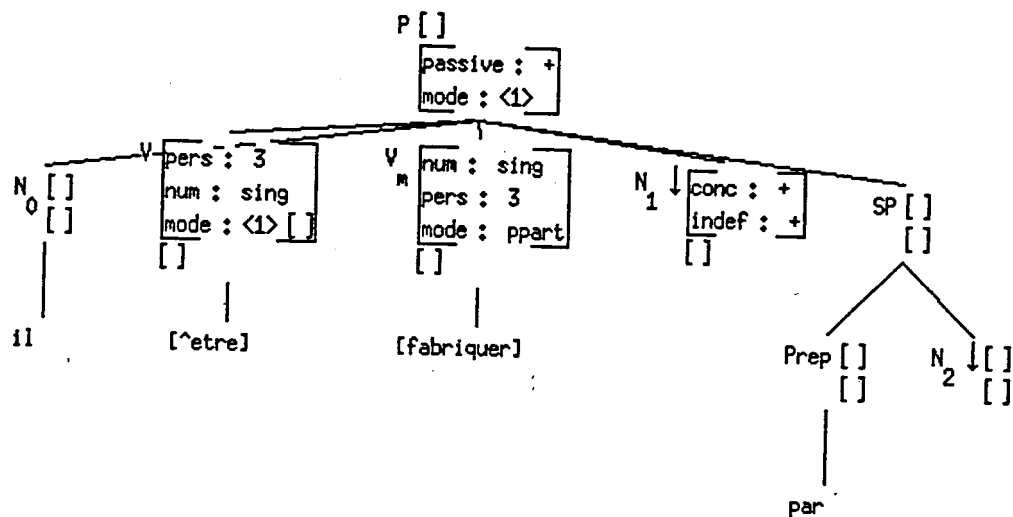
Passif

n0Vn1(pass)



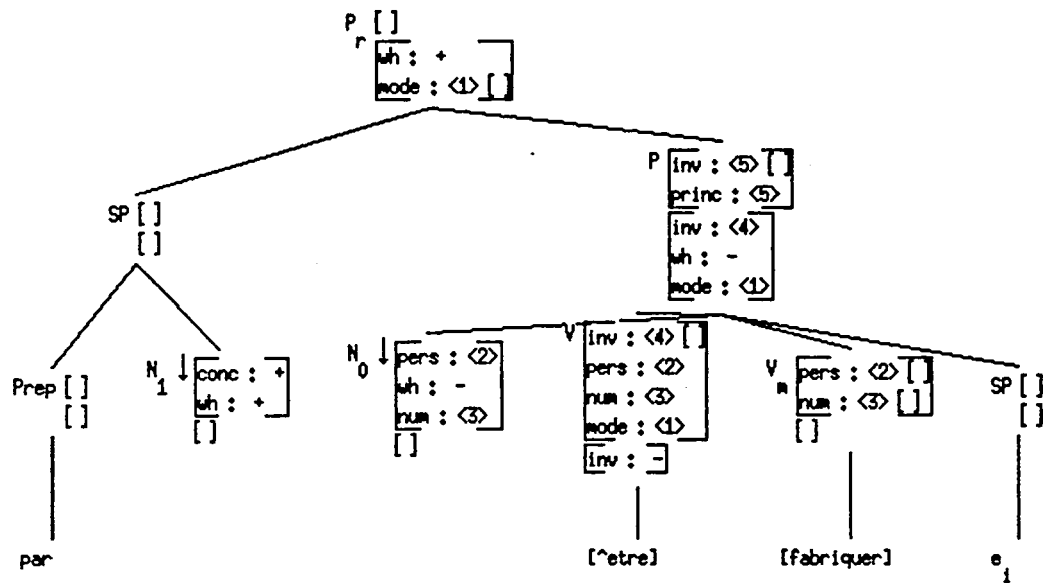
Des jouets sont fabriqués par Jean

n0Vn1(extr)



Il a été fabriqué trois tonnes de jouets le mois dernier

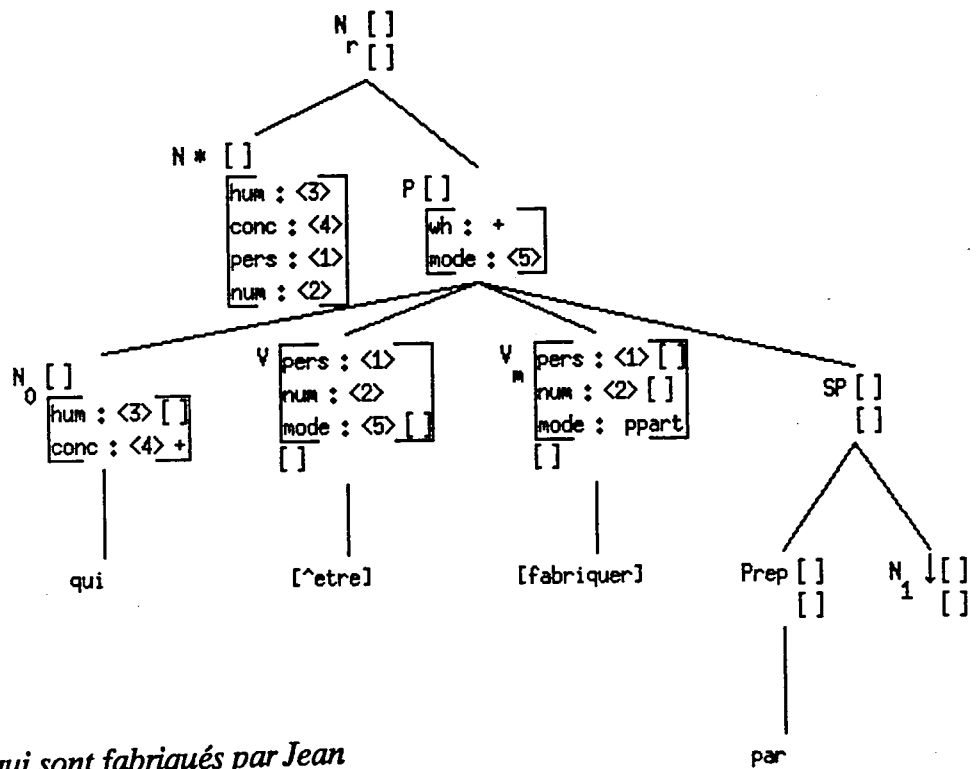
Question Wn0Vn1(pass)



Par qui des jouets sont-ils fabriqués?

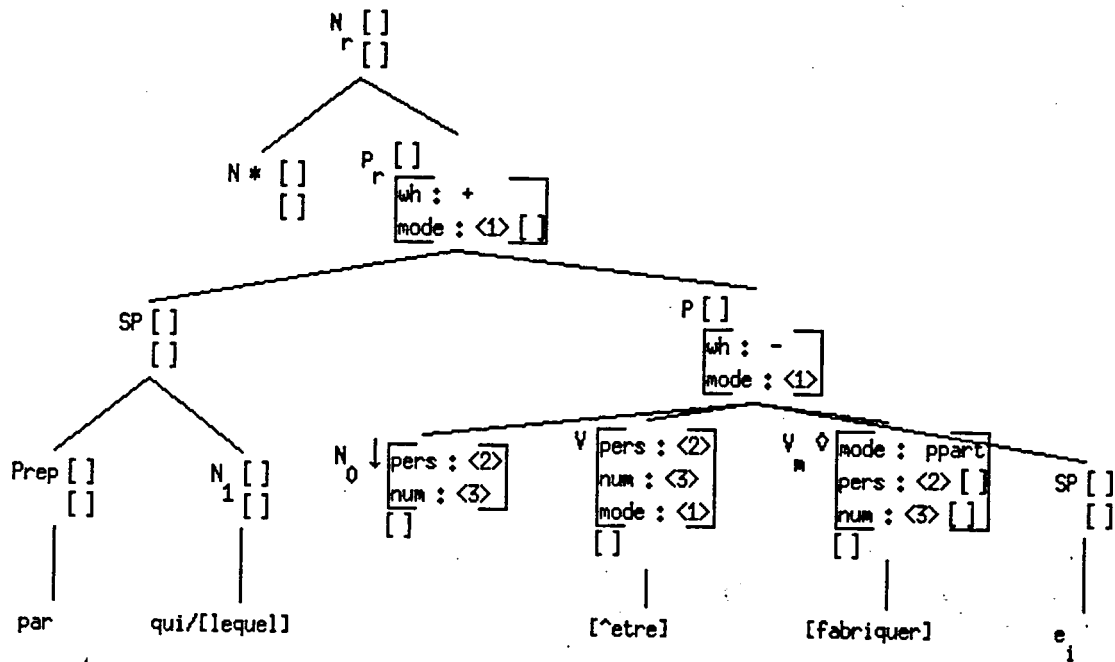
Relatives

R0n0Vn1(pass)



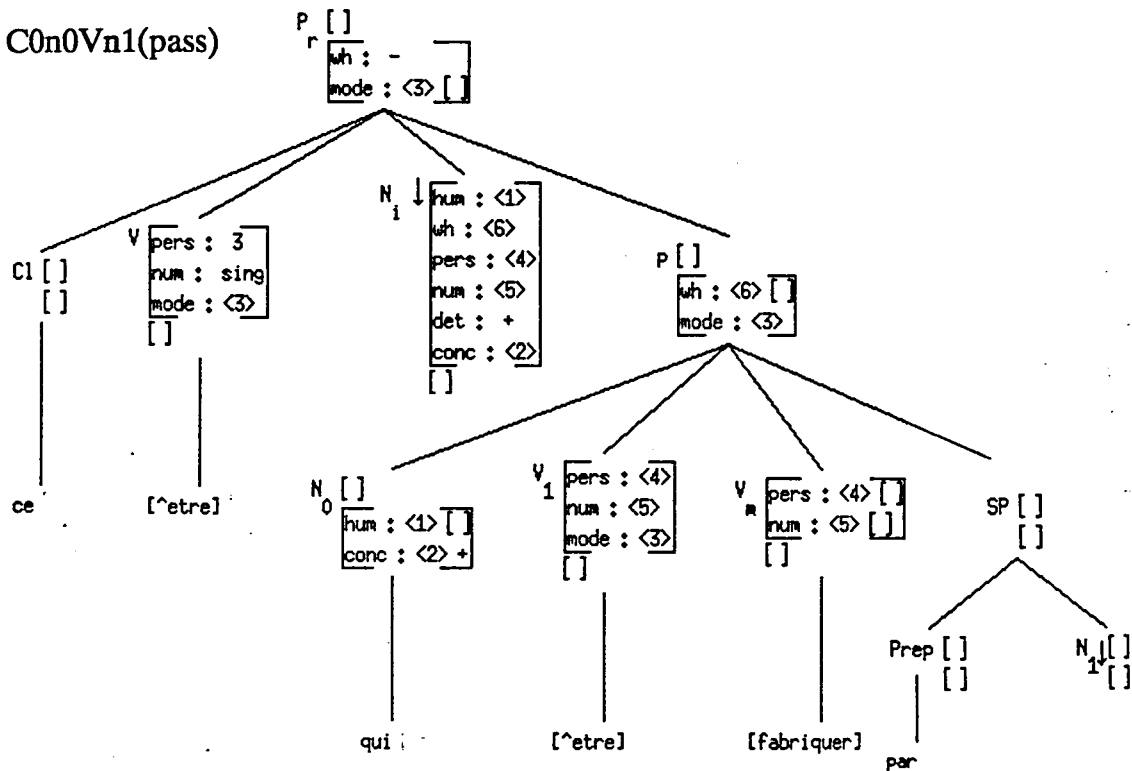
J'aime les jouets qui sont fabriqués par Jean

R1n0Vn1(pass)

*Je connais l'homme par qui ces jouets sont fabriqués*

Clivées

C0n0Vn1(pass)

*Ce sont des jouets qui sont fabriqués par Jean*

FAMILLE n0VdNpn1

entrée lexicale :

[agression],N;[commettre],V;contre,P:

n0VdN(pn1){N0:<hum> = +,N0:<det> = +,N1:<det> = +,sym=-}.

Arbres élémentaires	Traits	Exemples de phrases
n0VdNpn1		<i>Jean commet une agression contre Marie</i>
Npn0pn1	nominalisation	<i>Je désapprouve l'agression de Luc contre Marie</i>
Wn0VdNpn1(1)	question	<i>Quelle agression Jean a-t-il commise contre Marie ?</i>
Wn0VdNpn1(2)	question	<i>Quelle agression contre Marie Jean a-t-il commise ? (ou Une agression contre qui Jean a-t-il commise ?)</i>
W1n0VdNpn1	question	<i>Contre qui Jean a-t-il commis une agression ?</i>
R0n0VdNpn1	relative	<i>Je connais l'homme qui a commis une agression contre Marie</i>
Rn0VdNpn1(1)	relative	<i>Je désapprouve l'agression contre Marie que Jean a commise</i>
Rn0VdNpn1(2)	relative	<i>Je désapprouve l'agression que Jean a commise contre Marie</i>
R1n0VdNpn1	relative	<i>Je connais la fille contre qui Jean a commis une agression</i>
C0n0VdNpn1	clivée	<i>C'est Jean qui a commis une agression contre Marie</i>
Cn0VdNpn1(1)	clivée	<i>C'est une agression contre Marie que Jean a commise</i>
Cn0VdNpn1(2)	clivée	<i>C'est une agression que Jean a commise contre Marie</i>
C1n0VdNpn1	clivée	<i>C'est la fille contre laquelle Jean a commis une agression</i>
Cp1n0VdNpn1	clivée	<i>C'est contre Marie que Jean a commis une agression</i>
n0VdNpn1(pass)	passif	<i>Une agression contre Marie a été commise par Jean</i>
n0VdNpn1(extr)	extrap + passif	<i>Il a été commis une agression contre Marie</i>
Wn0VdNpn1(pass1)	question+passif	<i>Quelle agression a été commise par Jean contre Marie ?</i>
Wn0VdNpn1(pass2)	question+passif	<i>Quelle agression contre Marie a été commise par Jean ? (ou Une agression contre qui a été commise ?)</i>
W0n0VdNpn1(pass)	question+passif	<i>Contre qui une agression a-t-elle été commise par Jean ?</i>
W1n0VdNpn1(pass1)	question+passif	<i>Par qui une agression a -t-elle été commise contre Marie ?</i>
W1n0VdNpn1(pass2)	question+passif	<i>Par qui une agression contre Marie a -t-elle été commise ?</i>
Rn0VdNpn1(pass1)	relative+passif	<i>Je désapprouve l'agression contre Marie qui a été commise par Jean</i>
Rn0VdNpn1(pass2)	relative+passif	<i>Je désapprouve l'agression qui a été commise par Jean contre Marie</i>

R0n0VdNpn1(pass) relative+passif
commise par Jean

R1n0VdNpn1(pass1) relative+passif
commise contre Marie

R1n0VdNpn1(pass2) relative+passif
Marie a été commise

Cn0VdNpn1(pass1) clivée+passif
commise par Jean

Cn0VdNpn1(pass2) clivée+passif
contre Marie

C0n0VdNpn1(pass) clivée+passif
commise par Jean

C1n0VdNpn1(pass) clivée+passif
contre Marie

Cp0n0VdNpn1(pass) clivée+passif
commise par Jean

Je connais la fille contre qui une agression a été

Je connais l'homme par qui une agression a été

Je connais l'homme par qui une agression contre

C'est une agression contre Marie qui a été

C'est une agression qui a été commise par Jean

C'est la fille contre laquelle une agression a été

C'est par Jean qu'une agression a été commise

C'est contre Marie qu'une agression a été

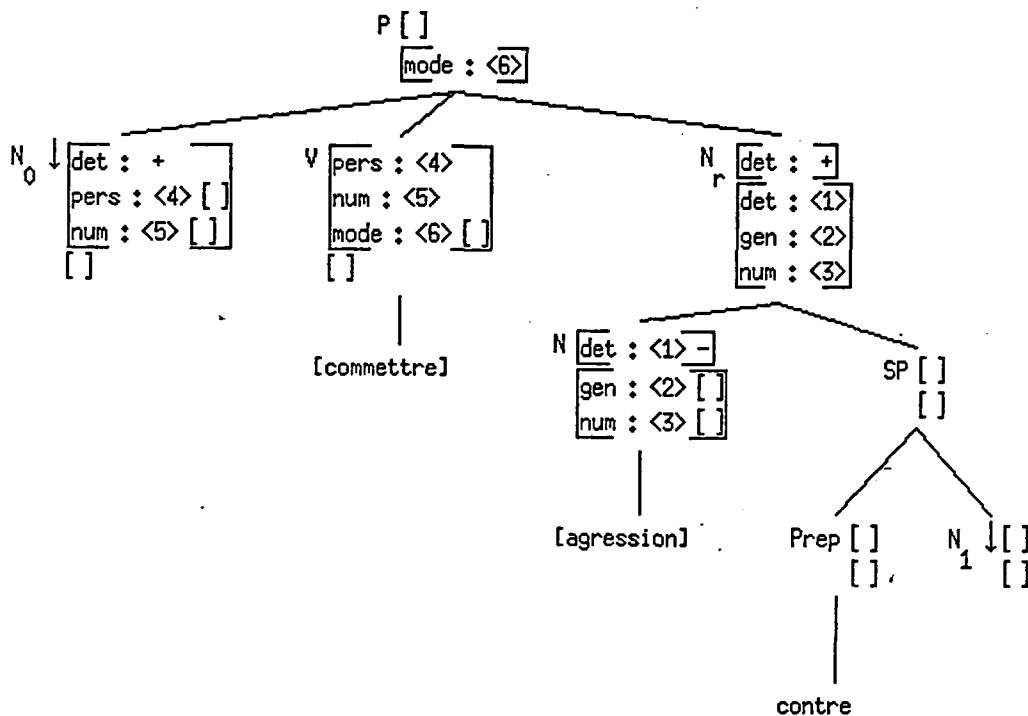
ARBRES ELEMENTAIRES DE LA FAMILLE N0VDNPN1

entrée lexicale :

[agression],N;[commettre],V;contre,P:

n0VdN(pn1){N0:<hum> = +,N1:<det> = +,N1:<hum> = +,sym=-}.

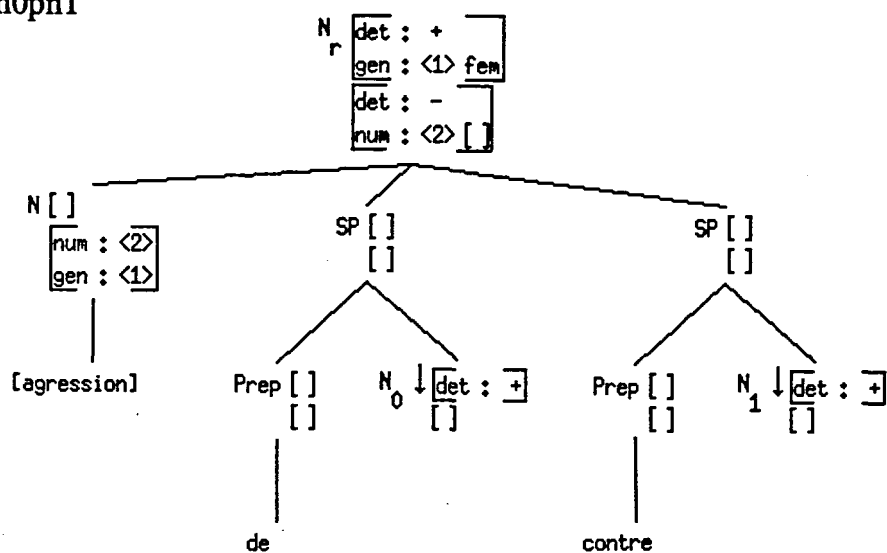
n0VdNpn1



Jean commet une agression contre Marie

SN complexe

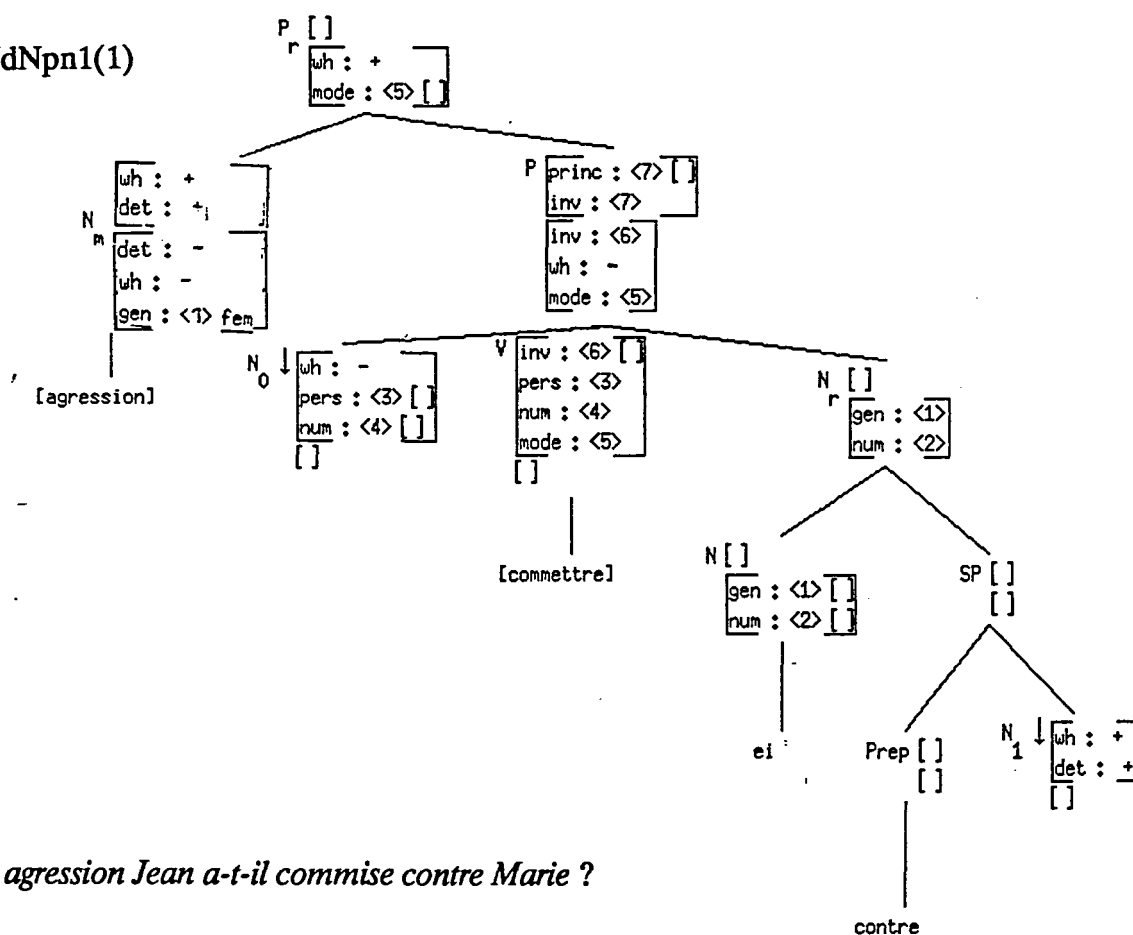
Npn0pn1



Je désapprouve l'agression de Luc contre Marie

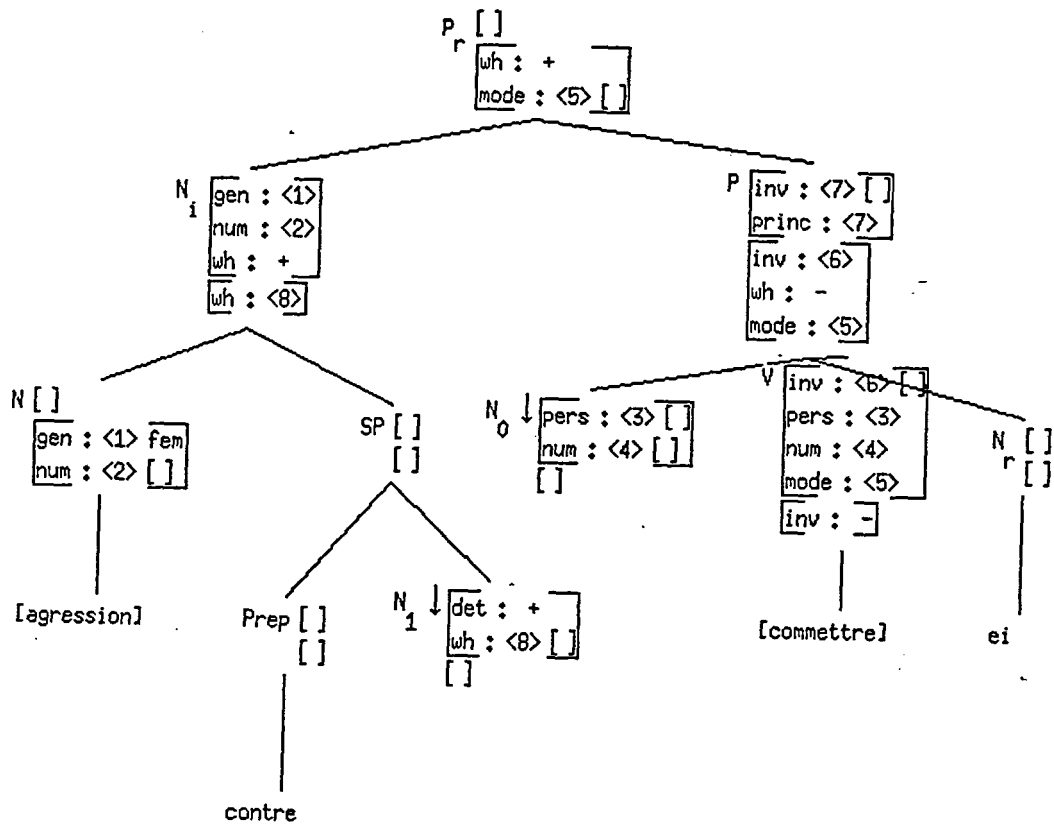
Questions

Wn0VdNpn1(1)



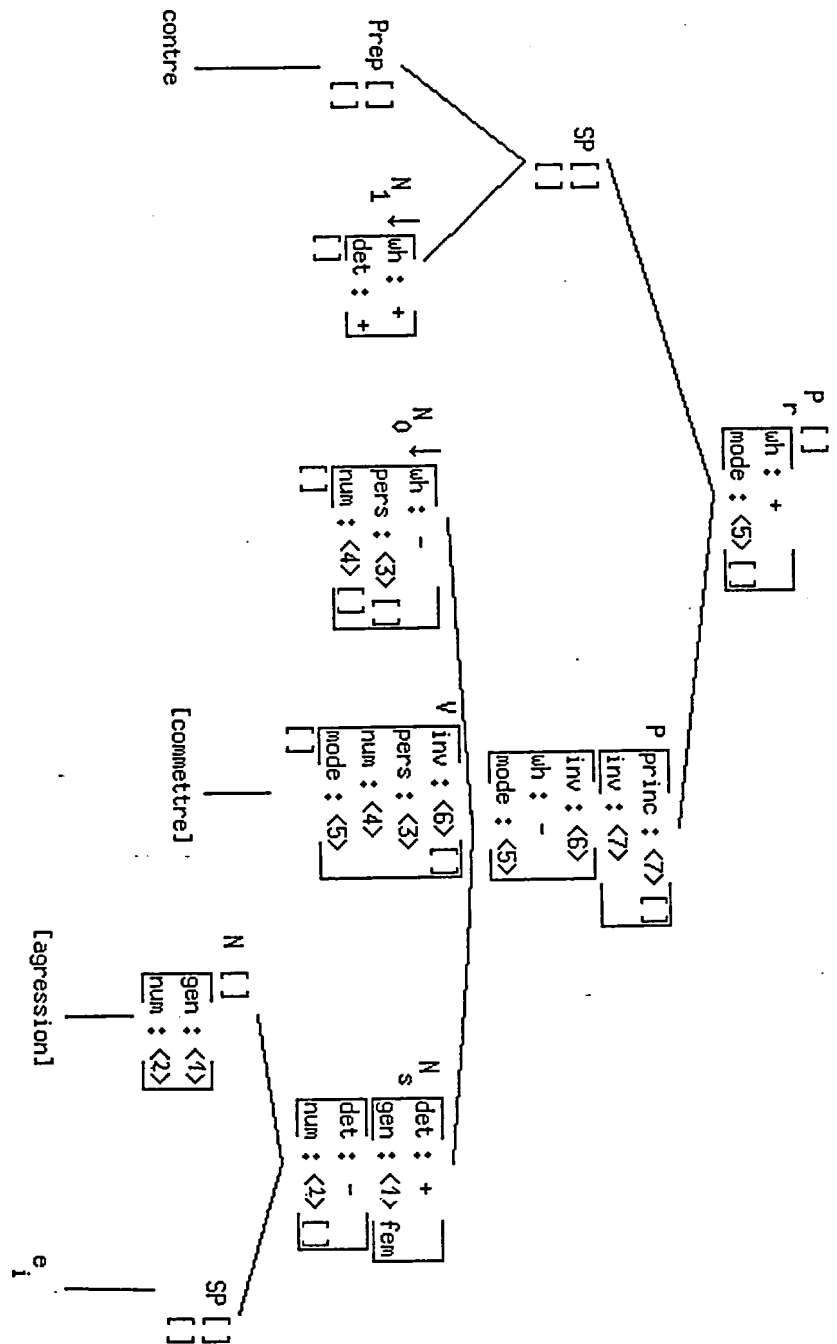
Quelle agression Jean a-t-il commise contre Marie ?

Wn0VdNpn1(2)



*Quelle agression contre Marie Jean a-t-il commise ? (ou
Une agression contre qui Jean a-t-il commise ?)*

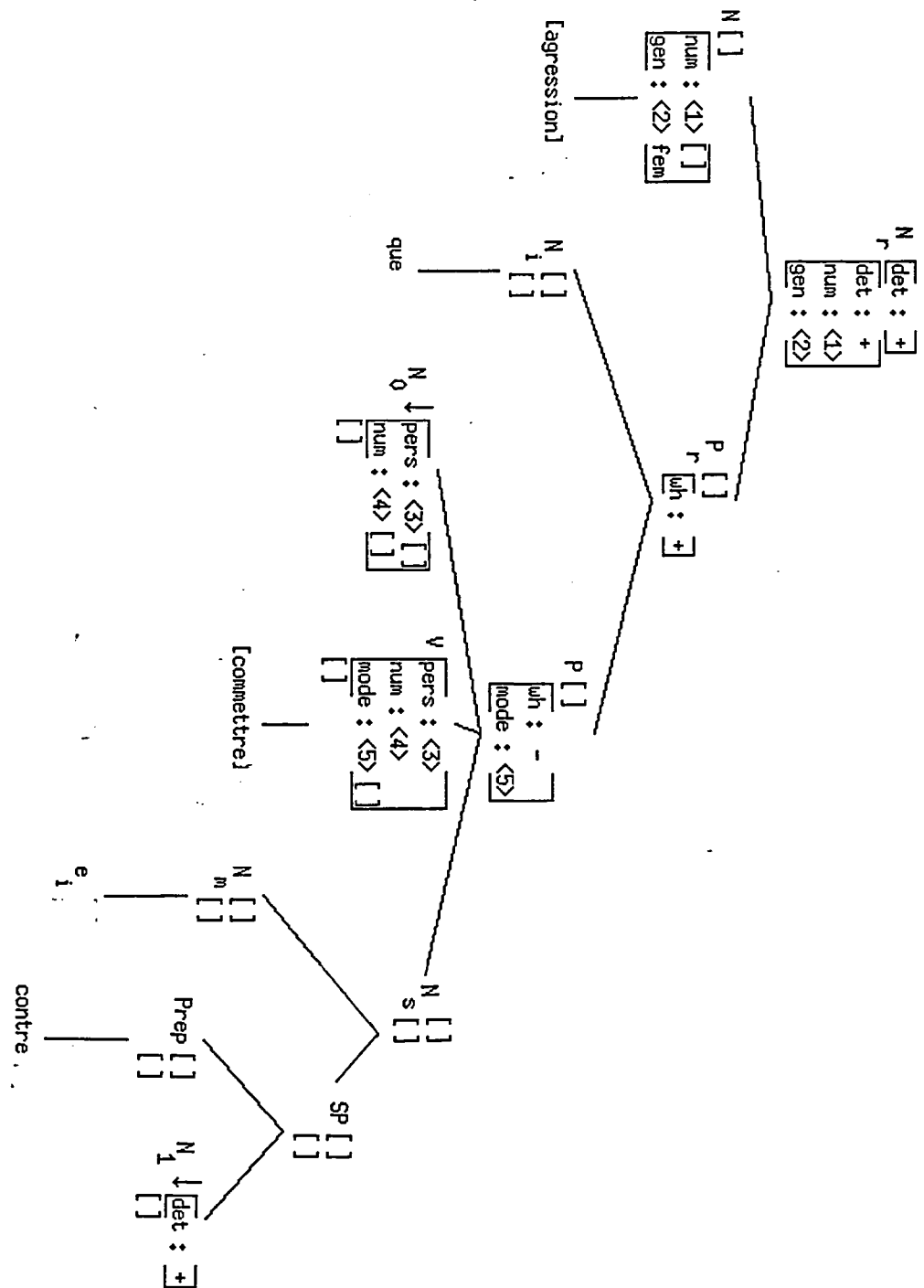
W1n0VdNpn1



Contre qui Jean a-t-il commis une agression ?

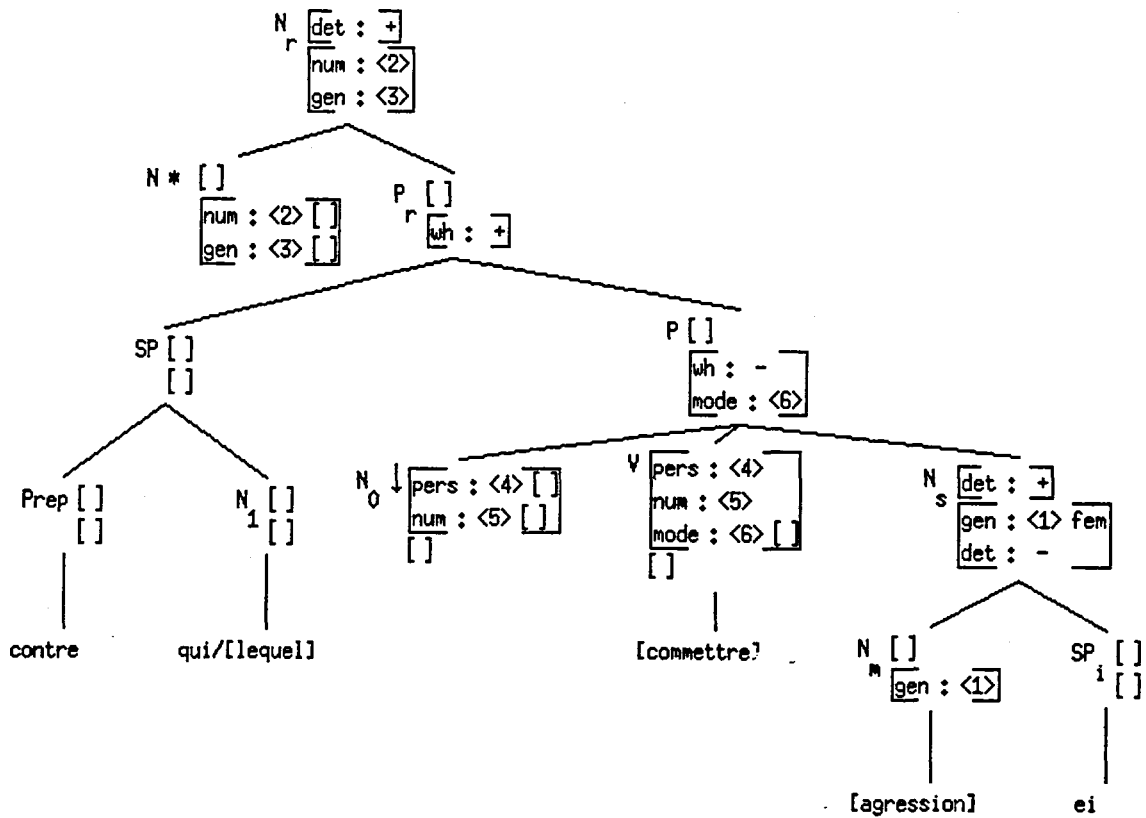
Relatives

Rn0VdNpn1(1)



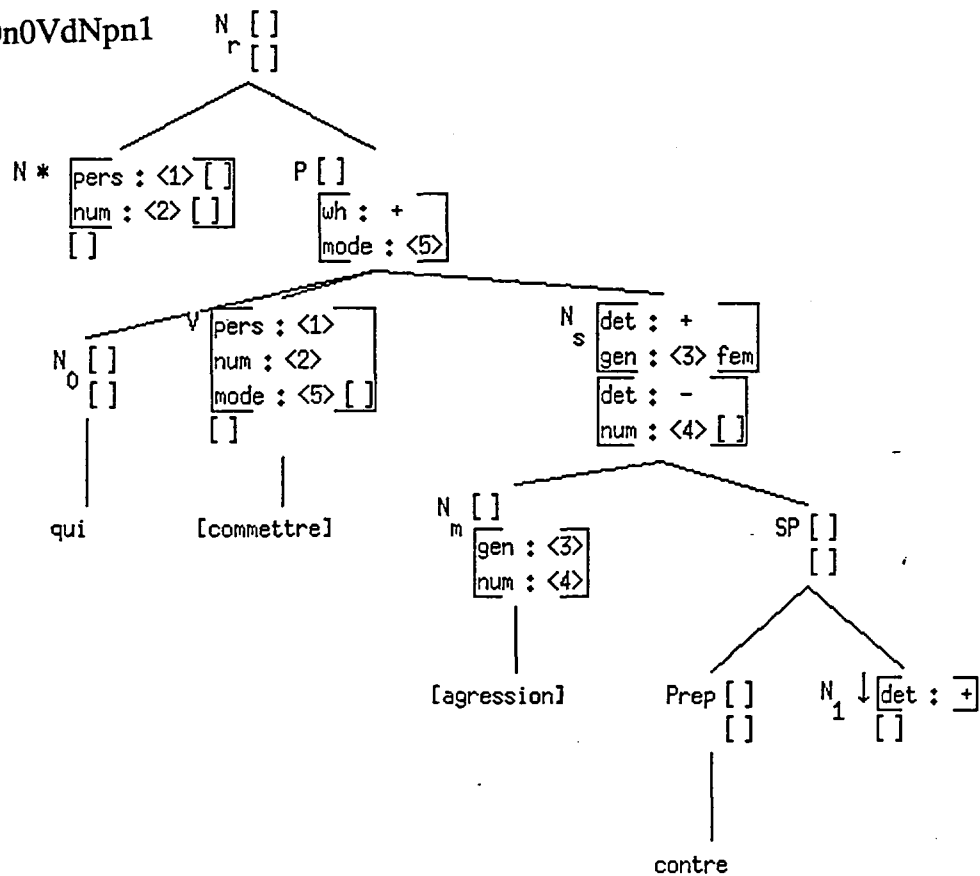
Je désapprouve l'agression que Jean a commise contre Marie

R1n0VdNpn1



Je connais la fille contre qui Jean a commis une agression

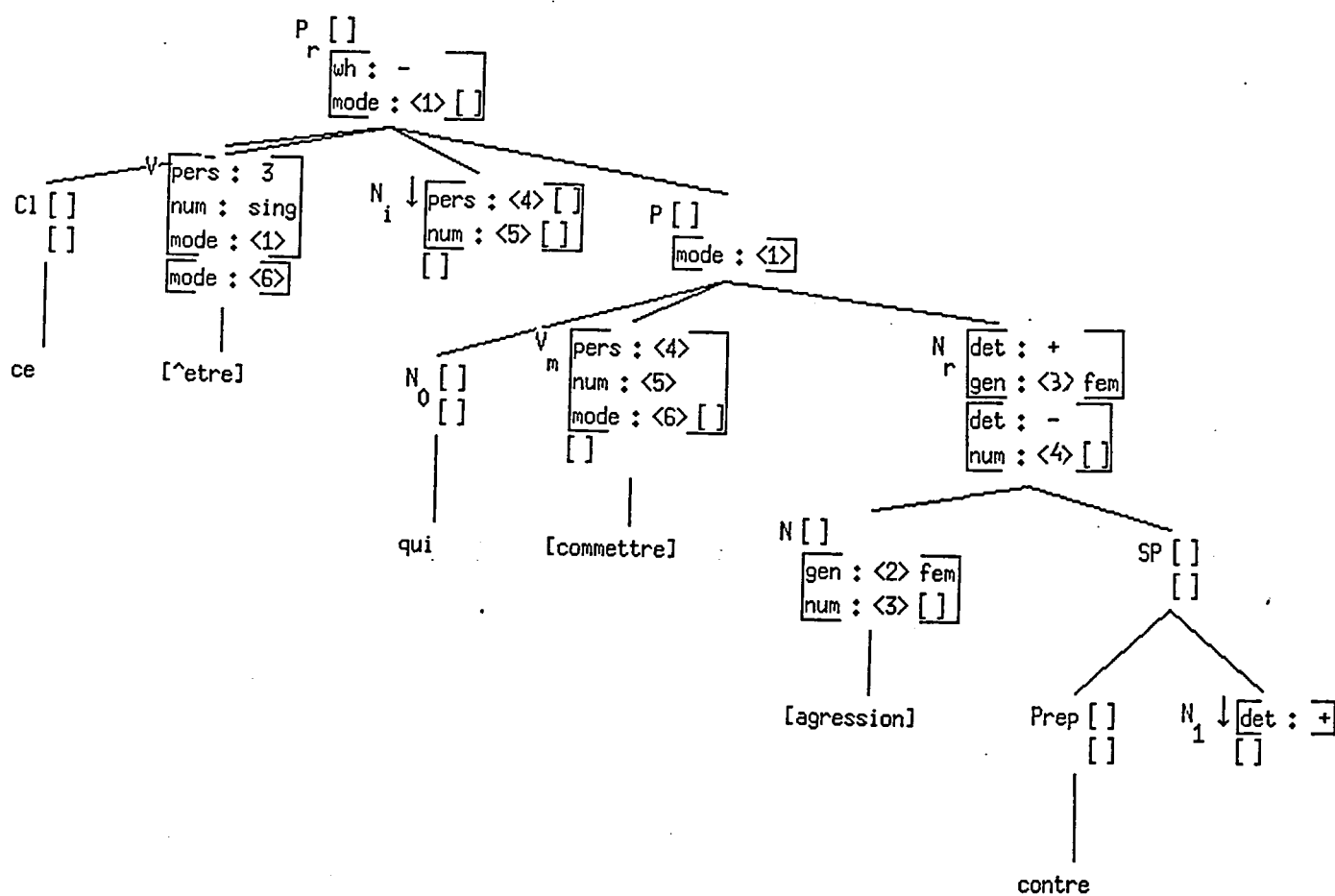
R0n0VdNpn1



Je connais la personne qui a commis une agression contre Marie

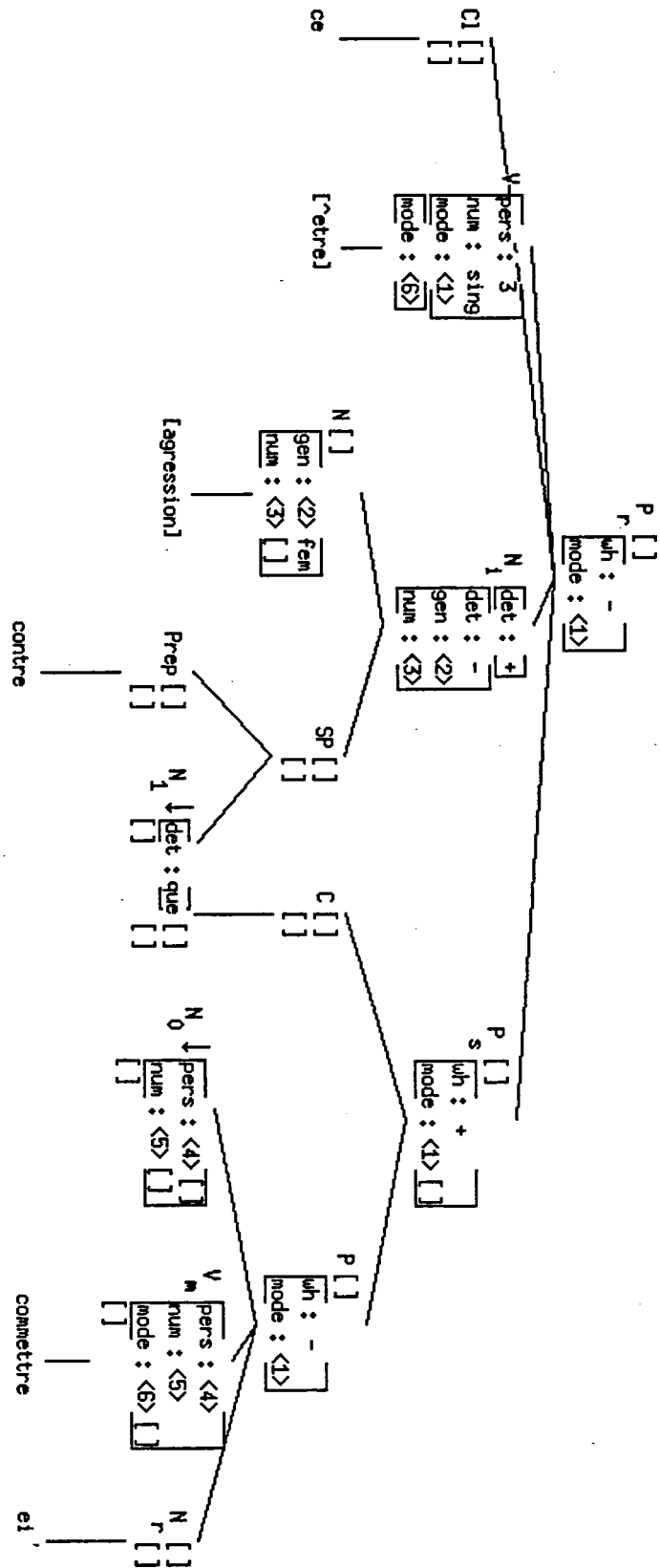
Clivées

C0n0VdNpn1



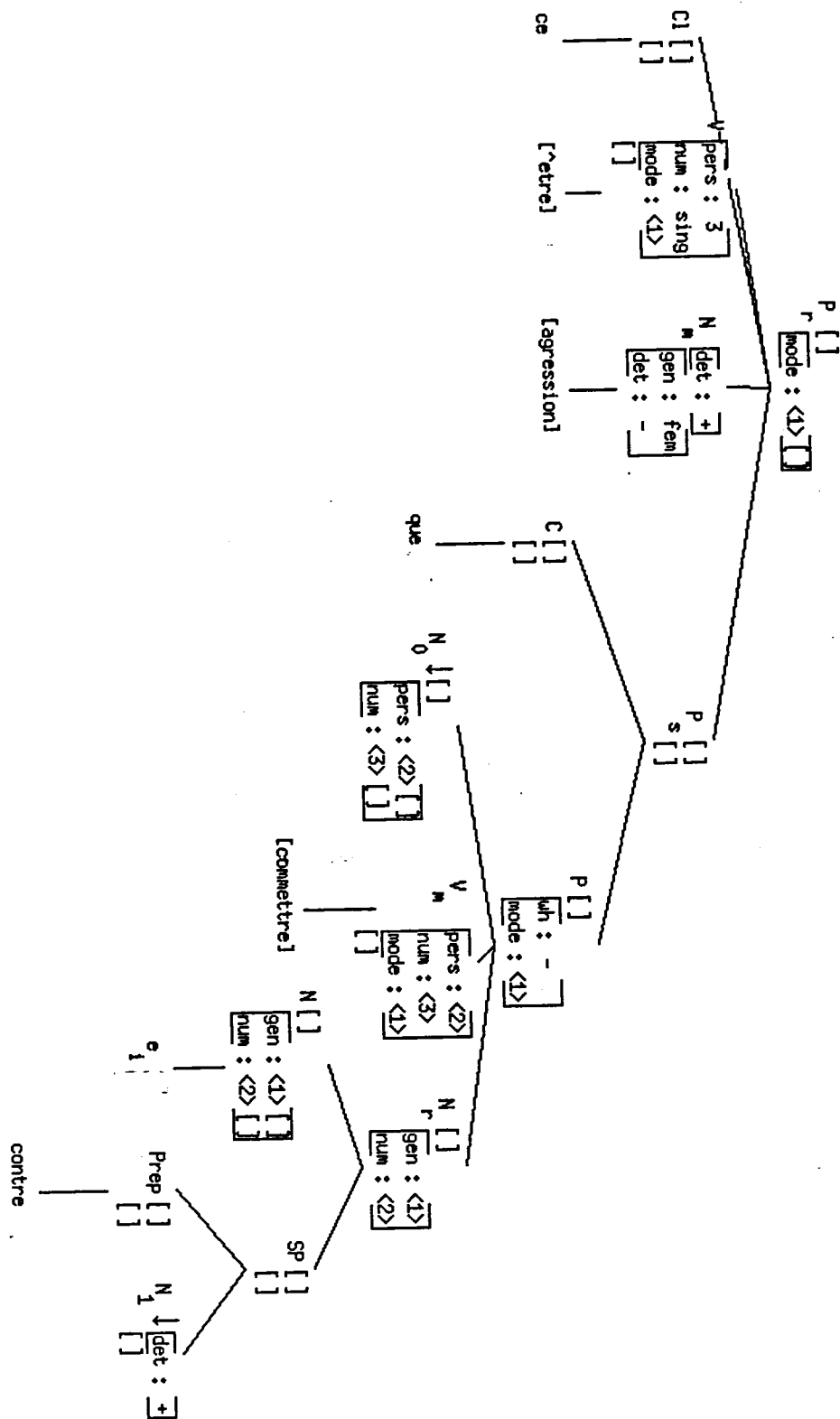
C'est Jean qui a commis une agression contre Marie

Cn0VdNpn1(1)



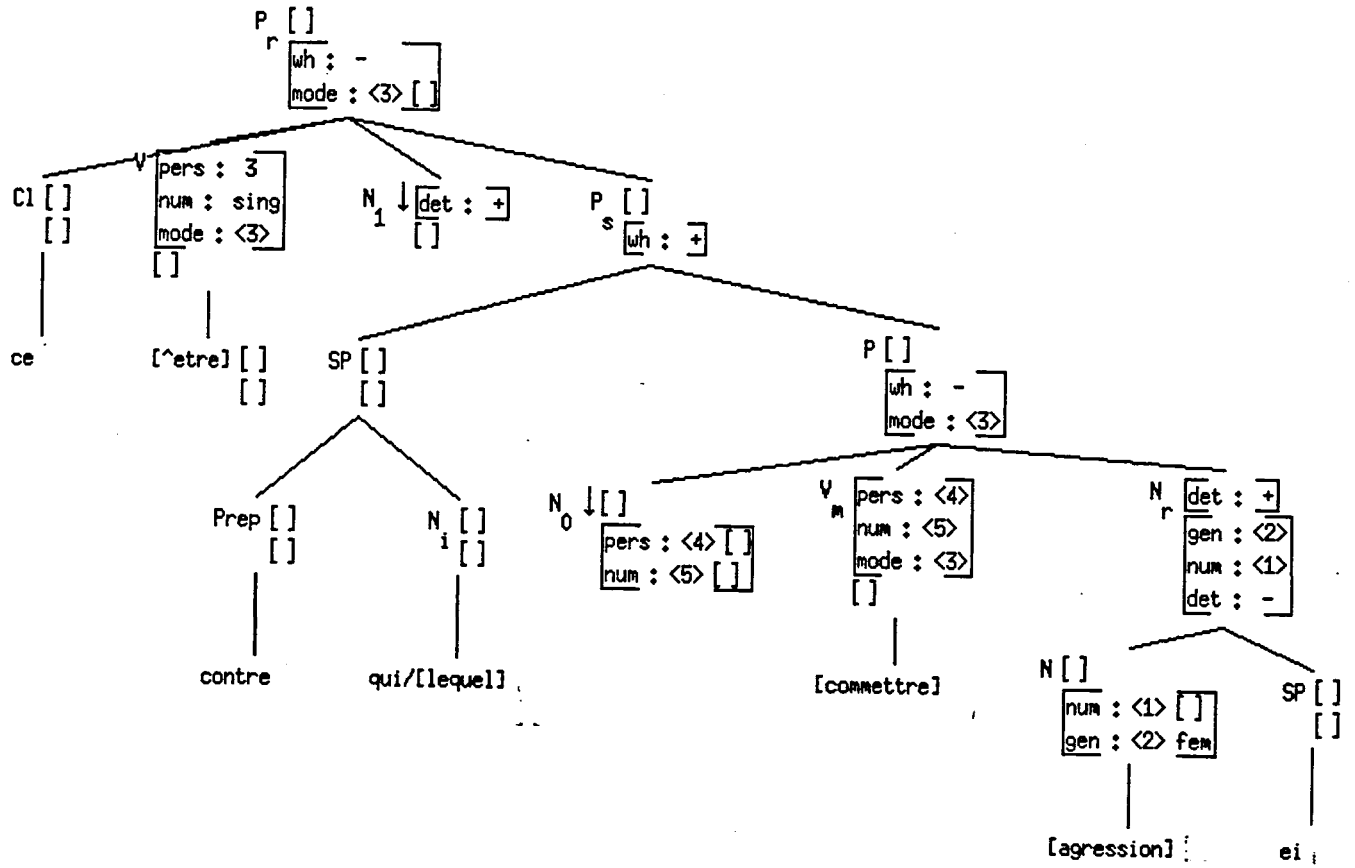
C'est une agression contre Marie que Jean a commise

Cn0VdNpn1(2)



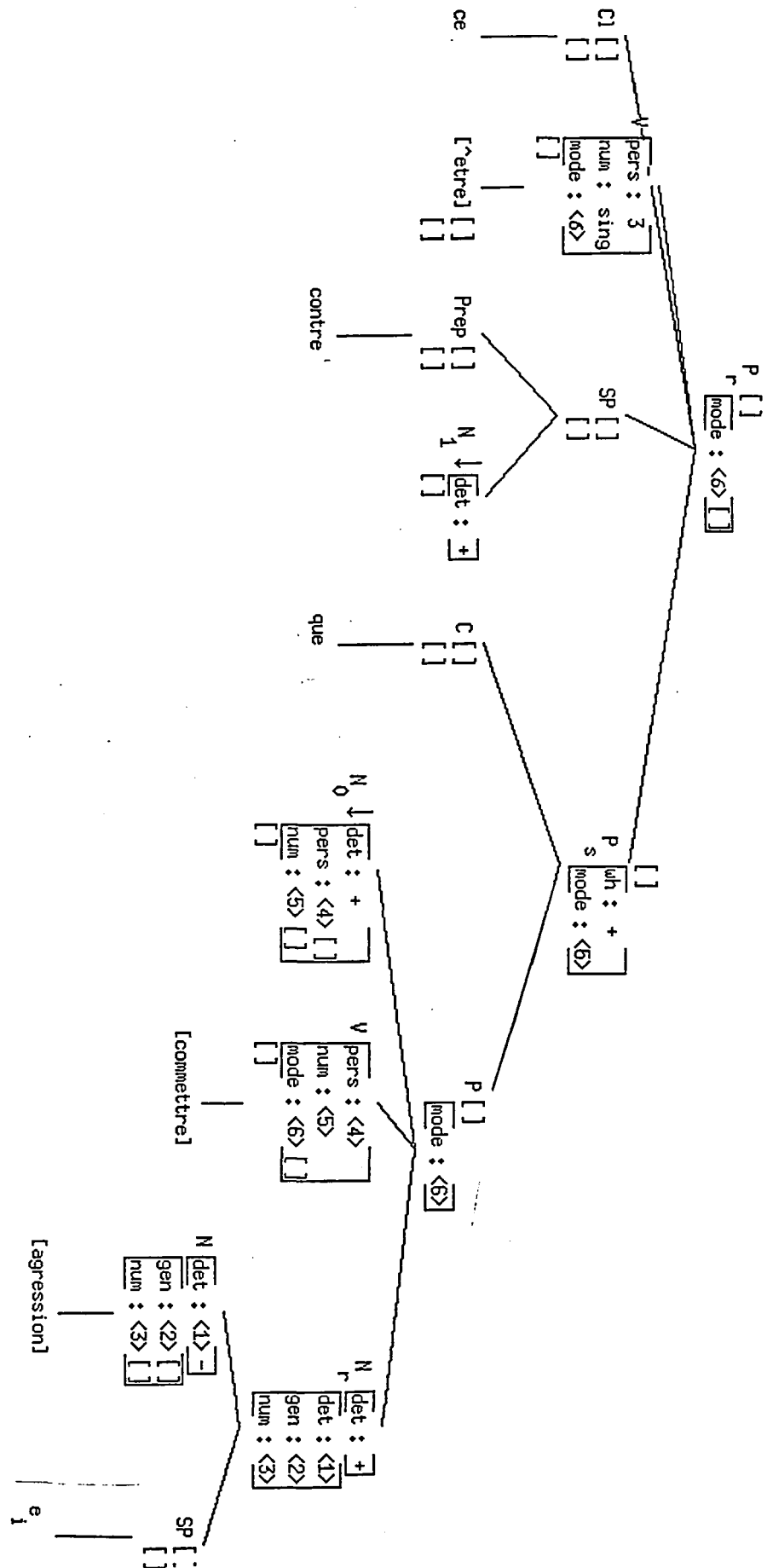
C'est une agression que Jean a commise contre Marie

C1n0VdNpn1



C'est la fille contre laquelle Jean a commis une agression

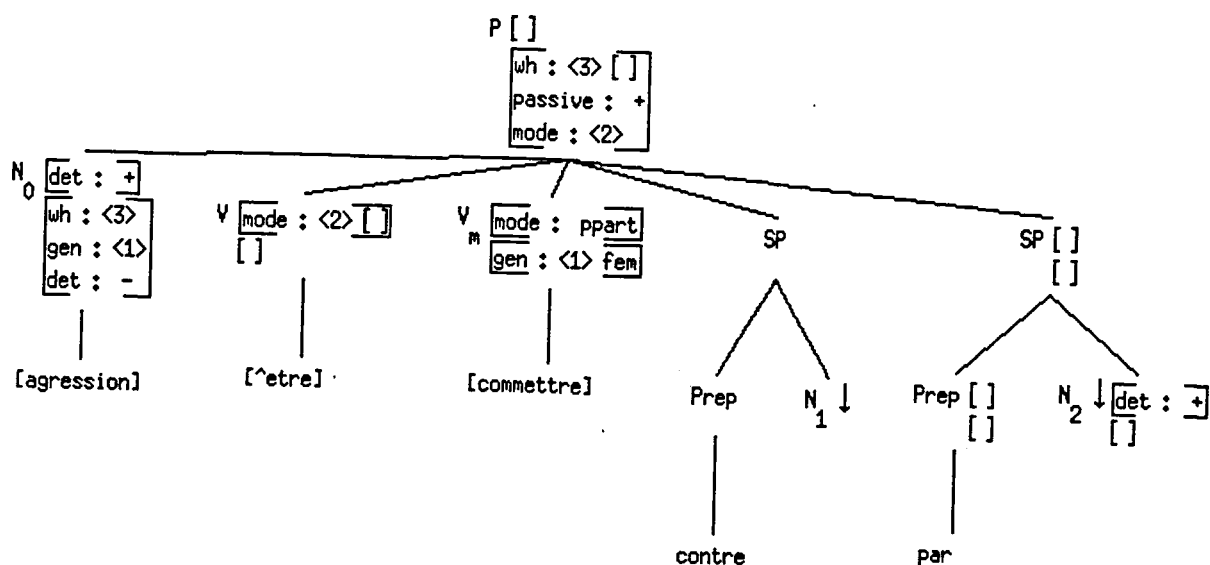
Cp1n0VdNpn1



C'est contre Marie que Jean a commis une agression

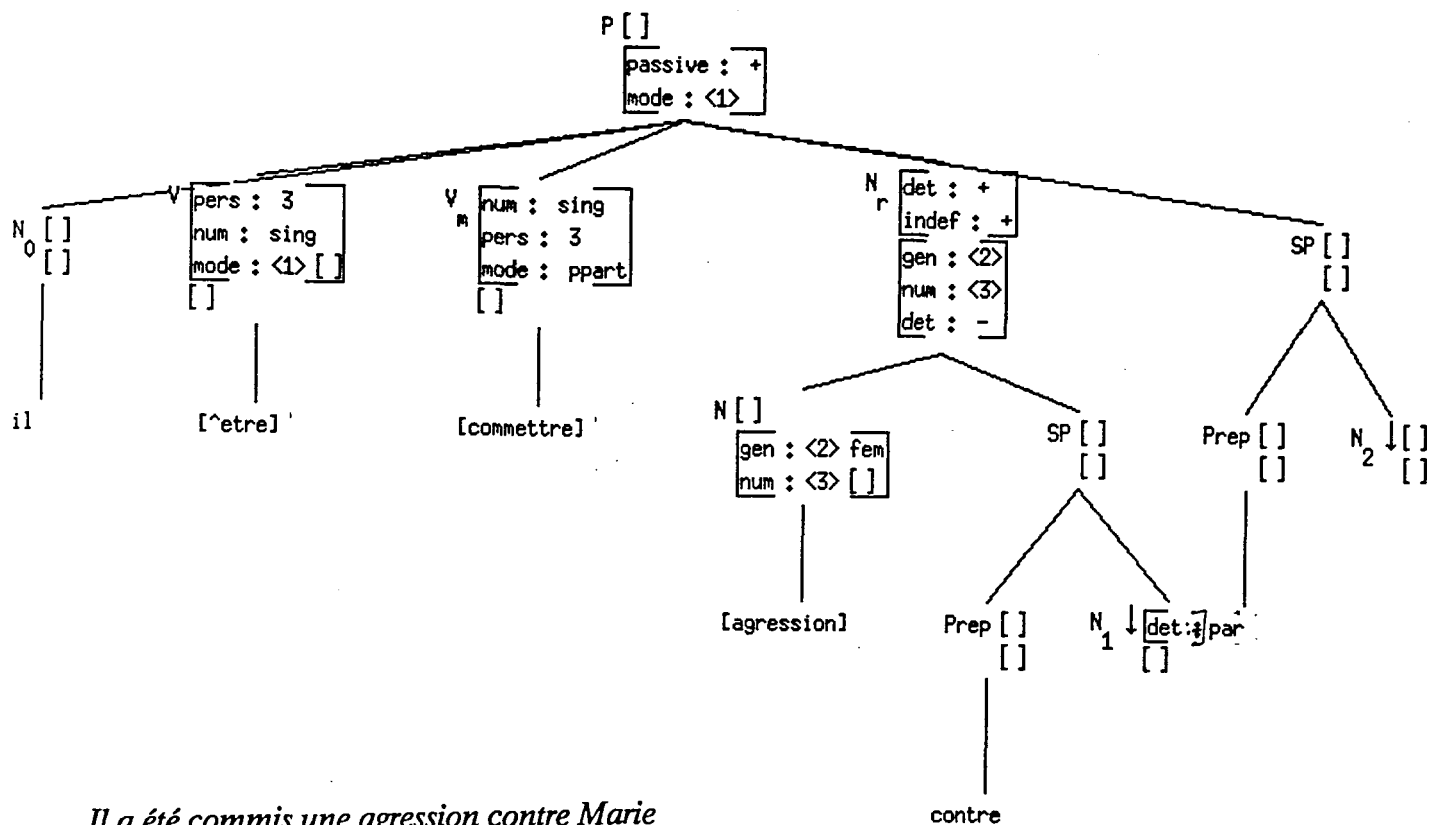
Passif

n0VdNpn1(pass)

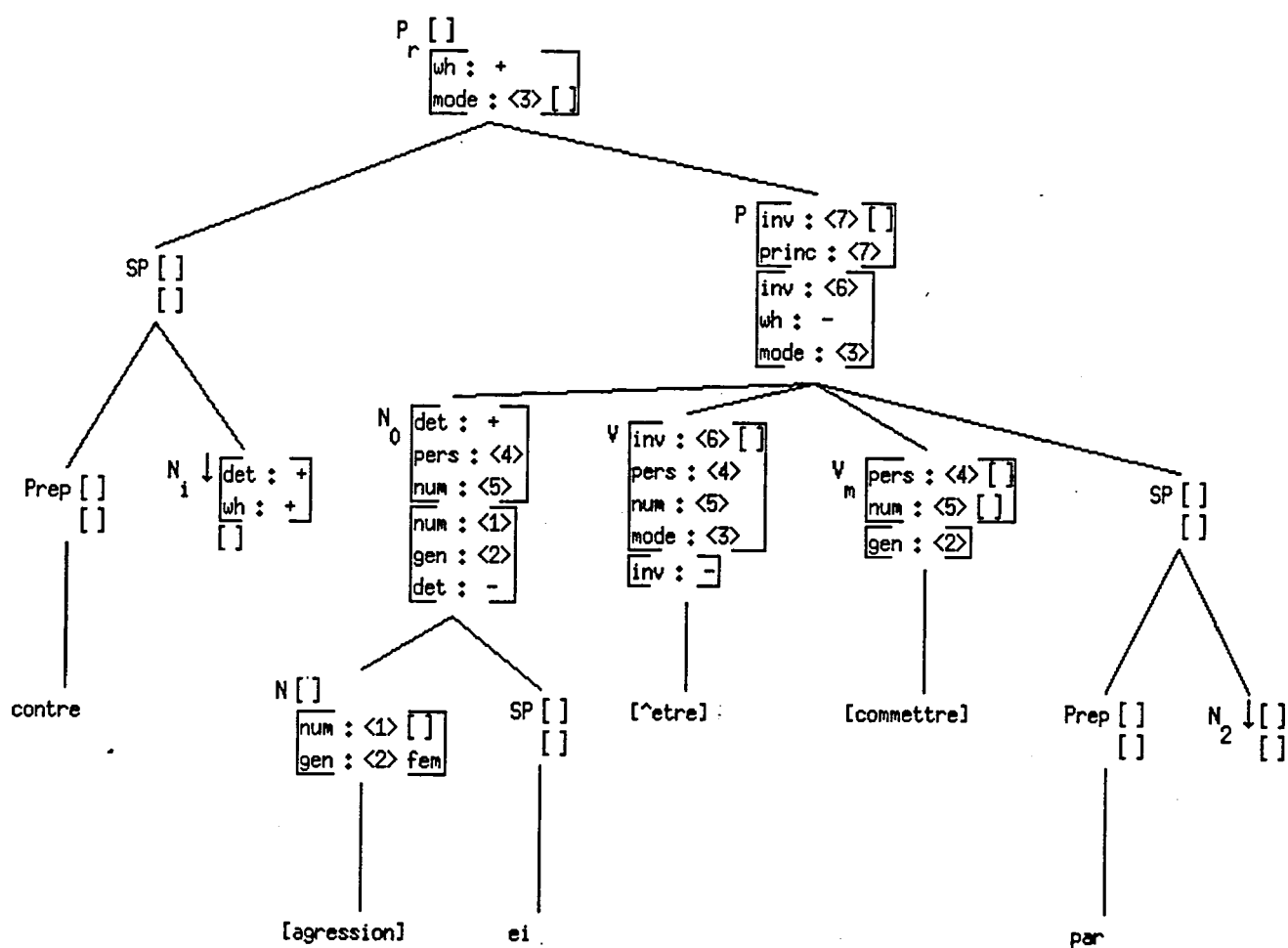
*Une agression contre Marie a été commise par Jean*

Extraposition + passif

n0VdNpn1(extr)

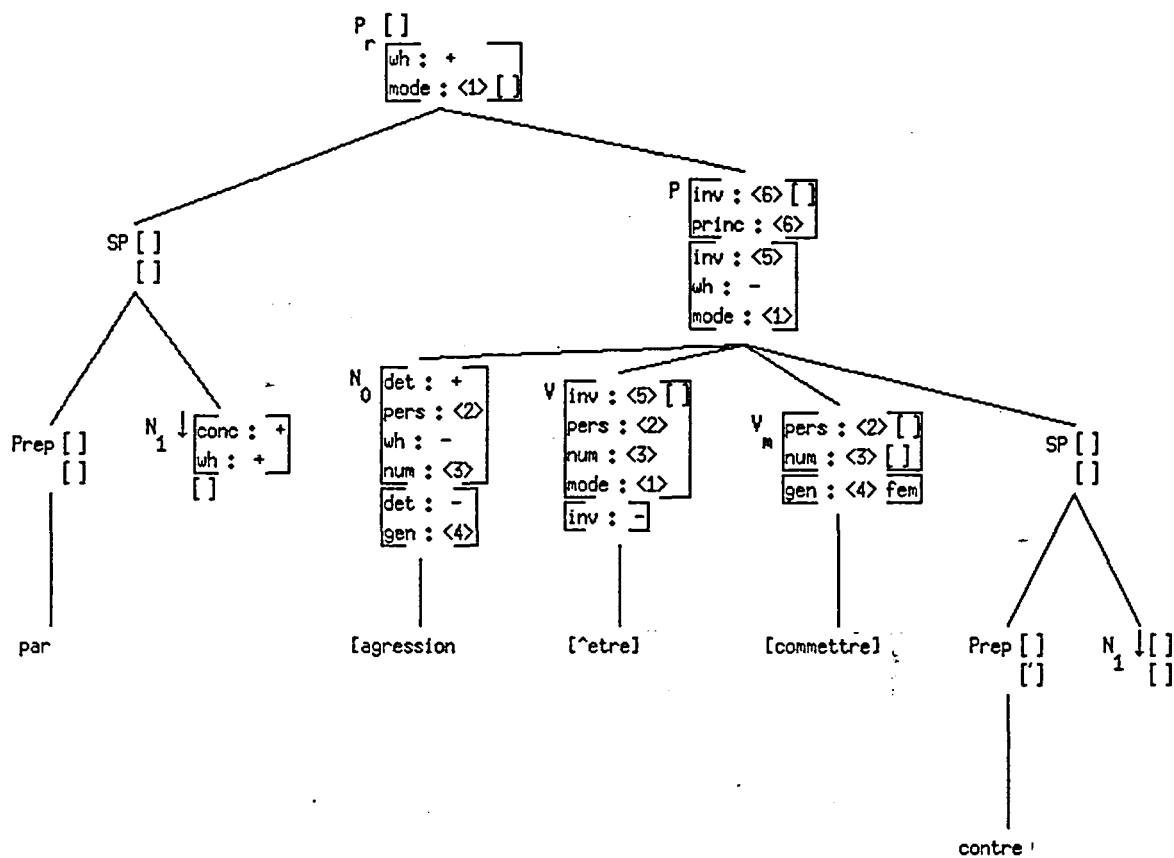
*Il a été commis une agression contre Marie*

Questions + passif
W0n0VdNpn1(pass)

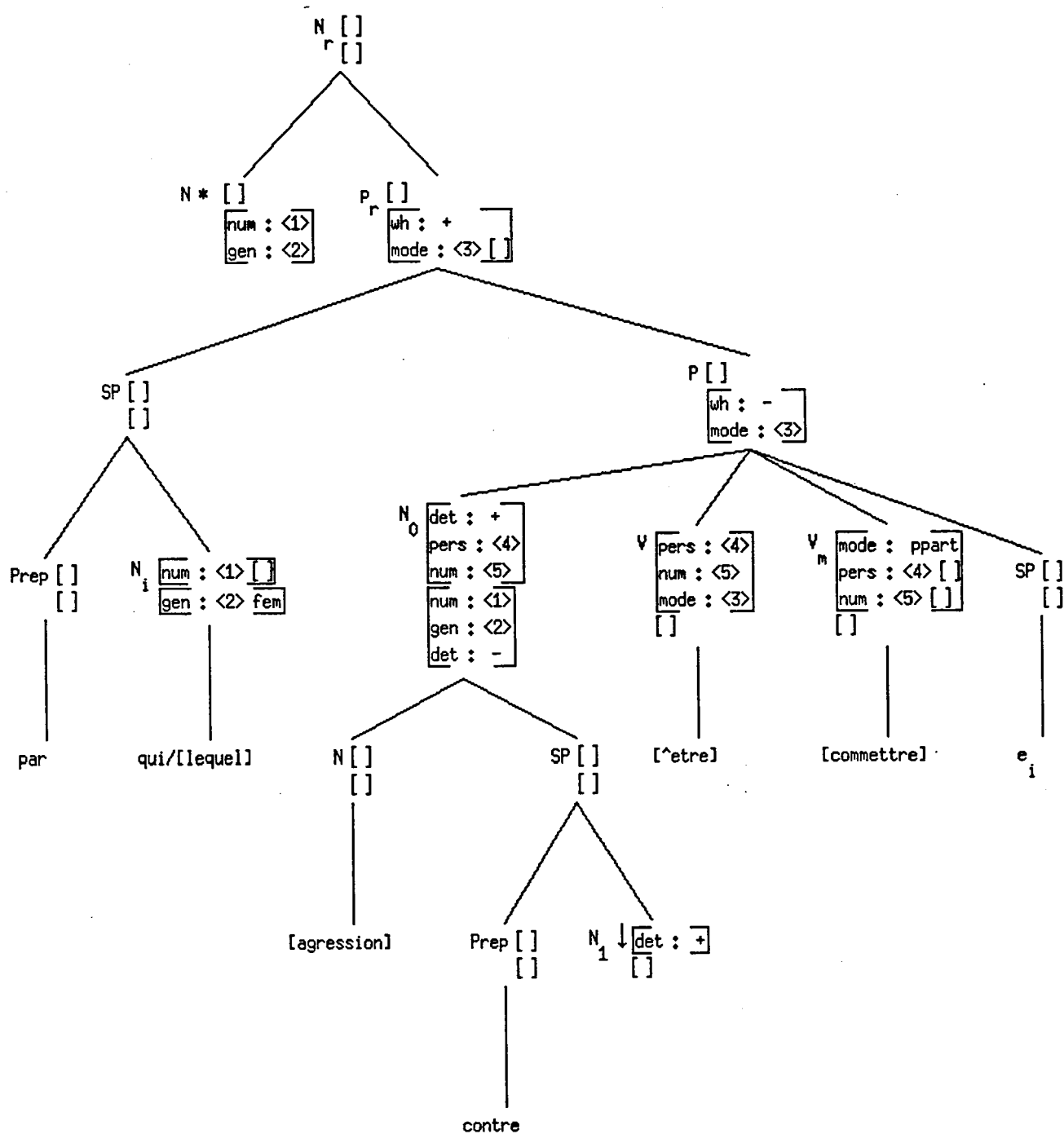


Contre qui une agression a-t-elle été commise par Jean ?

W1n0VdNpn1(pass1)

Par qui une agression a -t-elle été commise contre Marie ?

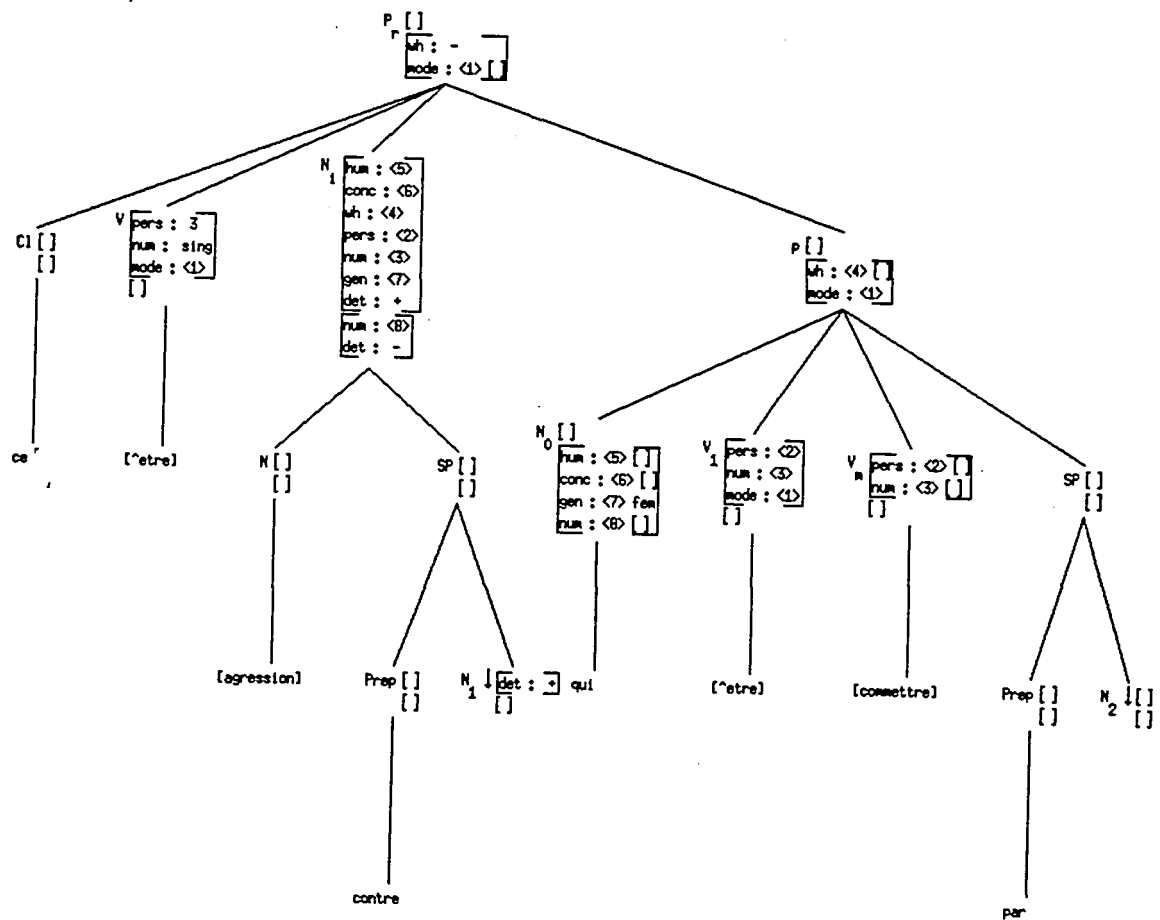
R1n0VdNpn1(pass2)

Je connais l'homme par qui une agression contre Marie a été commise

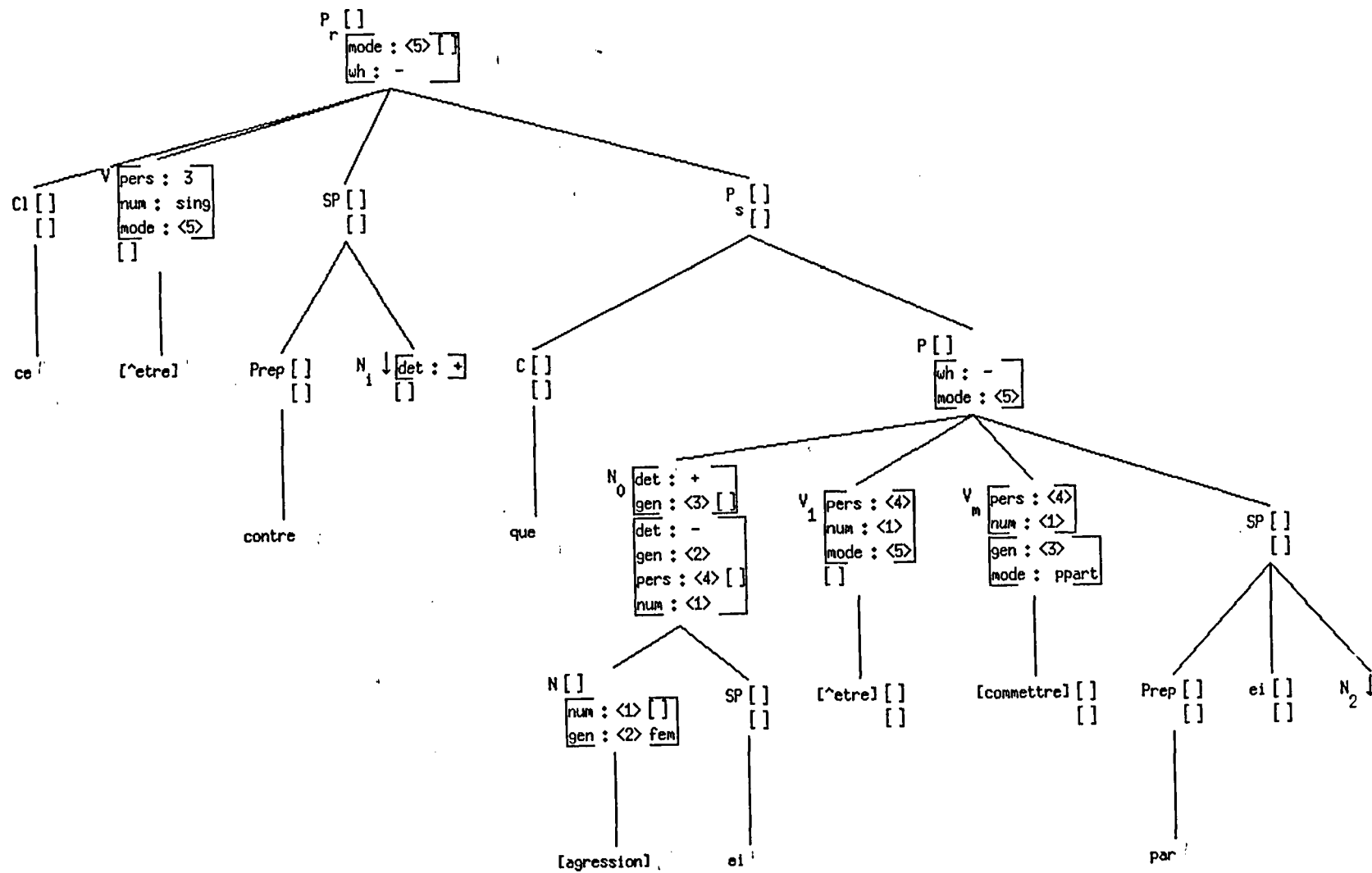
Clivée+passif

Cn0VdNpn1(pass1)

C'est une agression contre Marie qui a été commise par Jean

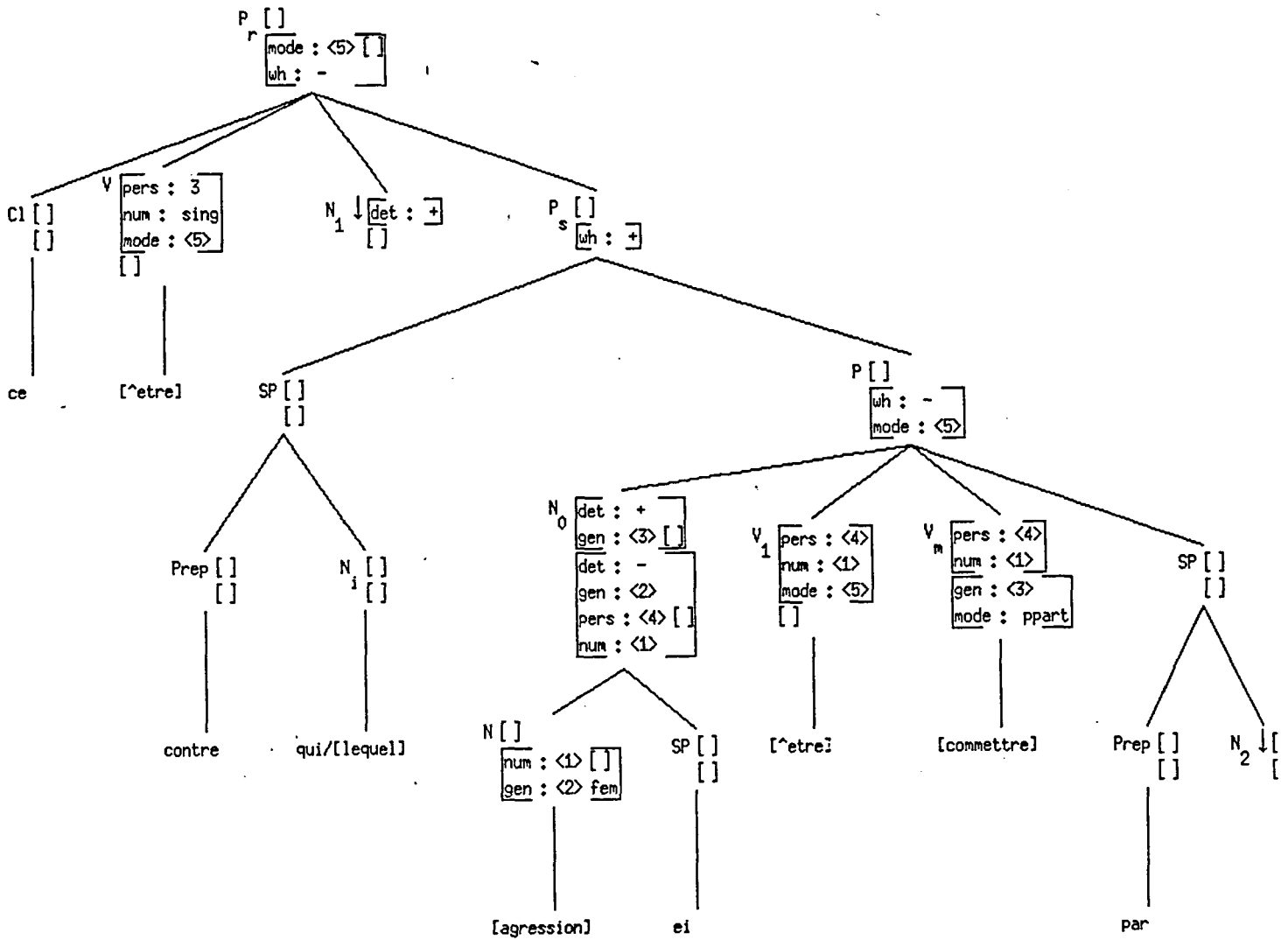


Cp0n0VdNpn1(pass)

C'est contre Marie qu'une agression a été commise par Jean

C0n0VdNpn1(pass)

C'est la fille contre laquelle une agression a été commise par Jean



FAMILLE N0VDN1

entrée lexicale :

[veste],N;[prendre],V;une,D: n0VdN(pn1){N0:<hum>=+,N1:<abs>=+,extrap=-,neutre=-}.

Arbres élémentaires

Traits

Exemples de phrases

n0Vn1

Jean prend une veste

R0n0Vn1

relative

Je connais la personne qui a pris une veste hier

R1n0Vn1

relative

J'ai été surprise par la veste que Jean a prise hier

C0n0Vn1

clivée

C'est Jean qui a pris une veste hier

C1n0Vn1

clivée

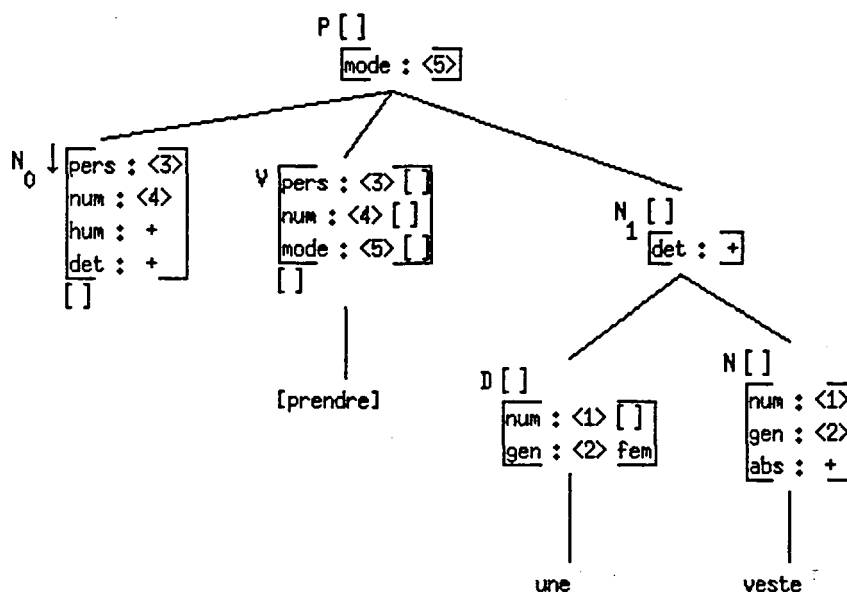
C'est une sacrée veste que Jean a prise hier

ARBRES ELEMENTAIRES DE LA FAMILLE N0VDN1

entrée lexicale :

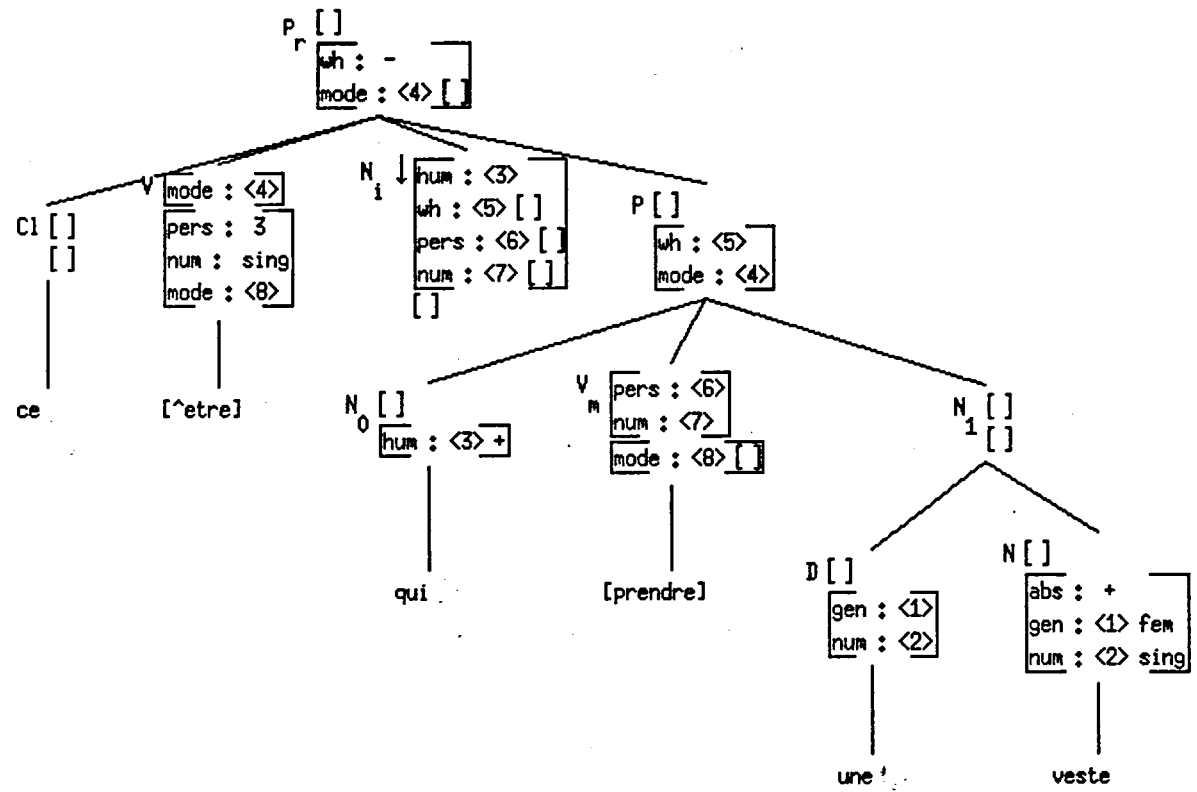
[veste],N;[prendre],V;une,D: n0VdN(pn1){N0:<hum>=+,N1:<abs>=+,N1.t:<qu>=-,extrap=-,neutre=-}.

n0Vn1

*Jean prend une veste*

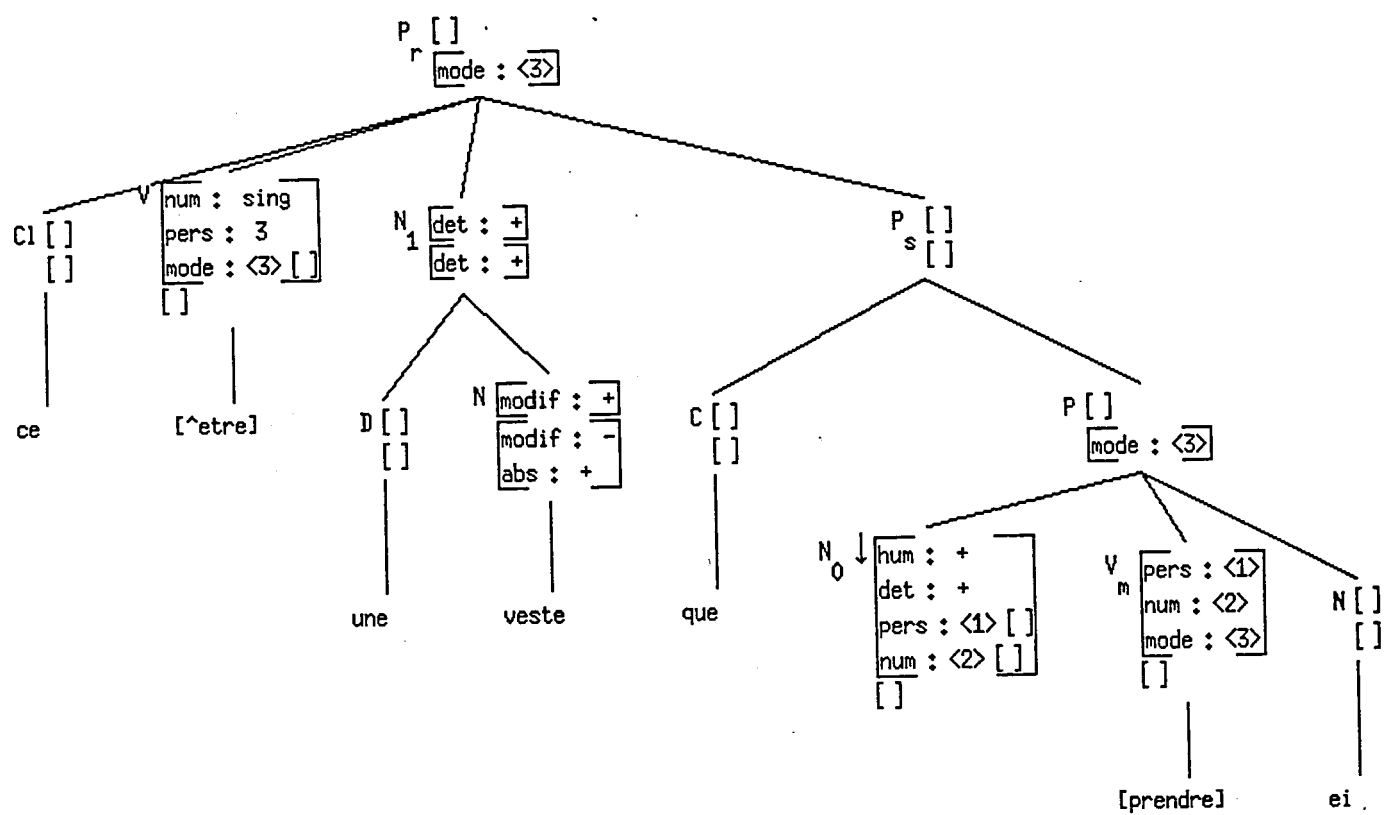
Clivées

C0n0Vn1



C'est Jean qui a pris une veste hier

C1n0Vn1

C'est une sacrée veste que Jean a pris hier

ANNEXE 3**EXTRAITS DES DICTIONNAIRES**

```

;;; -*- Package: LEXICON -*-
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%% GRAMMAIRE TAG DU FRANCAIS
%%% DICTIONNAIRE SYNTAXIQUE
%%%
%%%          900 VERBES = 1500 ENTREES
%%%
%%% Anne Abeille
%%% (1ere version :juillet 1988 avec l'aide de Beatrice Daille et Sabine
%%% Petillon)
%%% mis a jour AA decembre 1989 et aout 1990)
%%%
%%% Traits semantiques sur Ni : <abs>, <conc>, <hum>
%%% Traits mode sur Si ou Vi: <mode>=inf/subj/ind/ppart
%%% Traits d'auxiliaires : <aux>= ^etre/avoir
%%% par default, Ni:<det>=+, <aux>=avoir, <dative>=+, erg=-, extrap=-, sym=-
%%%
%%% autres traits utilises: <quant>, <dative>, <dative-h>, <acc>
%%% a rajouter: information sur les controleurs
%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

```

[abaisser],V : n0Vn1,

[abaisser],V; (a,Prep): s0Vn1, n0Vn1ps2{N1:<hum>=+}.

[abandonner],V : n0Vn1, n0V.

[abattre],V : n0V(n1), s0Vn1.

[abimer],V : n0V(n1), s0Vn1.

[abjurer],V: n0V(n1){N0:<hum>=+}.

[aboutir],V;a,Prep1;pour,Prep2 : n0V, s0Vpn1(pn2), n0Vpn1(pn2).

[abreuver],V;de,Prep: n0Vn1(pn2){N2:<num>=pl,N1:<hum>=+}.

[abriter],V; [se],Cl;(de,Prep) : n0ClV(pn1)(pn2){V.b:<refl>=+, p2:<loc>=+}.

[abuser],V;de,Prep : n0V(pn1){N0:<hum>=+}. %to overindulge %= abus

[abuser],V: n0Vn1, s0Vn1{N1:<hum>=+}. %to abuse

[accabler],V;(de,Prep) : s0Vn1, n0Vn1(pn2){N1:<hum>=+}.

[accepter],V ;(de,Prep);(de/que,C) : n0V(n1)(pn2){N2:<hum>=+, N0:<hum>=+},
n0Vs1{S1:<mode>=subj/inf}.

[accompagner],V : n0Vn1(pn2){N0:<hum>=+,p:<loc>=+}.

[accompagner],V: n0Vn1{N0:<hum>=-}. %accompagnement

[accomplir],V : n0Vn1.

[accorder],V : n0Vn1. %to tune (for music)

```

[accorder],V;a,Prep;(de/que,C) : n0Vn1(pn2)(N2:<hum>=+,N0:<hum>=+),
n0Vs1(pn2)(S1:<mode>=inf/subj).

[accorder],V; a/avec,Prep : n0Vn1pn2(sym=1).% to get well together
%=accord

[accoucher],V;de,Prep : n0V(pn1)(N0:<gen>=fem), n0Vn1(pn2)(N1:<gen>=fem,
N0:<hum>=+).

[accuser],V;(de,Prep): n0Vn1(pn2)(N1:<hum>=+), n0Vn1ps2(S2:<mode>=inf),
s0Vn1. %=accusation

[acheter],V ;(a/pour,Prep) : n0V(n1), n0Vn1(pn2)(N0:<hum>=+,N2:<hum>=+).
%= achat

[achever],V;(de,C) : n0Vn1(N1:<abs>=+), n0Vs1(S1:<mode>=inf). %to end up
%= achievement

[achever],V: n0Vn1(N1:<hum>=+), s0Vn1(N1:<hum>=+). %epuiser

[acqu'erer],V;(de,Prep) : n0Vn1(pn2). %=acquisition

[acquitter],V; [se],Cl;(de,P) : n0ClV(pn1)(V.b:<refl>=+,N0:<hum>=+).

[adapter],V; (a/avec,Prep): n0Vn1(pn2)(sym=1).

[admettre],V;(que,C) : n0Vn1(N1:<abs>=+,N0:<hum>=+),
n0Vs1(N0:<hum>=+,S1:<mode>=subj/ind).

[admettre],V : n0Vn1(pn2)(p:<loc>=+,N1:<hum>=+). %=admission

[admirer],V ;(que,C);(de,Prep): n0Vs1(N0:<hum>=+, S1:<mode>=subj), n0Vn1ps2
(N1:<hum>=+, S2:<mode>=inf).

[admirer],V;(pour,Prep): n0Vn1(pn2)(N2:<abs>=+,N1:<hum>=+).

[adopter],V : n0Vn1(N1:<hum>=+,N0:<hum>=+). %=adoption

[adopter],V ;(que,C): n0Vn1(N1:<abs>=+), n0Vs1(S1:<mode>=ind). %=adoption

[adresser],V;(a,Prep) : n0Vn1(pn2).

[adresser],V; [se],Cl;a,Prep: n0ClVpn1(V.b:<refl>=+, V.t:<ative>=-,
N0:<hum>=+).

[affaiblir],V : n0Vn1, s0Vn1.

[affaiblir],V; [se],Cl : n0ClV(V.b:<refl>=+).

[affecter],V: s0Vn1(N1:<hum>=+).

[affecter],V;(de,C): n0Vn1(N1:<abs>=+,N0:<hum>=+), n0Vs1(S1:<mode>=inf).

[affecter],V; a,P: n0Vn1pn2. %=affectation

[affirmer],V;a,P;(que,C): n0Vn1(pn2)(N0:<hum>=+), n0Vs1pn2
(S1:<mode>=inf,S1:<tense>=past),

```

n0s1(pn2){S1:<mode>:ind/subj}.

[agacer],V : n0Vn1, s0Vn1{N1:<hum>=+}.

[aggraver],V : n0Vn1, s0Vn1{N1:<hum>=-}.

[agir],V; [se],Cl; il,N0;(de,Prep);(que,C) : N0ClVpn1{V.b:<refl>=+},
N0ClVps1{S1:<mode>=inf,V.b:<refl>=+, N0ClVs1{s1:<mode>:subj}.

[agir],V;(sur,P) : n0V, n0Vpn1.

[agir],V: n0Vad(Ad:<man>=+,N0:<hum>=+).

[agrandir],V : n0Vn1.

[agresser],V : n0Vn1. %=agression

[aider],V;(a,Prep) : n0Vn1(ps2){S2:<mode>=inf}, s0Vn1.

[aimer],V ;(que,C);(a,Prep) : n0Vn1{N0:<hum>=+}, n0Vs1{S1:<mode>=subj,
V.t:<acc-le>=-}, n0V(p)s1{V.t:<dative>=-, S1:<mode>=inf}.

[ajouter],V;a,Prep;(que,C) : n0Vn1(pn2){sym=1},
n0Vs1(pn2){S1:<mode>=ind/subj}, s0Vpn1.

[alarmer],V : n0Vn1, s0Vn1{N1:<hum>=+}. %=alarme

[aligner],V;(a/avec,sur,Prep) : n0Vn1(pn2){sym=1}. % symetric

[aller],V : n0Vad{V.b:<aux>=^etre,Ad:<man>=+}. %to do fine/terrible...

[aller],V ;avec,Prep : n0Vpn1{V:<aux>=^etre}. % to suit

[aller],V{V.b:<mode>=ind/inf/subj} ;a,P : n0Vpn1{N1:<hum>=+}. % to fit

[aller],V : n0Vs1{V:<aux>=^etre,S1:<mode>=inf}, n0Vpn1{p:<loc>=+}. % to go

[aller],V;[se],Cl;en,Cl1: n0ClCl1V(pn1){V.b:<refl>=+,p:<loc>=+}.

[aller],V;chez,Prep : n0Vpn1{V:<aux>=^etre,N1:<hum>=+,N0:<hum>=+}. % to
visit

[allier],V;a/avec,Prep: n0Vn1pn2(sym=1). % symetric verb

[allumer],V : n0V(n1).

[alterner],V;(avec,Prep) : n0V(n1)(pn2){sym=0}. % symetric verb
%=alternance

[amener],V : n0Vn1(pn2){p:<loc>=+}.

[amener],V;a,Prep : n0Vn1ps2{ S2:<mode>=inf,N1:<hum>=+}.

[amuser],V : n0Vn1, s0Vn1{N1:<hum>=+}.

[amuser],V; [se],Cl;(a,Prep1),(avec,Prep2) :
n0ClV(pn1)(pn2){V.b:<refl>=+,N1:<abs>=+,N0:<hum>=+,sym=0},
n0ClVps1{V.b:<refl>=+,S1:<mode>=inf}.

```


;;; Package: LEXICON; Base: 10 -*-

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXX GRAMMAIRE TAG DU FRANCAIS

XXX DICTIONNAIRE SYNTAXIQUE

XXX

XXX 800 Noms = 1000 entrées

XXX

XXX Anne Abeill'e

XXX (1e version D'ecembre 1988 (mise a jour: Mars 1990,

XXX aout 1990)

XXX

XXX (avec les constructions 'a verbe support associ'ees)

XXX SN complexe toujours possible

XXX causatifs mettre et donner, pas de variantes aspectuelles

XXX pas de noms composés

XXX

XXX traits utilis'es: genr, num, hum, abs, conc

XXX count=+/-;mass=+/-; modif=+/-; <part>=+/-, <def>=+/-

XXX avec les conventions : hum=+ -> mass=-;abs=+ -> count=-; <gen>=+ =>

<def>=+, <num>=sing

XXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

[abandon],N(N:<num>=sing);[^etre],V;a,Prep;[le],D: n0VPDN1(N0:<loc>=+).

[abandon],N;(de,Prep1); (par,Prep2): NPdn(pn1)(pn2)(N:<abs>=+). %abandonner

[abb'e],N: NPdn(N:<hum>=+).

[abeille],N : NPdn(N:<mass>=-).

[abime],N : NPdn().

[abondance],N;de,Prep; [avoir],V : n0VNpn1(N1:<abs>=+).

[abondance],N(N:<num>=sing);[^etre],V;en,Prep:
n0VPN1(pn2)(p2:<loc>=+,N0:<num>=pl/<def>=+). %abonder

[abord],N(N:<num>=pl);(de,Prep): NPdn(pn1)(N1:<loc>=+).

[abord],N(N:<num>=sing);(de,Prep): NPdn(pn1)(N1:<hum>=+).

[aboutissement],N;de,Prep;[^etre],V : n0VdNpn1(N1:<abs>=+, N0:<abs>=+).

[abri],N(N:<num>=sing);[^etre],V;a,Prep;[le],D;de,Prep1: n0VPDN(pn1)().

[abri],N;contre,Prep: NPdn(pn1)(N:<hum>=-).

[absence],N;de,Prep: NPdn(pn1)(N1:<num>=sing). %N1:common noun

[absence],N;de,Prep: NPdn(pn1)(pn2)(N2:<hum>=+).

[absence],N;[avoir],V: n0VdN1(N0:<hum>=+). %to be absentminded

[accident],N;[avoir],V: n0VdN1(N1:<mass>=-).

[accord],N(N:<num>=sing);[son],D;[donner],V;(a,Prep1);(pour,Prep2):
 n0VDN(pn1)(pn2)(N1:<hum>=+,N0:<hum>=+),
 n0VDN(pn1)(ps2)(N1:<hum>=+,N0:<hum>=+,S2:<mode>=inf),
 NPdn(pn1)(N.t:<poss>=+,N:<mass>=-).

[accord],N;avec,Prep;[conclure],V: n0VdNpn1(N1:<hum>=+).

[accusation],N;[porter]/[lancer],V;contre,Prep: n0VdNpn1(N0:<hum>=+,
 N1:<hum>=+).

[achat],N;[faire],V;de,Prep: n0VdNpn1(N0:<hum>=+).

[achat],N(N:<num>=pl);[faire],V;[un],D: n0VDN.

[ach'evement],N(N:<num>=sing);[^etre],V;[le],D;de,Prep: n0VDNpn1(N1:<hum>=-
)

[acquisition],N;[faire],V;([le],D;de,Prep): n0VdN(N0:<hum>=+),
 n0VDNpn1(N0:<hum>=+,N1:<conc>=+,N:<num>=sing).

[acte],N(N:<num>=sing);[faire],V;de,Prep: n0VNpn1(N1:<abs>=+,N1:<det>=-),
 n0VNps1(N0:<hum>=+,S1:<mode>=inf).

[actif],N : NPdn.

[actif],N(N:<num>=sing);[^etre],V;a,Prep;[le]/[son],D;(de,Prep1):
 n0VPDNpn1(N.t:<poss>=+, N1:<hum>=+, s0VPDNpn1(N1:<hum>=+).

[action],N : NPdn. %en bourse

[action],N(N:<num>=sing);[lancer],V;contre,Prep: n0VdNpn1(N0:<hum>=+).

[action],N(N:<num>=sing);[avoir],V;sur,Prep;[un],D: n0VDNpn1. %agir sur

[action],N(N:<num>=sing);[exercer],V;sur,Prep;[un]/[son],D:
 n0VN1pn1(N0:<hum>=+).

[activit'e],N;[avoir],V;(de,Prep):n0VdN(pn1)(N0:<hum>=+,N1:<abs>=+).

[adieu],N(N:<num>=sing);[dire],V;a,Prep: n0VN1(pn2)(N0:<hum>=+).

[adieu],N(N.b:<num>=pl);[faire],V;[son]/[un],D;a,Prep:
 n0VDN1pn2(N0:<hum>=+,N1:<hum>=+, D:<poss>=- N.t:<modif>).

[administration],N(N:<num>=sing);de,Prep1;par,Prep2:
 NPdnpn1(pn2)(N2:<hum>=+).

[admiration],N(N:<num>=sing);[avoir],V;[du]/[un],D;pour,Prep:
 n0VDNpn1(N0:<hum>=+, D:<part>=-N.t:<modif>). %= admirer

[admiration],N(N:<num>=sing);[avoir],V;[le],D;de,Prep:
 n0VDNpn1(N1:<hum>=+). %=etre admiré par

[admiration],N(N:<num>=sing);[faire]/[susciter]/[provoquer],V;[le],D;de,Pre
 p: n0VDNpn1(N1:<hum>=+).
 %= etre admir'e par

[admiration],N(N:<num>=sing);[^etre],V;en,Prep;devant,Prep1:n0VPN(pn1)(N0:<hum>=+).
% =admirer

[adoption],N;de,Prep;[faire],V;[le],D: n0VDNpn1(N1:<hum>=+).
%= adopter Nhum

[adresse],N;(de,Prep): NPdn(pn1). %courrier

[adresse],N;[avoir],V;[du]/[un],D:
n0VDN(D:<part>=-N.t:<modif>,N:<abs>=+,N0:<hum>=+). %etre adroit

[adresse],N(N:<num>=sing);[^etre],V;a,Prep;[le],D;de,Prep1:
n0VPDNpn1(N1:<hum>=+). %etre adressé à

[affaire],N;[avoir],V;avec/a,Prep: n0VNpn1(N0:<hum>=+,N1:<hum>=+).

[affaire],N(N:<num>=pl); de,Prep: NPdn(pn1)(N1:<hum>=+).
% les affaires de Jean

[affectation],N;[avoir]/[recevoir],V: n0VdN(pn1)(N0:<hum>=+,Prep:<loc>=+).
% etre affecte loc
% etre affecte a un travail

[affection], N(N:<num>=sing);[avoir],V;[du],D;pour,Prep:
n0VDNpn1(N0:<hum>=+). %n0VPN1pn2(V=donne/accorde/apporte,Prep2=a).

[affirmation],N : n0VN1pn2(V=fait, p2=de).

[agacement],N(N:<num>=sing);[avoir],V;[du]/[un],D:
n0VDN(N0:<hum>=+,D:<part>=-N.t:<modif>,caus=donner).

[agent],N: NPdn(N:<hum>=+).

[agent],N: NPdn(N:<hum>=-,N.t:<modif>=+). %+ adj

[agrandissement],N;[faire],V;de,Prep : n0VdN(pn1)(N1:<conc>=+).

[agression],N;(contre,Prep);[faire]/[commettre],V :
n0VdN(pn1)(N:<count>=+,N0:<hum>=+, N2:<hum>=+).

[aide],N(N:<num>=sing);[^etre],V;[un],D;pour,Prep: n0VDNpn1(N1:<hum>=+).

[aide],N(N:<num>=sing); [donner],V;[du]/[un],D;a,Prep: n0VDNpn1
(D:<part>=-N.t:<modif>).

[aile],N : NPdn.

[aile], N(N:<num>=pl);[avoir],V;[du],D: n0VDN1(N0:<hum>=+, caus=donner).
% N0 a des ailes; X donne des ailes à N0

[aile],N(N:<num>=sing);sous,Prep;[^etre],V;[le],D;de,Prep1:
n0VPN1pn2(N0:<hum>=+,N2:<hum>=+).

[aimant],N: NPdn(N:<conc>=+).

[air],N(N:<num>=sing);[avoir],V;[le],D;de,Prep : VDNpv1(V1:<mode>:inf), %
il a l'air de pleuvoir =sembler

```

;;; Package: LEXICON; Base: 10 -*-
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% GRAMMAIRE TAG DU FRANCAIS %%
%%
%% DICTIONNAIRE SYNTAXIQUE
%%
%% 300 Adjectifs = 400 entrées
%%
%% epithetes et attributs, avec phrases copulatives
%% adjectifs a mont'ee
%%
%% Traits utilisées: <conc>, <hum>, <abs>, <num>, <det>
%% Pas d'adjectifs composés
%% Pas de "tough" constructions
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

[absolu], A; ([^etre],V) : BNa(N:<abs>=+), n0Va(A:<abs>=+).

[absurde],A; ([^etre],V) ;(de,C);(de la part de ,Prep):
n0Va1(N0:<hum>=+,S1:<mode>=inf), BNa(N*:<abs>=+) s0Va(pn2)(N2:<hum>=+).

[actuel], A; ([^etre],V) : n0Va, BNa, BaN.

[admirable], A; ([^etre],V) : AA, BaN(N*:<abs>=+), BNa.

[affreux],A; ([^etre],V; pour,Prep) : BNa, BaN, s0Va(pn1)(N1:<hum>=+).

[agréable], A; ([^etre],V;a/pour,Prep) : BNa, BaN, n0Va,
s0Va(pn1)(N1:<hum>=+).

[agressif],A;([^etre],V;envers,Prep): BNa(pn1), n0Va(pn1).

[aimable], A; ([^etre],V;envers/avec,Prep) : BaN, BNa(pn1),
n0Va(pn1)(N0:<hum>=+).

[allemand], A; ([^etre],V) : BNa, n0Va.

[américain], A; ([^etre],V) : n0Va, BNa.

[amoureux], A; ([^etre],V;de,Prep) : n0Va(pn1)(N0:<hum>=+),
BNa(pn1)(N*:<hum>=+).

[ancien], A : BaN(N*:<hum>=+). % un ancien maire

[ancien], A; ([^etre],V) : n0Va(N0:<hum>=-), BNa(N*:<hum>=-),
BaN(N*:<hum>=-).

[anglais], A; ([^etre],V) : n0Va, BNa.

[autre], A : BaN.

[autre], A; ([^etre],V) : BNa(N*:<abs>=+), n0Va(N0:<abs>=+).

[aveugle], A; ([^etre],V);('a,Prep): n0Va(pn1)(N0:<hum>=+,N1:<abs>=+), BNa.

[bas], A; ([^etre],V) : n0Va(N0:<hum>=-), BNa(N:<conc>=+). %peu 'elev'e

[bas], A; [^etre],V : n0Va(N0:<hum>=+). %ce malade est bien bas

```

[bas], A; ([^etre],V : n0Va(N0:<hum>=+). %ce malade est bien bas

[bas],A; (^etre,V); (de la part de,Prep): s0Va(pn2)(N2:<hum>=+),
BNa(N:<abs>=+). % vil.

[beau] ,A; (^etre,V) : n0Va, BNa, BaN.

[b^ete],A; (^etre,V);(de,C): n0Vas1(S1:<mode>=inf,N0:<hum>=+),
n0V(N0:<conc>=-), BNa(N*:<conc>=-).

[bizarre], A; ([^etre],V);(de,C) : n0Va, BNa,
n0Vas1(N0:<hum>=+,S1:<mode>=inf).

[blanc],A; (^etre,V) : BNa(N:<abs>=-), n0Va(N0:<abs>=-).

[bleu],A; (^etre,V): BNa(N*:<conc>=+), n0Va(N0:<conc>=+).

[bon], A; ([^etre],V;pour/envers/avec,Prep);(de,C) : n0Va(pn1)(N0:<hum>=+),
BNa(pn1)(N*:<hum>=+), n0Vas1(N0:<hum>=+,S1:<mode>=inf). % un homme bon

[bon], A; ([^etre],V);(pour,Prep) : s0Va(pn1)(N1:<hum>=+).

[bon], A; ([^etre],V;en,Prep) : n0Vapn1(N0:<hum>=+,N1:<det>=-,N1:<abs>=+);
BNa(pn1)(N*:<hum>=-). %il est bon en maths

[bon], A ([^etre],V) : n0Va, BaN(N*:<hum>=-). un bon spectacle, un bon
gâteau

[brave], A : n0Va, BNa. %courageux

[brave], A : BaN(N*:<hum>=+).

[bref], A; ([^etre],V) : n0Va(N0:<conc>=-), BaN(N*:<abs>=+), BNa. %brève
rencontre, orateur bref

[brillant], A ; ([^etre],V);(en,Prep) : n0Va(pn1)(N0:<hum>=+,N1:<det>=-,
N1:<abs>=+), BaN, BNa. %brillant individu, resultat brillant

[brillant], A ; ([^etre],V) : n0Va, BNa(N*:<conc>=+).
% lumiere brillante

[calme], A ; ([^etre],V) : n0Va, BNa.

[capable], A ; ([^etre],V);(de,Prep); (envers,Prep2) :
n0Vaps1(S1:<mode>=inf), BNaps1, n0Vapn1pn2(N2:<hum>=+, N0:<hum>=+,
N1:<abs>=+).

[capable], A ; ([^etre],V): n0Va(N0:<hum>=+); BNa(N*:<hum>=+).

[c^elebre], A ; ([^etre],V) : n0Va, BNa, BaN.

[cens'e],A;[^etre],V: 8vaV(V*:<mode>=inf). % raising adj

[certain], A; ([^etre],V); (de/que,C) :
n0Vaps1(N0:<hum>=+,S1:<mode>=inf/ind), BNas1(N*:<hum>=+,S1:<mode>=inf/ind).
% Jean est certain de venir

[certain], A; ([^etre],V): s0Va; BNa(N*:<abs>=+). %un resultat certain

```

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXX GRAMMAIRE TAG DU FRANCAIS
XXX
XXX DICTIONNAIRE SYNTAXIQUE
XXX
XXX Prepositions, Conjonctions et Adverbes = 400 entrees
XXX
XXX D'ecembre 1988
XXX mis a jour aout 90
XXX
XXX adverbos compos'es
XXX pas de phrases a verbe support
XXX
XXX Traits utilis'es : CQ, <modif>, <det>, <abs>, <hum>
XXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

```

'a, Prep : BNpn1(N1.t:<det>=-), BVpdn, BSps1(S1:<mode>=inf).

'a, Prep : BVpdn(N:<conc>=+,SP:<loc>=+), BSpdn(N:<conc>=+,SP:<loc>=+),
ASPPn(N:<conc>=+,SP:<loc>=+).

d',Prep;[abord],N(N:<num>=sing): BSpn(N.b:<det>=+), BVpn(N.b:<det>=+).

d'es,Prep;[abord],N(N:<num>=sing); [le],D: BSpdn(N.b:<det>=+),
BVpdn(N.b:<det>=+).

absurdement, Ad: BSad(Ad:<man>=+), αAdv(Ad:<man>=+).

[accord],N(N:<num>=sing);en,Prep;avec,Prep1: BVppnp1(N.t:<det>=-).

afin,Prep; que/de,C: BSps1(S1:<mode>=inf/subj).

agr'eablement,Ad : BVad(Ad:<man>=+).

ailleurs,Ad : BVad, αAdv(Adv:<loc>=+).

aimablement,Ad;(envers,Prep) : BVad(pn1)(N1:<hum>=+),
BSad(pn1)(N1:<hum>=+), αAdv(pn1)(N1:<hum>=+,Ad:<man>=+).

par, Prep; ailleurs,Ad: BVpad.

d',Prep; ailleurs,Ad : BVpad, BSpad.

[allure],N(N:<num>=sing); 'a,P;[tout],D: BVpdn.

[année], N(N:<num>=sing): BSdn(N.b:<det>=-, N.t:<det>=+).

apr'es,Prep; (que,C): BVpn(N:<det>=+), BNpn(N:<det>=+),
BSps1(S1:<mode>=inf/ind).

apr'es,Ad: BVad.

en,Prep;arri'ere,N: BVpn, BPPn(SP:<loc>=+).

'a,Prep;[le],D;[arriv'ee],N(N:<num>=sing): BVpdn.

assez,Ad : BVad, BadA(A:<modif>=-), BadSP.

'a,Prep;[aube],N(N:<num>=sing);[le],D: BVpdn.

aussit^ot,Ad: BVad, BSad.

autrefois,Ad: BSad, BVad.

autrefois,P;(que,C) : BSad, BVad, BSps1(S1:<mode>=ind).

'a,Prep;[le],D;[avance],N(N:<num>=sing): BVpdn(N.b:<det>=+).

par,Prep;[avance],N(N:<num>=sing): BVpdn(N.b:<det>=+).

d',Prep;[avance],N(N:<num>=sing): BVad(N.b:<det>=+).

avant,Prep; (que,C): BVpn(N:<det>=+), BNpn(N:<det>=+),
BSps1(S1:<mode>=inf/subj).

avant,Ad: BVad.

en,Prep;avant,N: BVpn, AdvSPpn(SP:<loc>=+).

avec,P: BNpn(N:<det>=+), BVpn(N:<det>=+/N:<abs>=+).

aveugl'ement,Ad: BVad, AdvAd(Ad:<man>=+).

par,Prep;[avion],N(N:<num>=sing): BVpn(N.b:<det>=+), BNpn(N.b:<det>=+).

[avis],N;sur,P;de,P1: BVppnp1.

bas, N;en,P : BVppn.

l'a-bas, Ad : BVad, BSad, AdvAdv(Adv:<loc>=+).

beaucoup,Ad : BVad(CQ=-).

b^etement,Ad: BVad, AdvAd(Ad:<man>=+).

b^etement,Ad: BSad(CQ=-).

bien ,Ad: BadA, BadSP, BVad, AdvAd(Ad:<man>=+).

bient^ot, Ad : BVad, BSad.

bizarrement,Ad: BVad, AdvAd(Ad:<man>=+).

bizarrement,Ad: BSad(CQ=-).

[bonheur],N(N:<num>=sing);par,Prep;e/un,D: BVpn(N.b:<modif>=-,
N.t:<det>=N.t:<modif>).

[bonheur],N(N:<num>=sing);avec,Prep: BVpn(N.t:<det>=-).

bus,N;par,Prep;[le],D: BVpdn.

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% GRAMMAIRE TAG DU FRANCAIS
%%
%% DICTIONNAIRE SYNTAXIQUE
%%
%% Phrases fig'ees
%%
%% 300 entr'ees
%%
%% Anne Abeille et Beatrice Daille
%%
%% majuscules :figé; minuscules : libre
%% 0 : sujet; 1,2,3 :compléments
%% arguments nominaux : ni, arguments phrastiques : si
%% adjectifs: a; adverbess : ad
%%
%% Traits syntaxiques: -=interdit; +=obligatoire
%%
%% D1:<Poss-pers>=N0:<pers>, D1:<Poss-num>= N0:<num> => passif=-
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

```

[accuser],V;[choc],N : nOVN1(passif=-,N1.t:<det>=+;N1.b:<det>=-).

[^age],N(N:<num>=sing);[^etre],V;en,Prep;de,Prep1:
nOVNPn1(N0:<hum>=+,S1:<mode>=inf).

[aile],N(N:<num>=pl);[un],D;[pousser],V;a,Prep: DNOVpn1(N1:<hum>=+).

[aises],N;[prendre],V;[son],D: nOVDN1(NOH,D:<Poss-pers>=N0:<pers>, D:<Poss-num>= N0:<num>}

[aller],V;[son],D;[train],N(N:<num>=sing): nOVDN1(D:<Poss-pers>=N0:<pers>,
D:<Poss-num>= N0:<num>}).

[apporter],V;[du],D1;[eau],N1(N:<num>=sing);[a],Prep2;[le]/[son],D2;[moulin
],N2(N:<num>=sing): nOVDN1PdN2(N2:<poss>=+).

[attendre],V;[enfant],N: nOVN1(N1.t:<det>=+;N1.b:<det>=-).

[avoir],V;[don],N(N:<num>=sing);[le],D;de,Prep:
nOVDN1(ps2)(S2:<mode>=inf,N0:<hum>=+, passif=-).

[avoir],V;[vent],N(N:<num>=sing);de,Prep: nOVN1pn2(passif=-N1.t:<det>=-).

[b^atir],V;[du],D1;[ch^ateau],N1(N:<num>=plur);
en,Prep2;Espagne,N2: nOVDN1PN2(N0:<hum>=+, passif=-).

[battre],V;[le],D;[mesure],N(N:<num>=sing): nOVDN1.

[battre],V;[le],D;[campagne],N(N:<num>=sing, passif=-): nOVDN1(N0:<hum>=+).

[battre],V;[son],D;plein,N(N:<num>=sing): nOVDN1(N0:<abs>=+,D:<Poss-pers>=N0:<pers>, D:<Poss-num>= N0:<num>}).

[battre],V;[son],D;coulpe,N(N:<gen>=fem,<num>=sing): nOVDN1(NOH,D:<Poss-pers>=N0:<pers>, D:<Poss-num>= N0:<num>}).


```
;;; -*- Mode: LISP; Syntax: Common-lisp; Package: LEXICON; Base: 10 -*-
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
%%
```

```
%% GRAMMAIRE TAG DU FRANCAIS
```

```
%%
```

```
%% DICTIONNAIRE MORPHOLOGIQUE
```

```
%%
```

```
%% formes flechies des verbes, noms, adjectifs et determinants
```

```
%%
```

```
%% version abregee: 3e personne, ni passe simple, ni conditionnel, part  
passe invariable
```

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
abaisse:[abaisser],V(V.b:<mode>=ind<:subj,V.b:<num>=sing,V.b:<temps>=pres).
```

```
abaissent:[abaisser],V (V.b:<mode>=subj/ind,V.b:<num>=pl).
```

```
abaiss'e:[abaisser],V(V.b:<mode>=ppart).
```

```
abaisser:[abaisser],V(V.b:<mode>=inf).
```

```
abaissait:[abaisser],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=past).
```

```
abaissaient:[abaisser],V(V.b:<mode>=ind,V.b:<num>=pl,V.b:<temps>=past)
```

```
abaissera:[abaisser],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=fut).
```

```
abaisseront:[abaisser],V(V.b:<mode>=ind,V.b:<num>=pl,V.b:<temps>=fut).
```

```
abandon:[abandon],N(N:<num>=sing, <genre>=masc).
```

```
abandons:[abandon],N(N:<num>=pl, <genre>=masc).
```

```
abandonne:[abandonner],V(V.b:<mode>=ind/subj,V.b:<num>=sing,V.b:<temps>=pre  
s).
```

```
abandonnent:[abandonner],V(V.b:<mode>=subj/ind,V.b:<num>=pl).
```

```
abandonner:[abandonner],V(V.b:<mode>=inf).
```

```
abandonn'e:[abandonner],V(V.b:<mode>=ppart).
```

```
abandonnera:[abandonner],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=fut).
```

```
abandonneront:[abandonner],V(V.b:<mode>=ind,V.b:<num>=pl,V.b:<temps>=fut).
```

```
abandonnait:[abandonner],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=past).
```

```
abandonnaient:[abandonner],V(V.b:<mode>=ind,V.b:<num>=pl,V.b:<temps>=past).
```

```
abat:[abattre],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=pres).
```

```
abatte:[abattre],V(V.b:<mode>=subj,V.b:<num>=sing).
```

```
abattent:[abattre],V(V.b:<mode>=subj/ind,V.b:<num>=pl).
```

```
abattra:[abattre],V(V.b:<mode>=ind,V.b:<num>=sing,V.b:<temps>=fut).
```

LISTE DES FIGURES

Liste des notations utilisées.....	2
Chapitre 1	
Fig. 1 Arbres initiaux et arbres auxiliaires.....	12
Fig. 2 Adjonction et substitution.....	13
Fig. 3 exemple d'adjonction.....	13
Fig. 4 exemple de substitution.....	14
Fig. 5 simulation de la substitution par l'adjonction.....	14
Fig. 6 arbres dérivés et arbres de dérivation.....	15
Fig. 7-8 exemples d'unification en grammaire CF.....	17-18
Fig. 9 structures de traits à réentrée.....	18
Fig. 10 unification en TAG (adjonction).....	19
Fig. 11 mise à jour des traits mode, personne et nombre (adjonction).....	20
Fig. 12 unification en TAG (substitution).....	20
Fig. 13 mise à jour des traits personne et nombre (substitution).....	21
Tableau comparatif de l'usage de l'unification en GPSG et TAG.....	23
Fig. 13bis grammaire Tag générant $\{a^{n_b}b^{n_c}c^{n_d}d^{n_e}\}$	24
Fig. 14 grammaire TAG générant $\{ww\}$	25
Fig. 15 grammaire TAG générant des dépendances emboîtées.....	26
Fig. 16 grammaire TAG générant des dépendances croisées.....	26
Tableau comparatif des propriétés formelles des grammaires CF, TAG et CS.....	28
Fig. 17 arbres hors-contexte de <i>caresser, trouver</i>	34
Fig. 18 arbres de <i>N0 prendre une veste, N0 tomber à l'eau</i>	34
Fig. 19 arbres lexicalisés de G2 approximant $S \rightarrow S S$, $S \rightarrow a$ (grammaire CF).....	36
Fig. 20 arbres non générés par G2.....	36
Fig. 21 arbres TAG équivalents à $S \rightarrow S S$, $S \rightarrow a$	36
Fig. 22 exemples d'arbres d'arbres générés par cette grammaire TAG.....	37
Fig. 23 arbres lexicalisés de <i>Jean, dort, beaucoup</i> (avec substitution seulement).....	37
Fig. 24 arbres lexicalisés équivalents à $P \rightarrow SN V$, $V \rightarrow V Adv$, $V \rightarrow dort$; $Adv \rightarrow beaucoup$ (avec substitution seulement).....	38
Fig. 25 arbres lexicalisés de <i>dort, beaucoup</i> (avec adjonction).....	38
Fig. 26 arbres lexicalisés de <i>dort, beaucoup</i> (avec substitution et *)......	39
Fig. 27 schémas de la furcation (droite et gauche).....	40
Fig. 28 arbres de <i>dort, beaucoup</i> avec furcation.....	40
Fig. 29 arbres de <i>peut, dormir</i> avec furcation.....	41
Fig. 29bis arbres de <i>peut, dormir</i> avec adjonction	42
Fig. 30 <i>Paul pense que Jean aime Marie</i> <i>Qui Paul pense que Jean aime</i> (avec furcation).....	43
Fig. 31 <i>Paul pense que Jean aime Marie</i> <i>Qui Paul pense que Jean aime</i> (avec adjonction).....	44
Fig. 32 <i>I think, my friend, that they are wrong</i>	47

Fig. 33 arbres de <i>N0 voler Loc N1</i> et <i>N0 voler N1 à N2</i> , de <i>N0 affecter N1</i> et <i>N0 affecter de P1</i>	49
Fig. 34 arbres de <i>aimer, être aimé par</i>	50
Fig. 35 arbres de <i>attaquer, lancer une attaque contre, avoir une attaque</i>	50
Fig. 36 <i>Je pense qu'il est venu aujourd'hui</i> (arbres dérivés).....	52
Fig. 37 <i>Je pense qu'il est venu aujourd'hui</i> (arbres de dérivation).....	52
Fig. 38 <i>Jean prend une veste</i> (arbre dérivé).....	54
Fig. 39 <i>Jean prend une veste</i> (arbres de dérivation).....	55
Fig. 40 <i>Jean aime Marie</i> vs <i>Marie est aimée par Jean</i>	56

Chapitre 2

Fig. 1 arbres de syntagmes prépositionnels.....	61
Fig. 2 arbres de SP (syntagmes étendus).....	61
Fig. 3 arbres de <i>Bonjour, Non</i>	62
Fig. 4 arbres de phrases à complétive (<i>N0 penser que P, N0 promettre à N1 que P</i>).....	64
Fig. 5 <i>Marie voudrait du whisky et de l'eau</i>	69
Fig. 6 <i>Jean semble dormir</i>	86
Fig. 7 <i>l'amour passionné de Jean pour Marie</i>	97
Fig. 8 <i>Jean aime passionnément Marie</i>	97
Fig. 9 arbres de <i>N0 rêver de N1, N0 penser à N1, pour N, avec N</i>	101

Chapitre 3

Fig. 1 arbres de <i>dormir, aimer, donner à, penser que</i>	105
Fig. 2 arbres de <i>N0 voler Loc N1</i> et <i>N0 voler N1 à N2</i>	106
Fig. 3 arbres de <i>pouvoir, venir de</i>	106
Fig. 4 auxiliaires <i>avoir</i> et <i>être</i>	107
Fig. 5 accord sujet-verbe.....	110
Fig. 6 <i>Jean mange</i> vs <i>Jean semble manger</i>	112
Fig. 7 arbre de <i>P0 plaire à N1</i>	112
Fig. 8 arbres de <i>N0 en vouloir à N1, N0 s'évanouir</i>	113
Fig. 9 arbres de <i>Jean pense que Marie promettra à Paul que Bob partira</i>	114
Fig. 10 arbres de <i>N0 penser que P1, N0 craindre de P1, N0 craindre que P1</i>	114
Fig. 11 <i>Jean est parti</i> vs <i>Jean doit être parti</i>	116
Fig. 12 <i>Paul croit Jean parti</i>	117
Fig. 13 arbres des verbes à complétives et infinitives.....	118
Fig. 14 <i>N0 vouloir que P, N0 vouloir Pinf</i>	118
Fig. 15 arbres de <i>N0 compter sur N1, N0 chercher à ce que P1</i>	119
Fig. 16 arbres des clitiques.....	120
Fig. 17 arbres des clitiques postposés.....	121
Fig. 18 <i>N0 aimer que P</i> vs <i>N0 penser que P</i>	121
Fig. 19 <i>N0 penser à N1</i> vs <i>N0 assister à N1</i>	122
Fig. 20 arbre de l'auxiliaire <i>être</i>	122
Fig. 21 arbres de <i>pour</i>	123
Fig. 22 arbres de <i>selon</i>	124

Fig. 23 arbres de, <i>fier de, rêver de</i>	124
Fig. 24 arbres de <i>N0 mettre N1 Loc N2, N0 aller Loc N1</i>	125
Fig. 25 arbres de <i>femme, femme de, idée que</i>	126
Fig. 26 arbres de <i>forme de, fondement de</i>	126
Fig. 27 <i>un corps, des élèves curieux</i>	129
Fig. 28 <i>un garçon, du blé, (un + du) gâteau</i>	129
Fig. 29 arbres de <i>bien</i>	130
Fig. 30 arbres de <i>N0 croire en N1, N0 transformer N1 en N2</i>	131
Fig. 31 <i>Jean aime une femme</i>	132
Fig. 32 <i>beaucoup de (E+ce) pain</i> (substitution).....	133
Fig. 33 <i>N0 pousser un grand nombre de cris</i> (substitution).....	134
Fig. 34 <i>la plupart de ces habitants, trois de ces filles</i> (adjonction).....	134
Fig. 35 arbres de <i>beaucoup de N, la plupart de N</i>	135
Fig. 36 arbres de <i>la majorité de N, la plupart de N</i>	136
Fig. 37 arbres de <i>plein de N, une partie de N, chaque N</i>	137
Fig. 38 arbres de <i>petit, propre à N et propre</i>	138
Fig. 39 <i>(E+une) semblable aventure</i>	139
Fig. 40 <i>N0 devenir A, N0 rendre N1 A</i>	139
Fig. 41 arbres de <i>assez, conformément à, vraiment</i>	140
Fig. 42 arbres de <i>ici et bien</i>	141
Fig. 43 dérivations synchrones.....	145
Fig. 44 <i>N0 aimer N1 vs N0 être aimé par N1</i> (synchrones).....	146
Fig. 45 <i>A quelle femme pense Jean ? = Jean pense à quelle femme ?</i>	147
Fig. 46 <i>N0 aimer N1 vs N1 être aimé par N0</i> (non synchrones).....	147
Fig. 47 métarègle du passif.....	150
Fig. 48 arbres passifs (<i>N0 aimer N1</i>).....	155
Fig. 49 arbre de <i>être dit, être censé</i>	156
Fig. 50 arbres questions (<i>N0 aimer N1, N0 penser à N1</i>).....	156
Fig. 51 arbres relatives (<i>N0 aimer N1, N0 penser à N1</i>).....	158
Fig. 52 arbres relatives (<i>N0 rêver de P1, P0 plaire à N1, N0 penser que P1</i>).....	158
Fig. 53 arbres clivés (<i>N0 aimer N1</i>).....	159
Fig. 54 arbres clivés (<i>N0 penser à N1</i>).....	160
Fig. 55 arbres extraposés (<i>N0 arriver, P0 déplaire à N1</i>).....	161
Fig. 56 <i>N0 parler à N1 de N2, N0 parler de N2 à N2</i>	163
Fig. 57 arbres neutres (<i>N0 cuire N1, N0 augmenter N1 de N2</i>).....	165
Fig. 58 arbres symétriques (<i>N0 flirter avec N1</i>).....	166
Fig. 59 arbres de <i>dans</i>	168
Fig. 60 arbres de <i>où</i>	168

Chapitre 4

Fig. 1 arbres de <i>N0 faire une gaffe, N0 commettre une agression contre N1, N0 avoir envie de P</i>	174
Fig. 2 arbres de <i>N0 faire la cour à N1, N0 donner (à N1) la réponse que P</i>	175
Fig. 3-4 <i>N0 avoir du courage vs N0 avoir un courage remarquable</i>	177
Fig. 5 <i>N0 prendre une décision vs N0 prendre la décision de P</i>	177

Fig. 6 arbres de <i>avoir tendance à, avoir coutume de</i>	182
Fig. 7 arbres de <i>agression (de N0) (contre N1), de gaffe (de N0)</i>	183
Fig. 8 arbres clivés (<i>N0 commettre une agression contre N1</i>).....	184
Fig. 9 arbres clivés (<i>N0 rapporter N1</i>).....	185
Fig. 10 arbres clivés (<i>N0 commettre une agression contre N1, N0 faire la cour à N2</i>).....	186
Fig. 11 arbres relatives (<i>N0 commettre une agression contre N1</i>).....	187
Fig. 12 arbres relatives (<i>N0 être dans l'ignorance de N1</i>).....	188
Fig. 13 arbres questions (<i>N0 commettre une agression contre N1</i>).....	
Fig. 14 arbres questions (<i>N0 commettre une agression contre N1,</i> <i>N0 faire la rencontre de N1</i>).....	189 190
Fig. 15 arbres passifs (<i>N0 commettre une agression contre N1</i>).....	191
Fig. 16 arbres passif + question, passif + clivée.....	192
Fig. 17 <i>N0 avoir du poids vs N0 donner du poids à N2</i>	196
Fig. 17bis <i>Jean prend la photo de Marie</i> (arbre dérivé, arbres de dérivation).....	
Fig. 18 <i>N0 être en désaccord profond avec N1</i> <i>= N0 être profondément en désaccord avec N1</i>	197
Fig. 19 arbres de <i>N0 tomber à l'eau, N0 perdre les pédales,</i> <i>la moutarde monter au nez de N1</i>	198 206
Fig. 20 arbres de <i>il faut que P, il s'agit de P</i>	207
Fig. 21 <i>Jean prend une veste</i> (arbre dérivé, arbres de dérivation).....	208
Fig. 22 <i>N0 prendre le taureau par les cornes</i>	209
Fig. 23 <i>N0 dire ses quatre vérités à N1, N0 prendre son courage à deux mains</i>	210
Fig. 24 <i>N0 marcher sur les brisées de N1 vs N0 marcher sur ses brisées</i>	211
Fig. 25 arbre de <i>N0 casser les pieds à N1</i>	212
Fig. 26 <i>N0 casser du sucre sur le dos de N2</i>	212
Fig. 27 arbre de <i>N0 passer N1 par les armes, N0 prendre N1 en compte</i>	213
Fig. 28 <i>N0 attendre Det enfant</i>	213
Fig. 29 <i>Jean prend une sacrée veste</i> (arbre dérivé).....	214
Fig. 30 <i>Jean prend Marie en compte, Jean prend une sacrée veste</i> (arbres de dérivation).	214
Fig. 31 arbres passifs (<i>N0 prendre N1 en compte, N0 briser la glace</i>).....	
Fig. 32 arbres questions (<i>N0 prendre N1 en compte, N0 donner le ton à N2</i>	216
<i>N0 dire ses quatre vérités à N2</i>).....	217
Fig. 33 arbres relatives (<i>N0 attendre Det enfant, N0 prendre une veste</i>).....	218
Fig. 34 arbre extraposé (<i>des ailes poussent à N1</i>).....	220
Fig. 35 arbre clivé (<i>N0 prendre N1 en compte</i>).....	221
Fig. 36 arbres clivés (déterminants figés).....	222
Fig. 37 arbres clivés (<i>N0 prendre une veste, N0 chasser le N</i>).....	222
Fig. 38 <i>Jean a pris une sacrée veste = Jean a sacrément pris une veste</i>	225
Fig. 39 <i>N0 séparer le bon grain (individualiste) de l'ivraie (collectiviste)</i>	226

Chapitre 5

Fig. 1 arbres de <i>savoir (si) P, demander à N1 (si) P</i>	233
Fig. 2 arbres de <i>Où ? et Qui construit sa maison ?</i>	233
Fig. 3 <i>Connais-tu l'homme (qui aime Marie + que Marie aime) ?</i>	235
Fig. 4 arbres relatifs <i>le magasin où Jean achète une montre</i>	236

Fig. 5 une agression de (Jean + qui) contre (Marie + qui) ?.....	236
Fig. 6 arbre question (contre qui commettre une agression).....	237
Fig. 7 Jean désapprouve le projet d'aller où ?	
* Où Jean désapprouve-t-il le projet d'aller ?.....	238
Fig. 8 Où Jean a-t-il le projet d'aller ?.....	238
Fig. 9 Paul trouve Marie contente de quoi ?.....	240
Fig. 10 De quoi Paul est-il content ?.....	240
Fig. 11 Paul trouve Marie contente que tu connaisses Pierre	
* Qui Paul trouve-t-il Marie contente que tu connaisses ?	241
Fig. 12 Marie est-elle contente que tu connaisses Pierre	
Qui Marie est-elle contente que tu connaisses ?.....	242
Fig. 13 arbres questions et clivées de N0 penser à N1.....	243
Fig. 14 Bob est parti pour N.....	244
Fig. 15 parce que Marie aime qui ?	
* qui parce que Marie aime ?.....	244
Fig. 16 Jean aime le chocolat et Paul préfère le café	245
Fig. 17 sous-arbres avec trait SLASH en GPSG.....	249
Fig. 18 sous-arbres avec trait SLASH en GPSG.....	250
Fig. 19 Bob, Mary loves et Bob, John thinks Mary loves (en LFG).....	252
Fig. 20 arbres élémentaires de N0 aimer N1.....	253
Fig. 21 Qui Jean semble-t-il avoir tendance à apprécier ?.....	254
Fig. 22 arbres de N0 penser que P1, N0 promettre de/que P1 à N2	254
Fig. 23 Quelle fille Paul pense-t-il que Bob sait que Jean aime ?	255
Fig. 24 Où Jean cherche-t-il à ce que Marie aille ?.....	256
Fig. 25 arbres de N0 dormir, N0 aimer N1 (infinitives).....	256
Fig. 26 arbres de N0 vouloir Pinf, N0 demander Pinf.....	257
Fig. 27 N0 aimer N1 (infinitives).....	257
Fig. 28 Paul croit Jean parti.....	258
Fig. 29 Quelle fille Paul pense-t-il que Bob sait que Jean aime ? (arbre de dérivation).....	258
Fig. 30 Quelle fille Bob a-t-il l'impression que Paul est convaincu que Marie aime ?.....	259
Fig. 31 arbres de dans N.....	260
Fig. 32 Dans quelle maison Paul pense-t-il que Marie travaille ?.....	260
Fig. 33 arbres questions de N0 aimer N1.....	261
Fig. 34 arbres relatives de N0 penser que P1.....	262
Fig. 35 C'est Paul qui pense que Jean aime Marie	
* Qui est-ce Paul qui pense que Jean aime ?	263
Fig. 36 arbres élémentaires de P0 revenir à P1, N0 déduire P1 de P2.....	264
Fig. 37 arbre passif N0 montrer que P1.....	265
Fig. 38 arbres élémentaires de P0 ennuyer N1, Il ennuyer N1 de P2.....	265
Fig. 39 Paul demande (qui dort + où Marie dort).....	266
Fig. 40 Paul demande qui pense que Jean aime Marie	
* Qui Paul demande-t-il qui pense que Jean aime ?.....	267
Fig. 41 arbres élémentaires de N0 dire que P, N0 s'exclamer que P.....	268
Fig. 42 Jean rit pendant que Marie lit le journal.....	269
Fig. 43 arbre Qui Jean aime.....	270
Fig. 44 arbres élémentaires de est-ce que, -il, N0 demander P1.....	271

Fig. 45 contraintes d'inversion.....	271
Fig. 46 <i>Qui Marie pense-t-elle que Paul croit que Jean aime ?</i>	272
Fig. 47 <i>Je demande qui Paul croit que Jean aime</i>	273

Chapitre 6

Fig. 1 arbre dérivé de <i>Il le voit</i>	275
Fig. 2 arbre dérivé de <i>N0 en rêve</i>	275
Fig. 3 <i>N0 le mange</i>	276
Fig. 4 arbres élémentaires de <i>ils, elle</i> (inversés).....	277
Fig. 5 <i>Il part, Jean et Jean le donne à Marie, le livre</i>	277
fig. 6 arbres dérivés de <i>Jean semble partir, la branche semble casser</i>	280
Fig. 7 arbres élémentaires de <i>N0 casser, N0 partir</i> (infinitives).....	281
Fig. 8 arbres élémentaires de <i>N0 espérer P⁰inf, N0 permettre à N1 de P^Iinf</i>	282
Fig. 9 <i>Jean propose à ses enfants d'être élus à sa place</i>	283
Fig. 10 <i>Jean semble partir = Il semble que Jean part</i>	
(arbres dérivés, arbres de dérivation).....	284
Fig. 11 <i>Jean partira probablement</i>	287
Fig. 12 <i>Un portrait est arrivé au musée, de cette fille que tu connais</i>	288
Fig. 13 <i>Combien Paul vend-il de tableaux ?</i>	289
Fig. 14 arbres élémentaires de <i>et</i>	290
Fig. 15 <i>Jean aime Marie et déteste Jeanne</i>	292
Fig. 16 <i>Jean dort et rit</i> (arbre dérivé).....	292

Chapitre 7

Fig. 1 arbres élémentaires sélectionnés pour <i>Jean aime Marie</i>	300
Fig. 2 mêmes arbres avec structures de traits	301
Fig. 3 arbres élémentaires sélectionnés pour <i>l'idée que Marie parte,</i> <i>le gâteau que Marie rejette</i>	303
Fig. 4 arbres de la grammaire engendrant <i>aaaax</i> ou <i>aaaay</i>	304
Fig. 5 arbres élémentaires sélectionnés pour <i>Jean prend une veste</i>	306
Fig. 6 arbres élémentaires de la famille de <i>N0 briser la glace</i>	306
Fig. 7 arbres élémentaires sélectionnés pour <i>Jean travaille pour Marie</i>	309
Fig. 8 arbres élémentaires pour <i>Jean prend une veste</i>	309
Fig. 9 arbres de <i>très</i>	310
Fig. 10 arbres élémentaires sélectionnés pour <i>Jean cueille une fleur très rouge</i>	310
Fig. 11 arbres élémentaires sélectionnés pour <i>Jean mange une poire très ferme</i>	311
Fig. 12 arbres de dérivation de <i>Jean casse sa pipe</i>	313
Fig. 13 arbres élémentaires de <i>la nuit</i>	314
Fig. 14 arbres élémentaires de <i>souhaiter</i>	314
Fig. 15 arbres élémentaires de <i>avec</i>	315

Fig. 16 <i>Que Jean ait pris une veste surprend Marie</i> = <i>la veste que Jean a prise surprend Marie</i>	319
Fig. 17 arbres élémentaires synchrones de <i>N0 commettre une agression contre N1</i> et <i>N0 agresser N1</i>	320
Fig. 18 arbres de dérivation synchrones pour <i>Jean a commis une récente agression contre Marie</i> et <i>Jean a récemment agressé Marie</i>	321
Fig. 19 arbres élémentaires synchrones de <i>N0 faire allusion à N1</i> et <i>N0 allude to N1</i>	322



Anne Abeillé

Une Grammaire lexicalisée d'arbres adjoints pour le français

ERRATUM

Remerciements. 3e ligne, lire "active illie"

Introduction

p. 6 . 4e §: lire 1-capacité générative forte insuffisante

p. 10 . fin du 3e §: En particulier, (et non En partie)

Chapitre 1

p. 20. Figure 11: Après unification
aux noeuds $V <pers> \neq 2$ (et non $<pers> = 3$)

p. 24. Figure : les deux premiers arbres sont la grammaire TAG engendrant le langage $\{a^n b^n c^n d^n\}$, le troisième est un exemple d'arbre dérivé engendré par cette grammaire.

p. 34. Ligne 3: Aux arbres α_1 et α_3 ci-dessus correspondent...

p. 41. 3e §, avant-dernière ligne: partagent (et non partage) le même noeud

Chapitre 2

p. 72. lire: *les premiers* (au lieu de: les seconds) et *les seconds* (au lieu de: les premiers);
ligne 12

p. 75. lire: *la* radio (cinquième exemple)

Chapitre 3

p. 117. Note 12: * *Jean apprend à ce que P*

p. 118. Note 14, lire: en disjonction avec *de/à* (et non avec *que*).

p. 125. 5e §, lire : un SN ou un adverbe locatif (et non un SN de type concret ou ...)

p. 134. Figure 34. arbre de *trois*: pas de trait $<det>$ au noeud N racine.

p. 135. Figures 35 et 36. arbre de *la plupart de* : au noeud N racine $<def> = +$ (et non -).

p. 146. 2e ligne, lire: "N1 dans (1) et N0 dans (2)"

p. 151. section 3.3, 3e § : *Faire cela démange Luc*

Chapitre 4

p. 158. ligne 2, lire: "tu penses" (et non: du penses).

Figure 45, 2e arbre, lire : à en dessous du noeud Prep

p. 159. section 3.4, lire: *ce* (ou *c'*) et une forme de [être] (et non *c'est*)

Figure 46, 2e arbre: N1↓(et non N1)

|
ei

p. 160. Figure 47, 2e arbre noeud foyer : N (et non N↓<qu> = +)

|
qui/[lequel]

p. 168. fin 1er §, lire "les arbres de la question" (et non: les arbres de la question ou de la relative...)

Figure 53, 2e arbre: supprimer la branche Adv

|
ei

3e arbre: <qu> = + au noeud P intérieur, <qu> = - au noeud P feuille

p. 177. Figure 3, 2e arbre, sur le noeud N dominant *courage*, lire:

t:<modif> = ~ D:<part>

b:<modif> = -

1ere ligne, lire "impossible en (2)" (et non en (1))

3e ligne, lire "obligatoire en (1)" (et non en (2))

p. 181. lire: "Elles sont suivies de l'infinitif..."

p. 183. 3e §, lire "G. Gross 1990" (et non L. Danlos 1991)

p. 189. Figure 13, arbre 2: ajouter b:<qu> = N1:<qu> au noeud N dominant "agression"

p. 201. avant-dernier §, lire "pas de figement sujet-verbe par exemple" (et non: pas de figement sujet-objet)

p. 202 : b) M. Salkoff 1979 (et non a))

p. 203 : c) Les grammaires d'unification (et non b))

p. 205. section 2.2, avant-dernière ligne, omettre : (ni morphologique)

p. 208. entrée lexicale de *N0 prendre le taureau par les cornes*:
famille N0VDN1PDN2

p. 216. Figure 31, 1er arbre, lire : N2 (dernière feuille à droite)

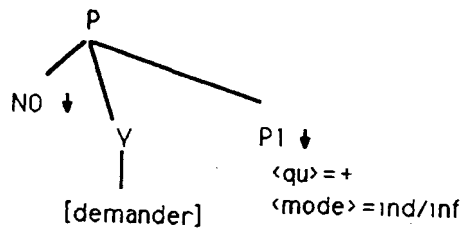
p. 219. section d), 1er exemple: *Il pend au nez de Luc que P*

p. 221. Note 28 : Les têtes lexicales sont notées $\diamond$.

p. 224. lire *le* glas (premier exemple)

Chapitre 5

p. 233 Figure 1: lire:



p. 243. Figure 15, arbre (2) : N (et non $N \downarrow <qu> = +$)
qui/[lequel]

Chapitre 6

p. 279. au lieu de "Nous mentionnerons...", lire : "Le même traitement pourrait s'appliquer aux sujets des noms prédicatifs".

p. 281. 2e §, 2e ligne, lire: ($\uparrow$ Subj) = ($\uparrow$ Obj)

p. 285. section 3, 2e § lire : Joshi 1987b (et non Joshi 1988)

p. 292. 3e §, 2e ligne, lire "correspondant à deux fonctions de même ..."
2) lire "à condition qu'ils soient de même catégorie.." (et non elles)

Chapitre 7

p. 304. Figure 4



3e ligne en-dessous, lire: on élimine 3 (dans le 1er cas) et 2 (dans le 2d cas)

p. 306. 2e §, 2e ligne, lire: "ou ceux de *N0* prendre *N1* à *N2*" (et non ainsi, ceux...)

p. 315. 3e §, 5e ligne, lire " (*Jean pense que P*) est beaucoup plus naturelle que celle à relative (*Jean pense faire W, malgré les éléments que Paul saura*)"

Bibliographie

manquent:

A. Borillo, 1971. "Remarques sur les verbes symétriques français", *Langue Française*, n°11.

N. Chomsky, 1986. *Barriers*, MIT Press.

D. Godard, 1988. "Sujets et compléments génitifs dans le groupe nominal", in Milner (ed), *Recherches nouvelles sur le langage*, Publications de l'ERA 642, Université Paris 7.

A. Joshi, 1987b. "The relevance of Tree Adjoining Grammar to generation", in G. Kempen (ed), *Natural Language Generation*, Dordrecht.

C. Leclère, 1971. "Remarques sur les substantifs opérateurs", *Langue Française*, n°11.

J-C. Milner, 1989. *Introduction à une science du langage*, Seuil.

Annexe 1

p. 337. valeurs du trait mode : ind,imp,subj,inf,ppart
manque aussi le trait mass = + :- portant sur N

Annexe 2

p. 345 (arbre Wn0Vn1), p.349 (arbre Wn0Vn1(pass)) : manquent traits <inv> et <princ> sur les noeuds P_T et P.

191 ARE