

EXTRACTIONS DANS UNE GRAMMAIRE DE DÉPENDANCE LEXICALISÉE À BULLES

Sylvain KAHANE*

Résumé - Abstract

Cet article propose un traitement des différents cas d'extraction en français par une grammaire de dépendance lexicalisée qui construit des arbres à bulles comme représentation syntaxique. Les arbres à bulles sont des arbres dont les nœuds peuvent être occupés par des bulles, qui peuvent contenir d'autres nœuds qui peuvent, à leur tour, être occupés par des bulles ou avoir leurs propres dépendants. Notre grammaire nécessite des opérations de combinaison des structures élémentaires plus complexes que d'autres grammaires lexicalisées comme TAG ou CG, mais évite la multiplication des structures élémentaires et permet des traitements linguistiques bien motivés.

The paper proposes a treatment of different cases of extraction in French by a lexicalized dependency grammar which builds bubble trees as syntactic representations. Bubbles trees are trees whose nodes can be filled by bubbles, which can contain others nodes which can, in their turn, be filled by bubbles or have their own dependents. Our grammar needs more complex operations of combination of elementary structures than other lexicalized grammars, such as TAG or CG, but avoids the multiplication of elementary structures and provides well motivated linguistic treatments.

Mots clefs - Keywords

Extraction, verbe à montée, grammaire de dépendance, grammaire lexicalisée, théorie Sens-Texte, dépendance syntaxique, dépendance sémantique, nucléus verbal, arbre à bulles.

Extraction, raising verb, dependency grammar, lexicalized grammar, Meaning-Text theory, syntactic dependency, semantic dependency, verbal nucleus, bubble tree.

* LaTTiCe/TALaNa (Univ. Paris 7) & Univ. Paris 10 - Nanterre (sk@ccr.jussieu.fr)
Le développement de ce formalisme est en partie conséquent à un début de collaboration avec A. Nasr. Ce travail a également bénéficié de discussions fructueuses et/ou de relectures de A. Abeillé, M.-H. Candito, K. Gerdes, J. Milićević, G.-J. Kruijff, I. Mel'čuk, A. Nasr, A. Polguère et un lecteur anonyme de la revue T.A.L.

INTRODUCTION

Dans cet article, nous nous intéressons au traitement des extractions et plus particulièrement de la relativisation. L'accent portera sur l'étude de la dépendance non bornée (= le fait que le groupe extrait peut ne pas dépendre directement du sommet de la relative) et du pied-piping (= le fait que le pronom relatif peut ne pas être le sommet du groupe extrait), illustrés par :

- (1) *Marie connaît la dame **sur le mari de laquelle** je pense que tu peux compter.* (le groupe extrait de la relative est en gras)

Nous proposons également une solution simple au fameux problème de l'alternance *qui/que* en français (Kayne 1975) :

- (2) a. *la personne que je pense **que** tu connais*
b. *la personne que je pense **qui** te connaît*

Nous étudierons aussi l'inversion (optionnelle) du sujet dans les extractions :

- (3) *le livre que veut donner **Marie** à son frère*

Le traitement des verbes à montée et copule sera également abordé pour introduire le formalisme et par contraste avec le traitement des extractions.

Notre description des extractions se fera dans le cadre d'une grammaire de dépendance lexicalisée inspirée de la Théorie Sens-Texte [= TST] et des Grammaires d'adjonction d'arbres [= TAG], à la suite de Nasr 1995, 1996. Notre grammaire se particularise par le fait qu'elle utilise comme représentation syntaxique une structure plus complexe que ces formalismes, l'arbre à bulles (Kahane 1997). Des nouvelles opérations de combinaison des structures élémentaires sont définies en conséquence.

Dans la Section 1, nous présentons l'organisation générale de la grammaire et notamment la distinction entre lexies (et grammèmes) profonds et de surface. Dans la Section 2, nous présentons les arbres à bulles, la représentation des extractions et une grammaire pour traiter les extractions. Une comparaison sera effectuée avec le traitement d'autres formalismes (TAG, CG, HPSG, Syntaxe X' et les grammaires de dépendance).

1. UNE GRAMMAIRE DE DEPENDANCE SANS BULLES

Retenant les postulats de la TST (Mel'čuk 1988 : 53), nous considérons 1) qu'une **grammaire**¹ est un système formel permettant d'assurer la correspondance entre des textes et des sens (= des représentations sémantiques) et 2) qu'un tel système est modulaire et comprend des modules de correspondance entre des niveaux de représentation intermédiaires, notamment un niveau morphologique et un niveau syntaxique. Nous nous limiterons ici à présenter un fragment de grammaire de la phrase pour le français écrit.

¹ Le terme *grammaire* est ici utilisé au sens chomskien de modèle linguistique.

1.1 Les différents niveaux de représentation

La **représentation morphologique** d'une phrase est la suite des représentations morphologiques des mots de la phrase ; la représentation morphologique d'un mot est une **lexie de surface** accompagnée éventuellement d'une liste de **grammèmes de surface**. Considérons la phrase :

(4) *Marie a parlé à un type étrange.*

La représentation morphologique de (4) est :

(5) MARIE_{sg} AVOIR_{ind,présent,3,sg} PARLER_{-é} À UN_{masc,sg} TYPE_{sg} ÉTRANGE_{masc,sg}

La **représentation syntaxique** d'une phrase est un **arbre de dépendance** non ordonné, à la façon des arbres de dépendance syntaxique de surface de la TST (Mel'čuk 1988), eux-mêmes largement inspirés des stemmas de Tesnière 1959. Les nœuds de la structure sont étiquetés d'une **lexie de surface** accompagnée d'une liste de grammèmes de surface² et les dépendances sont étiquetées par des **fonctions syntaxiques** (de surface).

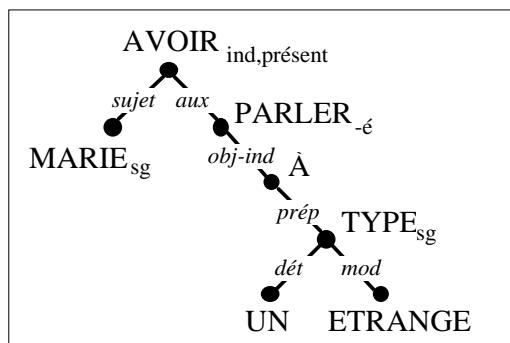


Figure 1. La représentation syntaxique de (4)

La **représentation sémantique** d'une phrase est un graphe orienté dont les nœuds sont étiquetés par une lexie profonde accompagnée d'une liste de grammèmes profonds. Une **lexie profonde** correspond à une lexie de surface ou à un groupe de lexies de surface qui forment sémantiquement un tout (= **phrasème**). Un **grammème profond** est un grammème non sémantiquement vide³; un grammème profond peut être exprimé par un grammème de surface ou par une expression complexe incluant des lexies de surface : c'est le cas, par ex., du passé composé (AVOIR_{ind,présent}+é) ou de la voix passive (ÊTRE+é). Une lexie profonde fonctionne comme un prédicat et

² La seule différence avec la représentation morphologique d'un mot est que les grammèmes d'accord ou de régime (comme les cas) ne sont pas présents. Cette remarque est surtout pertinente pour la synthèse, où ces grammèmes seront introduits par le module syntaxique.

³ Nous y inclus la voix, qui n'exprime pas réellement un sens, mais possède une valeur communicative certaine.

est reliée à ses arguments par des arcs pointant sur eux. Les différents arcs émergeant d'une lexie profonde sont numérotés de 1 à n par ordre d'oblicité croissante des arguments. Un tel arc, qui représente une relation prédicat-argument, est appelé une **dépendance sémantique** ; le prédicat est appelé **gouverneur sémantique** et l'argument **dépendant sémantique**.

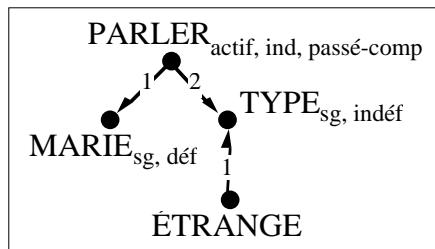


Figure 2. La représentation sémantique de (4)

Notre représentation sémantique est en fait un compromis entre les représentations sémantique et syntaxique profonde de la TST. D'une part, les nœuds sont étiquetés par des lexies profondes plutôt que par des sémantèmes (= le signifié d'une lexie profonde) : nous estimons que la distinction entre lexie profonde et sémantème faite par Mel'čuk 1988 n'est pas pertinente en raison de la correspondance biunivoque entre les deux⁴. D'autre part, à la structure de graphe vient se superposer une hiérarchie sur les nœuds, que nous appellerons, suivant Polguère 1990, la **dépendance communicative**, qui exprime l'importance communicative des sens exprimés dans la phrase et qui se retrouve encodée par la structure d'arbre au niveau syntaxique. Dans la tradition Sens-Texte, seuls les **sommets communi-catifs**⁵ (appelés nœuds dominants chez Polguère 1990) de certains sous-graphes (notamment le rhème et le thème) sont indiqués. Nous préférons indiquer toute la hiérarchie communicative. Ainsi, dans les figures, un gouverneur communicatif est placé au-dessus de ses dépendants communicatifs. Lorsque dépendance sémantique et dépendance communicative sont dans le même sens (flèche vers le bas), le dépendant est un **actant** (ou argument), tandis que lorsque ces dépendances sont inversées (flèche vers le haut), le dépendant communicatif (et gouverneur sémantique) est un **modifieur**. Avec cette structure hiérarchique, notre représentation sémantique s'apparente à un arbre de dérivation TAG (cf. Candito & Kahane 1998a pour une comparaison entre TAG et TST) et plus particulièrement à la représentation sémantique adoptée par

⁴ La représentation sémantique de la TST a initialement pour objectif d'être un invariant de paraphrase, ce qui nécessite d'associer à deux lexies synonymes le même sémantème. Nous préférons considérer le remplacement d'une lexie par une lexie synonyme comme une règle de paraphrase (voir Mel'čuk 1992 pour une présentation détaillée de ces règles).

⁵ Le sommet communicatif d'un syntagme est le sémantème qui en « résume » le sens (cf. Zwicky 1985). Ainsi, pour reprendre un exemple classique, un *verre à vin* est un *verre*, alors qu'un *verre de vin* est du *vin*. Remarquons que dans les deux cas, *verre* est pourtant le gouverneur sémantique de *vin*.

Danlos 1998 dans son formalisme de génération basé sur les TAG. Cf. également Dymetman & Copperman 1996.

Dans l'exemple précédent, toute dépendance communicative correspondait à une dépendance sémantique, ce qui permettait d'indiquer la dépendance communicative simplement par l'orientation des dépendances sémantiques (vers le haut ou vers le bas). Le cas des propositions relatives s'avère plus complexe (cf. Figure 3).

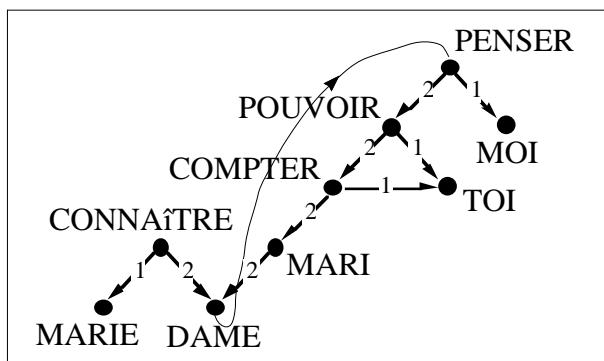


Figure 3. La représentation sémantique (sans grammèmes) de (1)

La proposition relative modifie son antécédent et dépend donc communicativement de lui, mais, par contre, il n'y a pas en général de dépendance sémantique entre l'antécédent (ici DAME) et le sommet hiérarchique de la relative (ici PENSER) : la dépendance communicative entre les deux devra donc être représenté par un lien propre (que nous représentons en fin et sans étiquette). Par ailleurs, l'antécédent occupe une position sémantique dans la relative (ici comme argument de MARI) et il est donc, en tant qu'élément de la proposition relative, dominé communicativement par le sommet hiérarchique de celle-ci. Notre représentation sémantique de la relative s'appuie sur l'étude détaillée de Kahane & Mel'čuk 1999. On notera également le nœud TOI qui est à la fois le premier argument de POUVOIR et COMPTER, créant ainsi un cycle dans le graphe sémantique.

Pour terminer, il faut dire que notre représentation sémantique est loin d'être complète et qu'il lui manque en particulier des informations sur la portée des opérateurs logiques ou sur la structure communicative comme l'organisation en rhème-thème (cf. Sgall *et al.* 1986 ou Mel'čuk à paraître pour une discussion sur ce dernier point dans le cadre des grammaires de dépendance).

Notre grammaire est composée de trois modules : un **module morphologique** assurant la correspondance entre phrases (= suites de mots) et représentations morphologiques, un **module syntaxique** assurant la correspondance entre représentations morphologiques et syntaxiques et un **module sémantique** assurant la correspondance entre représentations syntaxiques et sémantiques. Nous nous intéresserons à ces deux derniers

modules. Notons que dans la plupart des grammaires lexicalisées, comme TAG (Abeillé 1991, XTAG 1995, Candito 199199) ou les grammaires catégorielles [= CG] (Steedman 1985, 1997, Moortgat 1988, Morrill 1993), la correspondance entre niveaux morphologique et sémantique est effectuée par un unique module. Enfin, à l'instar des présentations de la plupart des modèles et à la différence des présentations de la TST, nous privilierons plutôt le sens de l'analyse.

1.2 Le module syntaxique (sans les bulles)

Notre formalisme repose sur la proposition de Nasr 1995 et 1996, elle-même inspirée du formalisme TAG (Joshi 1987). Il s'agit d'une **grammaire lexicalisée**, c'est-à-dire que la grammaire est réduite à un lexique où chaque lexie de surface est associé à une ou plusieurs structures élémentaires (où sont également encodées les informations grammaticales).

1.2.1 Structures élémentaires

Une **structure élémentaire** est une portion de représentation syntaxique enrichie d'un certain nombre de traits. Les **nœuds** de la structure sont étiquetés par une structure de traits comprenant des traits pour la lexie, les grammèmes correspondants et toutes les informations syntaxiques pertinentes (partie du discours, traits d'accord, ...). Les **dépendances** sont étiquetées par une structure de traits comprenant les traits fonction (pour la **fonction syntaxique**) et option (indiquant le caractère optionnel ou non de la dépendance), ainsi que les traits poids, quasi et nucleus que nous introduirons par la suite. De plus, les éléments, nœuds et dépendances, sont typés par un trait binaire type. Les éléments de type:1 et type:0 seront dit **noir** et **blanc** et représentés en noir et blanc pour les nœud et en gras et fin pour les dépendances. Les éléments noirs sont les éléments réellement introduits par la structure : il s'agit, dans notre grammaire, du nœud étiqueté par la lexie de référence, appelé **l'ancre**, et de certaines dépendances qui en sortent. Les éléments blancs peuvent être interprétés comme des **requêtes** de l'ancre⁶.

Pour faciliter la lecture de nos figures, seule est indiquée les valeurs des traits instanciés. Ainsi AVOIR_{(V)ind} correspond à la structure de traits [lex: AVOIR,

⁶ Notre formalisme permet par ex. de spécifier (par une dépendance blanche) qu'un nom doit avoir un déterminant. Le formalisme TAG ne permet pas de spécifier le contexte et ces questions sont résolues par l'introduction d'un trait ad hoc (Abeillé 1991). En fait, on peut montrer qu'une grammaire avec des dépendances blanches est fortement équivalente à une grammaire sans dépendances blanches, mais où de nouveaux traits ont été introduits dans l'étiquetage des nœuds. Remarquons également que notre formalisme n'empêche pas le nom MARIE de recevoir un déterminant, comme dans *la belle Marie*. Par contre, pour assurer qu'un nom possède un unique déterminant, il nous faut un outil supplémentaire. On peut adopter la solution préconisée dans le cadre de la TST consistant à typer les fonctions syntaxiques en *répétable* et *non répétable* et en introduisant un mécanisme vérifiant qu'un nœud ne possède pas deux dépendances d'une même fonction non répétable. Une autre solution, que nous n'étudierons pas, consiste à considérer que le déterminant gouverne le nom et non l'inverse (Hudson 1999).

cat: V, mode: ind, temps: t , pers: x , nbre: y] où t , x et y sont des variables. L'étiquette (N) du nœud *sujet* de la même structure vaut pour la structure de traits [lex: w , cat: N, gen: z , pers: x , nbre: y]. La cooccurrence des valeurs x et y des traits pers et nbre des deux nœuds assurent l'accord du verbe avec son sujet. Pour les différencier des grammèmes, les traits morphologiques utilisés pour l'accord, sont placés entre parenthèses ; par ex., le genre gen est un grammème pour l'adjectif, mais pas pour le nom. Une disjonction de valeurs est séparée par /.

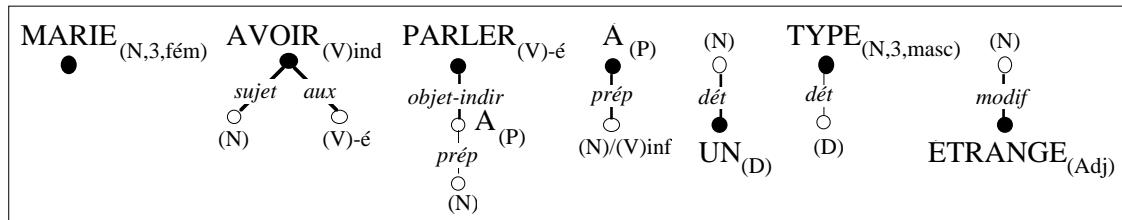
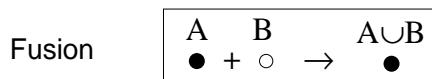


Figure 4. Des exemples de structures élémentaires

Remarquons que malgré la redondance certaine entre les structures de PARLER et À, nous pensons utile de considérer une structure élémentaire séparée pour la préposition À. Donnons deux raisons : 1) lors de l'analyse incrémentale d'une phrase comme *Pierre se demande à qui tu penses parler*, À pourra être lié à QUI avant que PARLER ne soit analysé (ce que confirme notre intuition linguistique) ; 2) dans une coordination, la préposition peut être répétée (*Marie parle à Pierre et à Jean*).⁷

1.2.2 Combinaison des structures élémentaires

Les structures élémentaires se **combinent** ensuite par une seule opération, la **fusion** d'un élément noir et d'un élément blanc, qui donne un élément noir ou la fusion de deux éléments blancs, qui donne un élément blanc. Deux éléments peuvent être fusionnés si leurs étiquettes A et B sont unifiables; l'élément résultant reçoit l'étiquette $A \cup B$, obtenue par unification des étiquettes A et B.



Le trait type possède un fonctionnement particulier pour l'unification ($0 \cup 0 = 0$, $0 \cup 1 = 1$ et $1 \cup 1 = \text{échec}$) assurant que deux éléments noirs ne puissent pas

⁷ Il pourrait même être préférable de spécifier le gouverneur dans la structure élémentaire d'une préposition vide, ce qui aurait l'avantage d'obtenir une structure parallèle à celle d'une préposition pleine (comme la préposition À dans *un livre à lire*). Le cas des prépositions locatives est de ce point de vue particulièrement intéressant : bien qu'on puisse considérer que la préposition est toujours pleine, celle-ci peut figurer dans le régime d'un verbe. Ainsi le complément locatif figure dans le régime de ALLER (*Pierre va à Paris*), mais pas de DORMIR (*Pierre dort à Paris*). On peut associer à la préposition locative À la même structure élémentaire dans les deux cas en indiquant que la dépendance vers le gouverneur de À peut être blanche ou noire (selon que la préposition est régie ou non et que la dépendance prép est ou non construite par le gouverneur de la préposition).

fusionner. La combinaison de deux structures est également contrainte par le fait que la structure résultante doit être un arbre.

Remarquons que l'opération de fusion subsume à la fois l'opération de substitution et une partie des opérations d'adjonction (la *sister-adjunction* pour être précis) de TAG : on parlera de **substitution** lorsque l'arc noir est introduit par le gouverneur syntaxique et d'**adjonction** lorsque l'arc noir est introduit par le dépendant syntaxique.

Enfin, des représentations morphologiques M et syntaxique S se **correspondent** si on peut associer à chaque lexie de M une structure élémentaire et si ces structures peuvent être combinées afin de former l'arbre S (Figure 5).

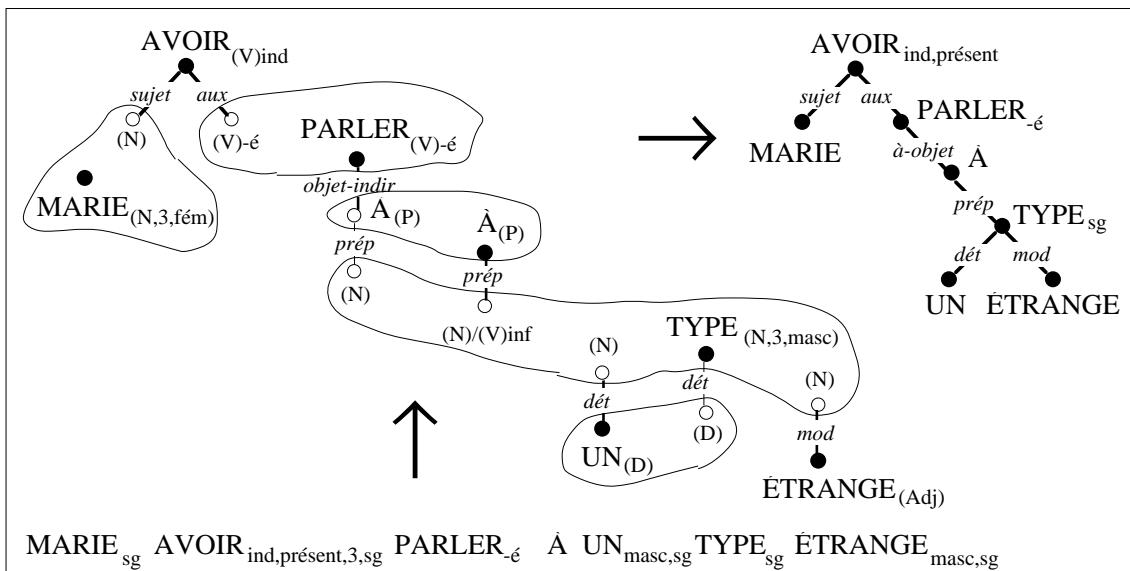


Figure 5. Un exemple de correspondance syntaxique

1.2.3 Ordre linéaire et projectivité

Il nous reste à voir comment est pris en compte l'ordre linéaire. Un arbre de dépendance assorti d'un ordre linéaire sur les nœuds est dit **projectif** si et seulement si, en plaçant les nœuds sur une ligne droite et tous les arcs dans le même demi-plan, on peut assurer que 1) deux arcs ne se coupent jamais et que 2) aucun arc ne couvre la racine de l'arbre (Figure 7)⁸. Cette propriété est équivalente au fait que la **projection** de tout nœud x de l'arbre (c'est-à-dire l'ensemble des nœuds dominés par x , x compris) forme un segment continu de la phrase (Lecerf 1961).

⁸ Suivant Hudson 2000, nous représenterons la racine de l'arbre avec une dépendance gouverneur verticale (potentiellement infinie). La condition (2) se ramène alors à un cas particulier de (1).

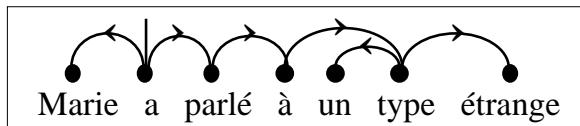


Figure 6. Un exemple d'arbre de dépendance projectif

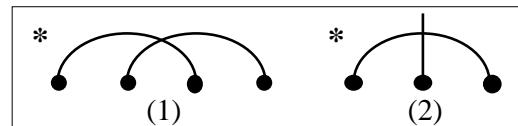


Figure 7. Les cas de non projectivité

La projectivité présente un intérêt immédiat : il suffit pour ordonner un arbre projectif de spécifier la position de chaque nœud par rapport à son gouverneur, ainsi que vis-à-vis de ses frères (Figure 6). Pour ce faire, nous suivrons la proposition de Mel'čuk 1967 (voir également Courtin & Gentilhal 1998) consistant à assigner à chaque dépendance un poids⁹ (par un trait poids) (Figure 8) et à assumer les **contraintes d'ordre locales** suivantes : si le poids d'une dépendance est négatif le dépendant doit se trouver à gauche de son gouverneur, s'il est positif, le dépendant doit se trouver à droite ; quant à deux frères, celui portant le poids le plus important (en valeur absolue) sera le plus éloigné du gouverneur.¹⁰ Ces contraintes sont **locales** dans la mesure où elles mettent en jeu une portion limitée de l'arbre, à savoir un nœud et ses dépendants directs.

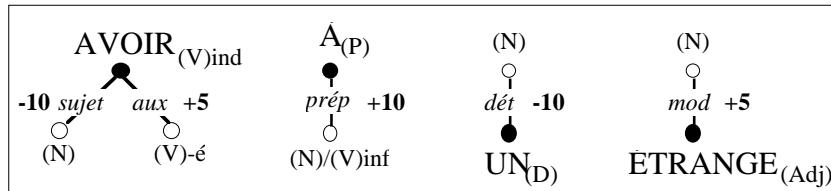


Figure 8. Des exemples de structures élémentaires avec poids pour la linéarisation

La correspondance peut maintenant être complètement définie : des représentations morphologique M et syntaxique S se **correspondent** si on peut associer à chaque lexie de M une structure élémentaire, si ces structures peuvent être combinées afin de former l'arbre S et si la structure produit MxS, qui est un arbre ordonné, vérifie la projectivité et les contraintes d'ordre locales.

⁹ D'autres solutions sont possibles. On peut par exemple écrire des règles de précédence linéaire comme en GPSG (Gazdar *et al.* 1985) ordonnant un nœud et son gouverneur ou bien deux dépendants. On peut également, comme Nasr 1996, en s'inspirant de Mel'čuk & Pertsov 1987, indiquer pour chaque nœud la position par rapport à son gouverneur (droite ou gauche) et la liste des autres dépendants autorisés à se trouver entre eux.

¹⁰ Nous ne discuterons pas davantage la question des poids par la suite. Notons simplement que le poids dépend de différents facteurs : la fonction syntaxique (en français par ex., les syntagmes possèdent une légère tendance à apparaître dans l'ordre d'obligé croissante), la taille et la nature du syntagme (un syntagme de taille importante a tendance à être plus éloigné de son gouverneur qu'un syntagme de petite taille; par ailleurs certaines positions n'autorisent que des syntagmes "lexicaux" : *urhaut mur* / **un haut de 3m mur* (Blinkenberg 1928, Abeillé & Godard 1998) et la saillance communicative (l'importance ou le rôle donné par le locuteur au syntagme). Le poids devrait donc être une fonction de ces différents paramètres.

Nous verrons dans la suite comment traiter un cas bien connu de construction non projective, l'extraction.

1.2.4 Un analyseur à pile associé à la grammaire

Le fait d'imposer la projectivité permet de construire facilement un automate à pile. Nous présentons ici de façon schématique l'analyseur décrit dans Nasr 1995, 1996. L'analyse d'une phrase requiert deux opérations : l'empilement et la réduction. L' **empilement** consiste simplement à mettre dans la pile une structure élémentaire à chaque fois qu'un mot est lu. La **réduction** consiste à fusionner deux cases adjacentes en combinant les arbres qu'elles contiennent, à condition que soient respectées la projectivité et les contraintes d'ordre locales. L'analyse réussit si, à la fin, seule une case de la pile est non vide et si celle-ci contient un arbre dont tous les nœuds sont noirs (Figure 9).

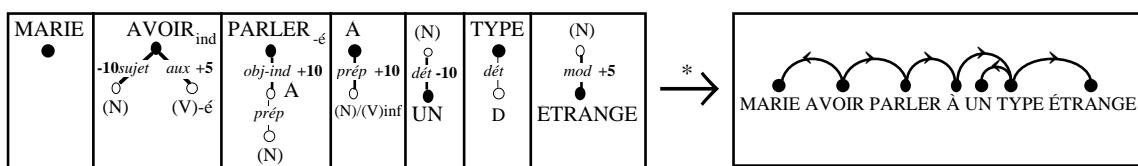


Figure 9. Schéma général de réduction de la pile

Cet « analyseur » s'apparente aux grammaires catégorielles. La grammaire de Nasr présente néanmoins à notre avis quelques avantages sur ces dernières : 1) elle construit directement une véritable représentation géométrique ; 2) elle n'est pas spécifiquement orientée vers l'analyse et peut donc être facilement utilisée pour la synthèse, car elle permet de combiner des éléments sans se soucier a priori de l'ordre linéaire ; 3) elle ne nécessite pas qu'un syntagme soit saturé pour le rattacher à son gouverneur (cf. Lecomte 1992 : 143, où le même argument est utilisé pour justifier un formalisme intermédiaire entre les CG classiques et une grammaire plus géométrique comme la nôtre). Nous poursuivrons la comparaison avec les CG lors de la présentation des extractions.

Nous ne discuterons pas ici de la mise en place d'une procédure réelle. Notons simplement que pour une analyse donnée l'ordre dans lequel les réductions sont faites n'a pas d'importance. On peut donc réaliser un analyseur incrémental qui privilégiera toujours la réduction avant l'empilement ou bien un analyseur du type CKY (Aho & Ullman 1972 : 314).

1.2.5 Une grammaire non lexicalisée équivalente

Notre formalisme permet de construire également une grammaire non lexicalisée, fortement équivalente à la grammaire lexicalisée que nous avons présentée. L'idée est de séparer tout ce qui peut l'être et de voir une structure élémentaire comme l'unification de structures encore plus élémentaires (une idée similaire est utilisée par Candito 1996, 1999 pour générer une grammaire TAG). Une position intermédiaire particulièrement intéressante est de considérer une grammaire dont les structures élémentaires contiennent exactement un élément noir. Une structure élémentaire dont l'élément noir est

un nœud sera dite **nodale** et une dont l'élément noir est un arc **sagittale**. Une structure élémentaire de la grammaire lexicalisée sera obtenue comme combinaison d'une structure nodale et de structures sagittales (Figure 10).

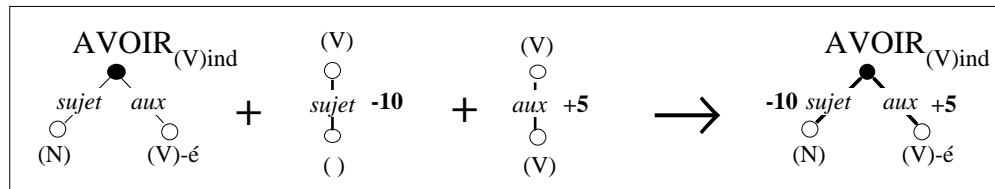


Figure 10. Une structure lexicalisée vue comme combinaison de structures nodale et sagittales

La partition en structures nodales et sagittales rejoint la partition de la grammaire (= modèle linguistique) en lexique et grammaire (au sens propre). Une structure nodale indiquera le régime de la lexie qui étiquette son ancre, tandis qu'une structure sagittale gérera les propriétés caractéristiques de la fonction syntaxique qui étiquette son arc noir, comme la linéarisation (= le poids) ou l'accord (par ex., la règle d'accord du verbe avec son sujet sera attachée à la règle sagittale *sujet*). Notons qu'une structure sagittale, bien qu'écrite dans un formalisme différent, est équivalente à une règle de correspondance Sens-Texte (Mel' čuk & Pertsov 1987).

1.2.6 Quasi-dépendance et “sujet” des formes verbales non finies

Nous allons introduire un enrichissement de notre formalisme, la quasi-dépendance, que nous illustrerons par le traitement du sujet des formes verbales non finies. Considérons :

(6) *Pierre commence à travailler.*

Nous considérons, comme le fait la TST, que dans une telle phrase PIERRE est le sujet de COMMENCER et qu'il n'y a pas de vraie dépendance syntaxique entre TRAVAILLER et PIERRE. Néanmoins, il y a une forme de « dépendance » entre le sujet de COMMENCER et le (V)inf et cette « dépendance » est bien syntaxique et non sémantique, puisque le sujet de COMMENCER et le (V)inf ne forme pas nécessairement deux unités sémantiques séparés, comme le montre les exemples suivants :

- (7) a. *Il commence à pleuvoir.*
 b. *La moutarde commence à me monter au nez.*

En (7)a, IL est un élément purement syntaxique qui n'a pas de correspondant sémantique, et en (7)b, le phrasème LA MOUTARDE MONTER AU NEZ correspond à un unique nœud sémantique.

Nous allons donc introduire dans la structure élémentaire d'un (V)inf une dépendance syntaxique *sujet* d'un type particulier, que nous appellerons une **quasi-dépendance**. Les quasi-dépendances se distinguent des dépendances par le fait qu'elles n'interviennent pas dans la linéarisation (cf. la solution proposée par Hudson 2000) : en conséquence, elles ne reçoivent pas de poids et ne sont pas prise en compte par la projectivité. Par contre, elles

reçoivent une étiquette fonctionnelle et un type. Formellement, dépendances et quasi-dépendances seront simplement différenciées par la valeur – ou + d'un trait quasi. Les quasi-dépendances seront représentées par une flèche discontinue.

La structure élémentaire d'un **verbe à montée**¹¹ comme COMMENCER indiquera que son sujet est aussi le "sujet" du (V)inf complément par une quasi-dépendance *sujet* blanche, qui s'unifiera avec la quasi-dépendance *sujet* noire du (V)inf, comme cela se ferait pour une véritable dépendance.

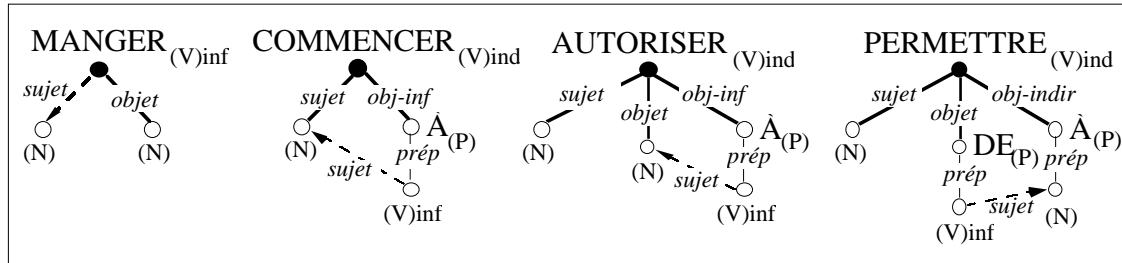


Figure 11. Structures élémentaires d'un verbe infinitif et de verbes à montée

Les **constructions attributives** (*ce livre est étrange, Max trouve ce livre étrange*) s'apparentent aux constructions à verbe à montée : le premier argument sémantique de l'adjectif, qui est le nom modifié lorsque l'adjectif est épithète (*un livre étrange*), est un complément du verbe copule. On a tout intérêt à maintenir entre l'adjectif et ce complément une quasi-dépendance, pour récupérer la dépendance sémantique ou pour assurer l'accord de l'adjectif sans qu'il soit nécessaire de passer par le verbe copule et donc d'y faire figurer des traits d'accord (comme le genre). Il est alors possible d'utiliser pour l'adjectif attribut la même structure que pour l'adjectif épithète. Il faudra autoriser une dépendance, ici modif(icative), à s'unifier avec une quasi-dépendance pour donner une quasi-dépendance. Les participes passés passifs (V)-é seront traités de la même façon que les adjectifs.

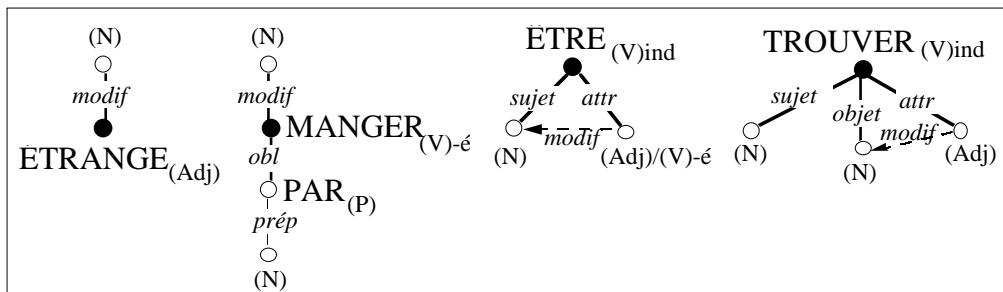


Figure 12. Structures élémentaires d'un Adj, d'un (V)-é et de verbes à copule

¹¹ Nous appelons verbe à montée un verbe L ayant pour complément un (V)inf dont le "sujet" est un autre complément de L (cf. Figure 11). On distingue parmi les verbes à montée en position sujet ceux qui contrôlent leur sujet (comme VOULOIR) et ceux qui ne le contrôlent pas (comme COMMENCER) (cf. Figure 13). Un argument syntaxique est dit **contrôlé** s'il correspond à un argument sémantique.

1.3 Le module sémantique

Comme nous l'avons dit, nous limitons la représentation sémantique à un graphe hiérarchisé indiquant les dépendances sémantiques (= relations prédicat-argument) et les dépendances communicatives. Les règles du module sémantique sont de deux types : les règles lexicales pour l'expression des lexies profondes et les règles grammaticales pour l'expression des grammèmes profonds.

Une **règle sémantique lexicale** indique la correspondance entre une lexie profonde et une configuration de lexies de surface. La correspondance entre actants sémantiques et nœud syntaxiques est indiquée par les indices x , y , etc. Les grammèmes profonds qui seront traduits par les règles grammaticales sont simplement reportés dans la configuration syntaxique (partie droite) et précédés d'une flèche (\rightarrow) indiquant qu'ils doivent être traduits (voir ci-dessous) ; pour un verbe français, il s'agit du m(ode), du t(emps) et de la v(oix). Comme en TAG, nous indiquons par un trait spécial, ici $*$, les nœuds qui ne peuvent recevoir d'adjonction.

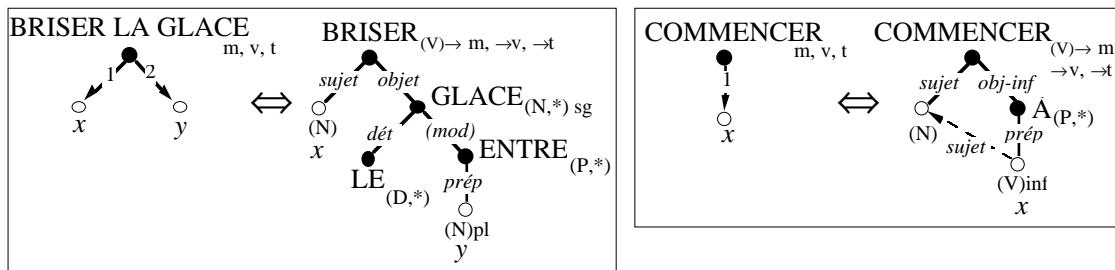


Figure 13. Deux règles sémantiques lexicales

Une **règle sémantique grammaticale** traduit un grammème profond en un grammème de surface (dans ce cas la règle est généralement triviale) ou en une configuration plus complexe. Par exemple, pour l'expression du mode infinitif, il faut indiquer que le grammème profond inf est exprimé par le grammème de surface inf et que le sujet devient une quasi-dépendance. Pour l'expression de la voix passive, il faut indiquer que le grammème profond est exprimé par **ÊTRE** + -é, que le nœud *objet* devient *sujet* et que le nœud *sujet* devient un complément *obl(ique)* optionnel en PAR ou DE.

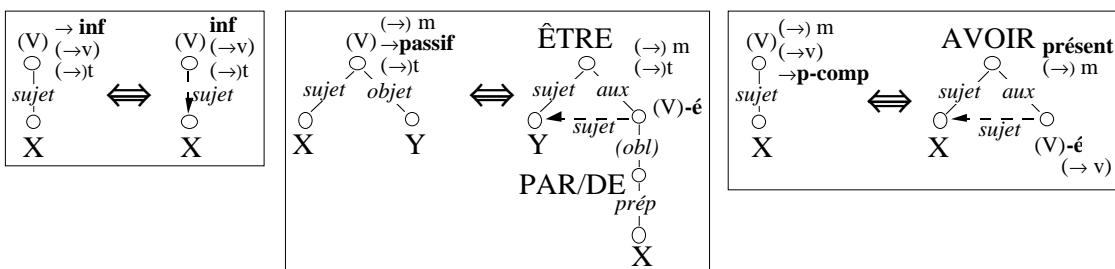


Figure 14. Trois règles sémantiques grammaticales (mode infinitif, voix passive, temps passé composé)

Remarquons que les règles sémantiques grammaticales ne sont pas des règles de correspondance entre une configuration sémantique et une configuration syntaxique mais entre des configurations syntaxiques. L'ordre dans lequel les règles sont appliquées n'a pas d'importance ; c'est la raison pour laquelle les → figurent entre parenthèses, selon que le grammème profond correspondant a déjà été traduit ou non.

2. UNE GRAMMAIRE A BULLES

Avant de proposer une grammaire à bulles, nous allons présenter, suivant Kahane 1997, la notion d'arbre à bulles et son utilisation dans la représentation des extractions.

2.1 Définition des arbres à bulles et projectivité

L'utilisation d'arbres à bulles remonte au moins à Tesnière 1959 (1967 : 46), dont le stemma comprend des bulles appelées nucléus où sont regroupés, pour des raisons différentes des nôtres, des éléments comme l'auxiliaire et le participe, la copule et l'adjectif ou l'article et le nom. La première étude formelle d'une structure de type arbres à bulles se trouve dans Gladkij 1968. Notre présentation reprend Kahane 1997 avec quelques modifications et précisions.

Intuitivement, un **arbre à bulles** est un arbre dont certains nœuds sont occupés par des bulles contenant elles-mêmes des nœuds pouvant à leur tour être occupés par des bulles ou avoir des dépendants. Formellement, un arbre à bulles est un quadruplet $(X, B, \rightarrow, \phi)$ où X est l'ensemble des **nœuds de base** (correspondant aux nœuds lexicaux dans le cas d'une représentation syntaxique), B est l'ensemble des **bulles**, ϕ est une application qui associe à chaque bulle son **contenu** (qui est une partie non vide de X) et \rightarrow est une **relation de dépendance** sur B (définissant un arbre sur B). Nous imposons que, pour tout nœud de base x , il y ait une bulle de contenu $\{x\}$ (la bulle la plus emboîtée de contenu $\{x\}$ sera identifiée avec le nœud de base x). Nous imposons de plus que deux bulles soient toujours disjointes ou incluses l'une dans l'autre¹², ce qui nous donne deux situations pour la dépendance entre deux bulles: la **dépendance pure** et l'**emboîtement**.

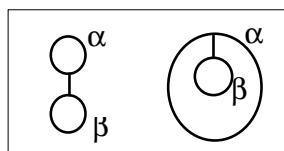


Figure 15. Une dépendance pure et un emboîtement (pour $\alpha \rightarrow \beta$)

¹² On impose en fait une propriété plus forte, dont nous n'avons pas de formulation simple : si deux bulles α et β ont au moins un nœud en commun, alors l'une des deux est dominée par l'autre, par exemple $\alpha \rightarrow^* \beta$, et toutes les bulles de la chaîne entre α et β sont incluses dans α . Cela signifie que, si une chaîne de dépendance sort d'une bulle, elle ne peut y revenir.

Nous appelons **projection** d'une bulle α la réunion des contenus des bulles dominées par α , α compris, pour la relation de domination \rightarrow^* obtenue par transitivation de \rightarrow . Un arbre à bulle est dit **linéairement ordonné** s'il est assorti d'un ordre linéaire sur ses nœuds de base. Un arbre à bulles linéairement ordonné est dit **projectif** si la projection de toute bulle forme un segment continu. Comme dans le cas des vrais arbres, la projectivité peut être exprimée plus géométriquement ; elle est équivalente à la conjonction des trois conditions suivantes : (1) deux bulles ou un arc et une bulle ou deux arcs ne se coupent jamais, (2) une bulle ou un arc ne couvre jamais la racine de l'arbre et (3) un dépendant d'une bulle α ne se trouve jamais au-delà d'un dépendant d'une bulle β contenant α (Kahane 1997; les conditions (1) et (2) sont énoncées dans Gladkij 1968) (Figure 16).

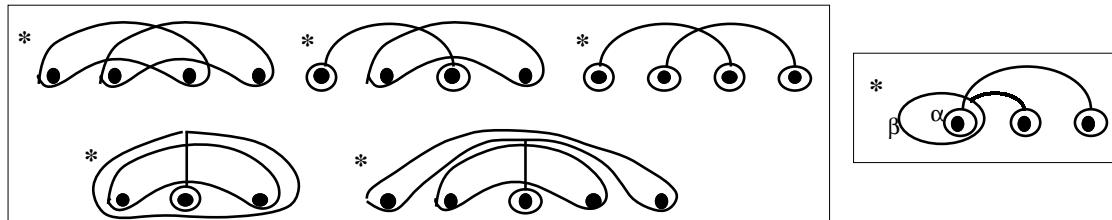


Figure 16. Les conditions (1) et (3) de non-projectivité pour un arbre à bulles

Il peut être utile d'affaiblir la notion de projectivité. Nous utiliserons pour décrire l'inversion du sujet dans les extractions une projectivité faible sans la Condition 3. Par contre la condition (3) est utilisée de façon essentielle dans la description de la coordination (Kahane 1997).

2.2 Représentation de l'extraction

Certains groupes de mots, que nous appellerons, suivant Tesnière 1959, des **nucléus**, peuvent se comporter comme un seul mot du point de vue de certains phénomènes syntaxiques comme l'extraction, mais aussi la cliticisation, la coordination, la négation ou la comparaison. Nous distinguerons deux familles de nucléus (cf. Kahane & Mel'čuk 1999 pour une déf. précise):

- un **nucléus verbal** est une chaîne de verbes (ou de tournures équivalentes) se comportant comme un seul verbe: *un livre que Pierre lit, a commencé à lire, a envie de lire, a l'air de trouver facile à lire, pense que Marie devrait lire, ...*¹³
- un **nucléus nominal** est une chaîne de noms (prépositions et déterminants compris) se comportant comme un seul nom: *une dame sur laquelle tu comptes, sur le mari de laquelle tu comptes; je me demande le mari de quelle dame tu as rencontré.*

¹³ Le traitement de l'extraction des compléments de nom des sujet et objet direct (*la dame dont le mari est mort; la dame dont je connais le mari; *la dame dont je parle au mari*) nécessite de considérer qu'un verbe et son sujet ou un verbe et son complément d'objet direct peuvent également former un nucléus verbal.

La nature du nucléus dépend de la langue (par ex. en anglais, à la différence du français, une préposition peut terminer un nucléus verbal (preposition stranding): *the girl you are talking to*; **la fille que tu parles à*) et du phénomène considéré (les chaînes V que V, possibles dans les extractions ou la coordination elliptique, sont plus difficiles pour la négation: *la personne à qui Jean veut que tu parles*; *Jean veut qu'on appelle la police et Marie les pompiers*; ?*Jean ne veut que tu parles à personne*).

Dans la suite, nous réservons l'étiquette vV (resp. vN) pour désigner un nucléus verbal (resp. nominal).

Un nucléus forme une portion connexe de l'arbre de dépendance (en fait une chaîne dans tous nos exemples).¹⁴ On peut donc représenter les nucléus par des bulles¹⁵ et attribuer, sous certaines conditions, des dépendances à ces bulles. Nous présentons Figure 17 et Figure 18 les représentations de (1) par l'arbre de dépendance proposé par la TST (Mel'čuk & Pertsov 1987, Kahane & Mel'čuk 1999) et par l'arbre à bulles qui en découle directement. Dans la suite, nous proposerons un autre arbre à bulles pour (1).

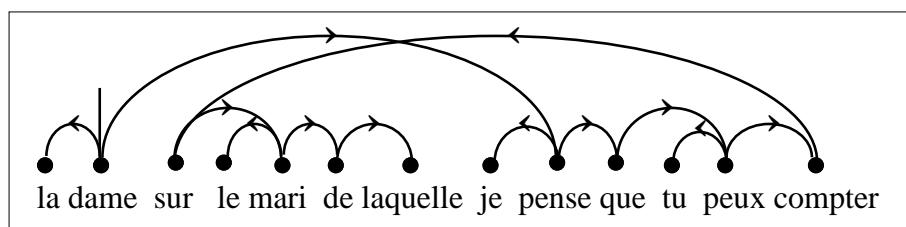


Figure 17. Un arbre de dépendance ordinaire (non projectif) pour (1)

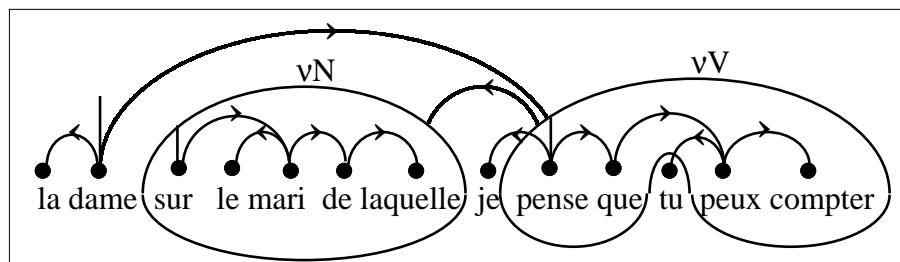


Figure 18. Un arbre à bulles (projectif) pour (1)

La considération des nucléus, et leur représentation par des bulles, a deux avantages majeurs.

¹⁴ Par contre, un nucléus ne forme pas a priori une portion connexe de la chaîne linéaire (cf. le nucléus verbal *pense* → *que* → *peux* → *compter* Figure 17 et Figure 18).

¹⁵ Nous distinguons clairement nucléus et bulles. La bulle est un objet formel qui permet de représenter l'objet linguistique nucléus, mais aussi d'autres objets linguistiques comme la chaîne coordinative (cf. Kahane 1997).

- Elle permet de décrire simplement les phénomènes où interviennent des nucléus. Par exemple, la relativisation peut être décrite ainsi : 1) le pronom relatif doit appartenir à un nucléus nominal gouverné par un nucléus verbal, lui-même gouverné par l'antécédent de la relative ; 2) le nucléus nominal est placé à l'avant de la relative.
- Alors que les phrases à extraction possèdent généralement un arbre non projectif, leur arbre à bulles est projectif. La projectivité de l'arbre à bulles permet d'assurer la linéarisation correcte par des règles de linéarisation *locales*, c'est-à-dire, comme dans le cas des arbres de dépendance, simplement par des poids sur les dépendances.

2.3 Une grammaire à bulles pour l'extraction

Nous allons proposer une grammaire pour la relativisation basée sur la représentation de la TST, puis nous argumenterons pour une représentation plus complexe à la Tesnière-Hudson et proposerons une grammaire en conséquence avant de l'étendre aux autres phénomènes d'extraction.

2.3.1 La question des éléments en position non canonique

Certains mots, comme les pronoms clitiques ou relatifs, peuvent occuper des positions argumentales, mais ne pas se trouver en position canonique. Dans une grammaire lexicalisée, le placement d'un élément est forcément assuré par la structure élémentaire d'une lexie. Dans notre grammaire, les éléments en position canonique sont placés par la structure élémentaire de leur gouverneur (cf. par ex. la structure de AVOIR Figure 8), mais nous pensons qu'il est préférable que les éléments en position non canonique soit placés par leur propre structure élémentaire (Kahane 2000). Certains formalismes, comme TAG, ne permettent pas qu'il en soit ainsi : tous les éléments sous-catégorisés sont placés dans la structure élémentaire de leur gouverneur, ce qui fait que, avec le croisement des phénomènes, un verbe triactanciel peut recevoir plusieurs centaines de structures élémentaires (Abeillé 1991, XTAG Group 1995, Candito 1999).

2.3.2 Structures élémentaires pour les relatives (première version)

La première question qui se pose est de savoir quel est l'élément de la relative qui va assurer le rattachement de la relative à son antécédent. Plusieurs solutions sont possibles. Par exemple, en TAG, dans l'analyse usuelle (Kroch 1987), c'est le verbe qui s'adjoint sur l'antécédent, mais il est possible de faire ce rattachement par le biais du pronom relatif (Kahane, Candito & De Kercadio 2000). C'est également l'analyse des CG, où le pronom relatif s'attache à la fois à l'antécédent et à la relative (cf. la catégorie $(N|N)/(S \uparrow NP)$ du pronom relatif WHICH dans Moortgat (1989: 113) signifiant que WHICH s'adjoint à un N et requiert un S auquel manque un NP).

Dans l'analyse proposée ici, non seulement le pronom relatif assure le rattachement à l'antécédent, mais c'est même lui qui contrôle la forme de la relative. Ainsi, la structure élémentaire d'un pronom relatif indique-t-elle que celui-ci appartient à un nucléus vN dépendant d'un nucléus vV dépendant lui-

même de l'antécédent. Plus précisément, c'est le pronom relatif qui introduit dans la structure de représentation les bulles vV et vN, ainsi que la dépendance entre l'antécédent et la relative (qui sont donc en noir). Un nœud est également prévu dans la bulle vV pour recevoir le verbe principal de la relative, qui doit être à l'indicatif ((V)ind). (Le fait qu'il s'agit du verbe principal est indiqué par la dépendance avec vV.)

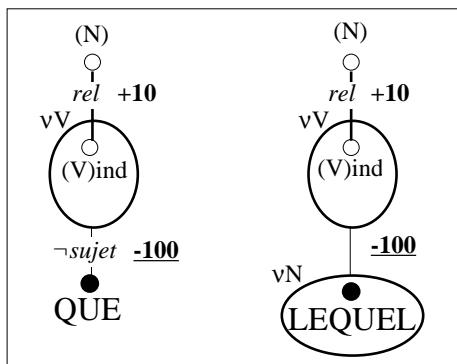


Figure 19. Structures élémentaires pour les pronoms relatifs QUE et LEQUEL (première version)

Il faut noter que c'est le pronom relatif qui assure le placement antéposé du nucléus vN par l'intermédiaire du poids prioritaire -100. Une **valeur prioritaire**, indiquée par un soulignement, vient se substituer aux autres valeurs au lieu de s'unifier avec.

Les pronoms QUE et LEQUEL se distinguent par le fait que QUE est directement lié au nucléus vV, alors que LEQUEL appartient à un nucléus vN lié au nucléus vV (ce qui traduit le caractère plus complémenteur de QUE par rapport à LEQUEL). La fonction syntaxique de la dépendance entre vV et vN est libre ; elle est indirectement contrôlée par la nature des nœuds qui se trouve dans vV et vN (cf. Figure 20).

2.3.3 Opération de combinaison avec les nucléus

Les opérations de combinaison avec nucléus reposent sur un marquage explicite des dépendances qui peuvent ou non figurer dans une bulle vV ou vN par un trait nucleus. Une dépendance qui peut être à l'intérieur d'un nucléus v recevra la valeur nucleus : iv. Une dépendance qui peut être attribuée à un nucléus v recevra la valeur nucleus : ev.

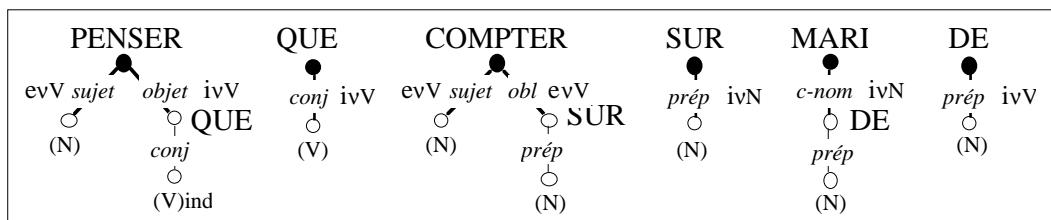


Figure 20. Structures élémentaires avec étiquetage des dépendances intra- et extranucléaires

Ce procédé permet d'encoder facilement les contraintes lexicales sur l'extraction : ainsi les verbes non ponts, comme *s'écrier que*, n'auront pas d'étiquette *ivV* sur la dépendance entre le verbe et la complétive (*le livre que Pierre a dit qu'il a lu* ; **le livre que Pierre s'est écrité qu'il a lu*).

Venons-en aux opérations de combinaison proprement dites. Lorsqu'un nœud A' fusionne avec un nœud A d'une bulle v , les nœuds reliés à A' sont placés par défaut à l'extérieur de v (Opération R1, Figure 21). Mais l'appartenance à v peut être propagée à tout nœud B' relié à A' par un lien intranucléaire *iv* (Opération R2, Figure 22). Enfin tout nœud B' relié à A' par un lien extranucléaire *ev* peut être promu sur le nucléus v (Opération R3, Figure 23).

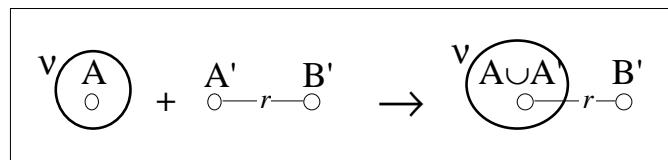


Figure 21. R1 : Placement par défaut d'un dépendant d'un nœud d'un nucléus v

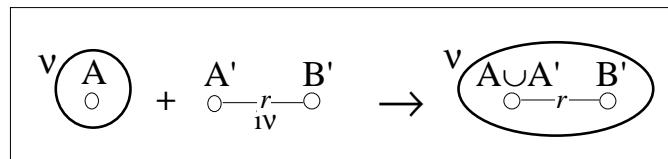


Figure 22. R2 : Propagation de l'appartenance à un nucléus v

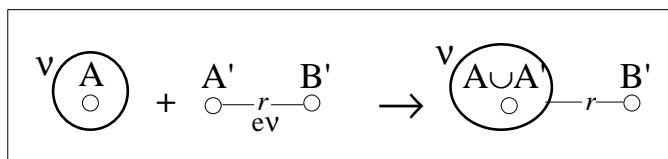


Figure 23. R3 : Promotion d'un dépendant sur un nucléus v

Dans notre représentation par arbre à bulles, nous perdons le fait que la dépendance promue sur le nucléus par l'opération R3 est à l'origine issue de A' . On peut facilement garder cette information par un enrichissement du formalisme de représentation en reliant toute dépendance d'un nucléus à un nœud intérieur. Nous y reviendrons aux §§ 2.3.5 et 2.3.7.

Les opérations R1 et R3 ne sont possibles que si B' est un dépendant de A' . Si B' est le gouverneur de A' et que A' fusionne avec un nœud A d'un nucléus v , alors soit B' est placé dans le nucléus v (et c'est l'opération R2 qui est déclenchée), soit B' devient le gouverneur de v et A' devient le sommet de v (Opération R4, Figure 24). Notons que le lien d'étiquette *r* entre B' et A' donne deux liens dans ce dernier cas : un lien d'étiquette *r* entre B' et v et un lien sans étiquette entre v et $A ∪ A'$ indiquant simplement que le nœud $A ∪ A'$ résultant de la fusion des nœuds A et A' est le sommet du nucléus v .

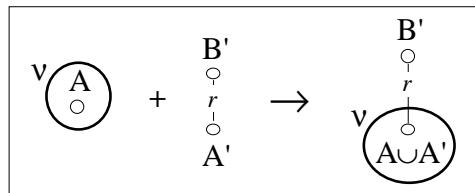


Figure 24. R4 : Placement par défaut du gouverneur d'un nœud d'un nucléus v

Avant de commenter la portée des opérations précédentes, nous allons les illustrer par l'analyse incrémentale de (1). Pour alléger nos figures, nous nous autoriserons dans la suite à étiqueter les nœuds par des mots, plutôt que par leur représentation morphologique (lexie + grammèmes).

Pour chaque mot du syntagme, une structure élémentaire est chargée (Figure 25). Le curseur situé à gauche se déplace vers la droite. A chaque fois qu'une nouvelle case est considérée, la réduction avec la case précédente est envisagée. Si cette première réduction aboutit, la réduction avec la case encore précédente est envisagée et ainsi de suite.

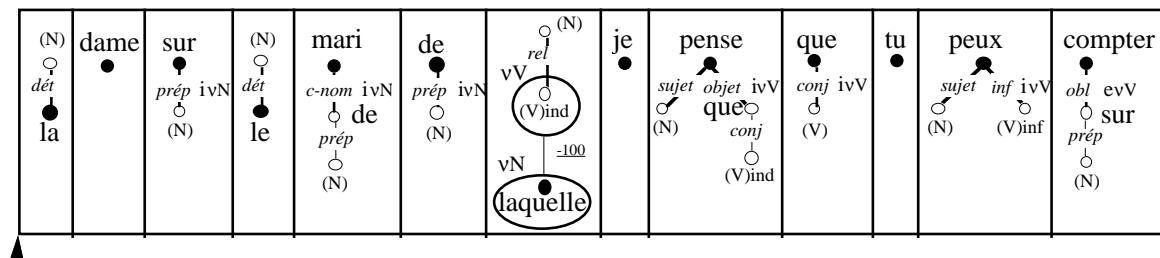


Figure 25. Etape 1: la pile avant la lecture du 1^{er} mot *la*

Lorsque le pronom relatif *laquelle* s'apprête à être considéré, le segment initial de la pile est réduit à deux cases contenant respectivement *la dame* et *sur le mari de (N)* (Figure 26).

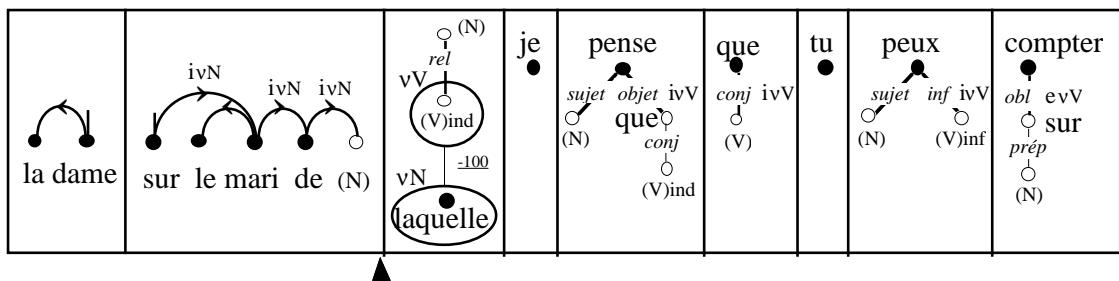


Figure 26. Etape 7: la pile avant la lecture du 7^{ème} mot *laquelle*

Le nœud ancré par *laquelle* peut fusionner avec la requête de *sur le mari de (N)*. Comme tous les liens de la chaîne *sur le mari de (N)* sont de type ivN, l'appartenance au nucléus vN se propage à l'ensemble de la chaîne (R2). Les cases occupées par *la dame* et *sur le mari de laquelle* pourront ensuite fusionner (Figure 27).

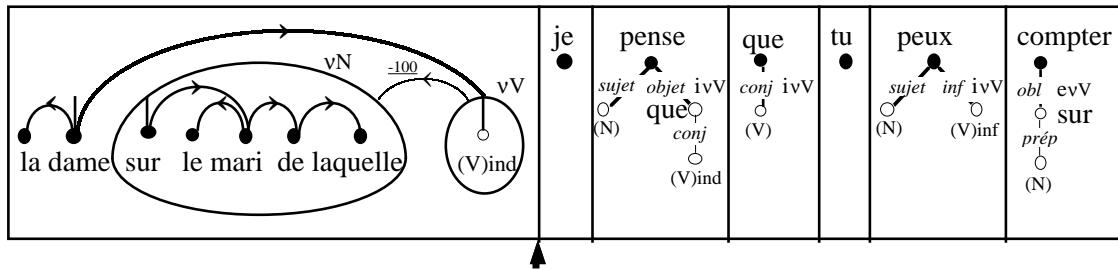


Figure 27. Etape 8: la pile avant la lecture du 8^{ème} mot *je*

Le verbe *pense* pourra, dès qu'il est considéré, venir fusionner avec le nœud (V)ind de la bulle vV et, par propagation, la requête *que* (V) de *pense* sera incluse dans la bulle vV (R2).

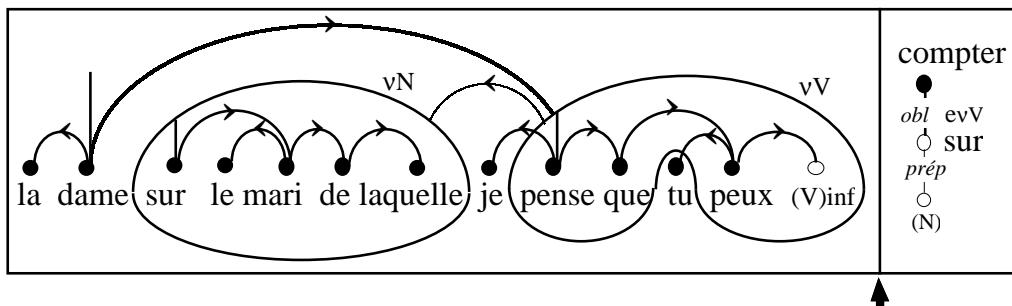


Figure 28. Etape 13 : la pile avant la lecture du dernier mot *compter*

Lorsque *compter* sera considéré (Figure 28), il pourra fusionner avec la requête (V)inf de la bulle vV (de la case précédente), tandis que sa propre requête (*sur* (N)) fusionnera avec le nœud *sur* de la bulle vN et le lien entre les deux nucléus. Il s'agit donc d'une double réduction faisant intervenir simultanément les opérations R3 et R4 : d'une part, la requête de *compter* est promu sur le nucléus vV (R3), d'autre part, cette requête vient gouverner le nœud *sur* du nucléus vN, qui en devient le sommet (R4). L'analyse du syntagme (1) a maintenant réussi. Il a fallu en particulier pour cela que le lien entre les deux nucléus devienne noir, c'est-à-dire que l'un des verbes du nucléus vV (à savoir *compter*) prenne un élément du nucléus vN comme dépendant.

Dans notre description des extractions, le groupe extrait, représenté par un nucléus vN, ne dépend pas de son gouverneur syntaxique, mais de la bulle vV. (Nous appelons gouverneur syntaxique de vN le nœud du nucléus vV qui a introduit la dépendance entre vV et vN et qui est le gouverneur dans la représentation par un arbre de dépendance ordinaire). Or la projection du nucléus vV est l'ensemble de la proposition relative et le nucléus vV domine directement le nucléus vN et le verbe principal de la relative : notre représentation est donc tout à fait comparable à la représentation de la Syntaxe X' (Chomsky 1977, Jackendoff 1977), où la proposition relative correspond à un nœud S' dans la projection duquel on trouve un nœud COMP pour le groupe extrait et un nœud S pour le reste de la relative. Par ailleurs, nos opérations peuvent être comparées aux opérations considérées par les

grammaires basées sur la Syntaxe X', comme HPSG (Pollard & Sag 1994) : l'opération R3 correspond à l'introduction d'un trait non local (ou trait Slash) (en étant promu sur le nucléus, le lien devient en quelque sorte non local), tandis que l'opération R2 correspond à la propagation de ce trait. Plusieurs descriptions proposées dans le cadre des grammaires de dépendance (Hudson 2000, Lombardo & Lesmo 1998, Kahane, Nasr & Rambow 1998) utilisent un appareillage équivalent au trait Slash de HPSG, permettant de remonter le groupe extrait sur le verbe principal de la proposition où a lieu l'extraction. Notre analyse diffère de ces analyses sur au moins un point : nous ne propagons pas le contenu du nucléus vN (ou du groupe extrait) ; nous l'attachons simplement au nucléus vV et c'est l'appartenance au nucléus vV qui se propage. En particulier, le nucléus extrait ne remonte pas toute la chaîne du nucléus vV en "passant" par chaque élément comme c'est le cas dans les traitements par un trait non local. Une idée similaire pourrait être utilisée en CG pour assurer les contraintes sur l'extraction en imposant lors de l'application de la Composition Fonctionnelle $A/B.B/C \rightarrow A/C$ que si A/B est de type ivV et B/C de type ivV (resp. evV), alors A/C soit de type ivV (resp. evV) et en assignant aux pronoms relatifs des catégories du type $(N/N)/(S/X)$ où S/X est de type evV .

2.3.4 Représentation sémantique et graphe de dérivation

De la même façon que cela est fait en TAG, on peut associer à l'analyse d'une phrase par notre grammaire un **graphe (hiérarchisé) de dérivation** (en TAG, il s'agit d'un arbre). Chaque nœud du graphe de dérivation représente une structure élémentaire et chaque lien la combinaison de deux structures élémentaires. Chaque lien est dessiné avec son gouverneur syntaxique en haut et son dépendant syntaxique en bas. De plus, chaque lien est orienté par une flèche qui indique laquelle des deux structures a introduit (en noir) la dépendance syntaxique qui les unit. En conséquence, une flèche vers le bas correspond à une substitution et une flèche vers le haut à une adjonction.

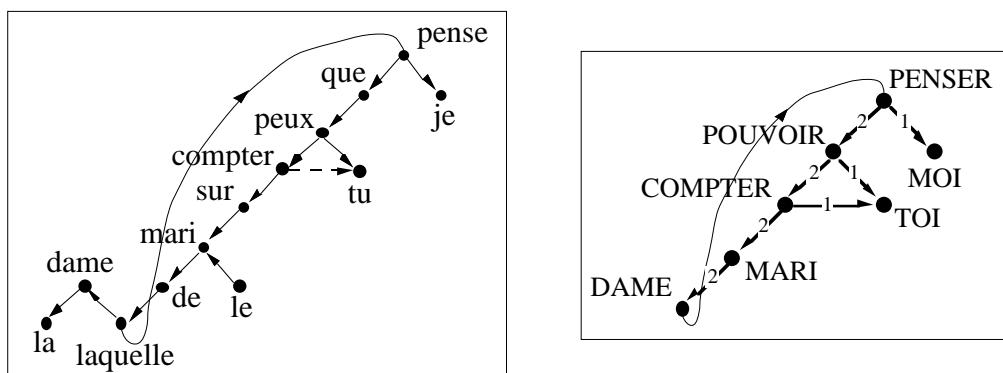


Figure 29. Le graphe de dérivation¹⁶ et la représentation sémantique (sans grammèmes) de (1)

¹⁶ Nous avons indiqué par un trait discontinu le lien correspondant à la quasi-dépendance *sujet* entre *compter* et *tu*, que nous n'avons pas représenté dans l'analyse pas à pas de (1) pour ne pas la surcharger inutilement (voir §1.2.6).

On note que le pronom relatif *laquelle* possède deux gouverneurs: il s'adjoint à l'antécédent *dame* et se substitue dans *[sur le mari] de*. Quant à la flèche de *laquelle* à *pense*, elle correspond à la substitution de *pense* dans la structure élémentaire de *laquelle* (il s'agit d'une flèche vers le bas, bien que *pense* soit placé au-dessus de *laquelle*, puisqu'il domine par ailleurs *laquelle*). Le graphe de dérivation n'est pas sans rappeler le stemma de Tesnière (1959 : 560), où les trois mêmes dépendances sont considérées pour un pronom relatif. Un graphe de dérivation assez similaire est obtenu dans un formalisme à la TAG proposé par Candito & Kahane 1998b.

Le graphe de dérivation est très proche de la représentation sémantique du syntagme : en effet, on passe du graphe de dérivation simplement en agrégeant avec leur voisin les nœuds qui correspondent à la même unité sémantique (c'est-à-dire *que* avec *pense*, *sur* avec *compter*, *de* avec *mari*, ainsi que *laquelle* avec *dame*). Autrement dit, le module syntaxique prépare parfaitement le terrain pour le module sémantique, qui se contentera d'effectuer les agrégations en question et la correspondance entre les dépendances syntaxiques et les dépendances sémantiques et communicatives. Tel que nous avons présenté le module sémantique au §1.3, celui-ci prend comme entrée, dans le sens de l'analyse, la représentation syntaxique. En fait, le module sémantique devrait plutôt s'appuyer sur le graphe de dérivation (à condition d'y encoder toutes les informations syntaxiques nécessaires). Ce point reste à l'étude et nous ne développerons pas davantage les règles sémantiques pour les extractions ici.

Dans les paragraphes qui suivent, nous allons revenir sur les structures élémentaires des pronoms relatifs.

2.3.5 Le pronom relatif comme sommet de la relative

Comme l'a très bien expliqué Tesnière 1959, le pronom relatif permet à un verbe fini de modifier un nom, c'est-à-dire d'occuper une position adjectivale. Ainsi le pronom relatif translate le verbe en adjectif¹⁷, c'est-à-dire qu'il sert d'écran au verbe par rapport au nom gouverneur qui attend un adjectif et doit pour cela s'interposer entre le verbe et le nom modifié. Il est alors, plus simple, comme le propose Tesnière 1959 : 560 de traiter le pronom relatif comme le sommet de la relative. Cette analyse est confortée par la description de l'alternance *que/qui* dans l'extraction du sujet d'une complétive (Kayne 1975):

- (8) a. *la personne que tu connais/qui te connaît*
 b. *la personne que je pense que tu connais/que je pense qui te connaît*
 c. *Pierre sait qui tu connais/qui te connaît*
 d. *Pierre sait qui je pense que tu connais/qui je pense qui te connaît*

¹⁷ Il existe d'autre translation du verbe en adjectif : les morphèmes du participe passé passif et du participe présent (*le livre regardé par Pierre* ; *la personne regardant ce livre*). Par contre, l'infinitif, comme la conjonction de subordination *que*, translate le verbe en nom et lui permet d'occuper des positions argumentales du verbe (*Pierre souhaite le départ de Marie/partir/que Marie parte*).

L'analyse la plus simple de ce phénomène repose sur un rapprochement entre "pronoms" relatifs et conjonctions de subordination, consistant à dire que dans les deux cas, *que* et *qui* sont des formes d'une même lexie QU- : lorsque QU- introduit une proposition sans sujet, celui-ci marque cette absence et prend la forme *qui*, alors que lorsqu'il introduit une proposition avec un sujet, il prend la forme non marquée *que*. Pour rendre compte de cela, nous introduisons un trait binaire *suj*(*et*) sur chaque verbe fini, indiquant s'il porte ou non un sujet, et un trait nom(inatif) sur QU- tel que $QU_{\text{nom:}+} = \text{qui}$ et $QU_{\text{nom:}-} = \text{que}$. De plus, le trait nom de QU- et le trait *suj* du V qu'il introduit doivent avoir des valeurs opposées, que nous notons x et \bar{x} , où x est une variable égale à + ou – (Figure 30).

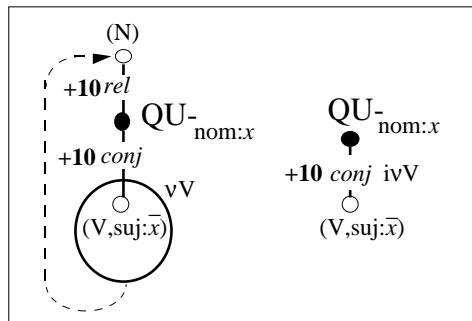


Figure 30. Structures élémentaires pour le "pronom" relatif QU- et la conjonction de subordination QU-

Nous considérons donc que, à l'instar de la conjonction de subordination, le "pronom" relatif QU- gouverne le verbe qu'il translate et est le sommet de la relative. Il s'en suit que la dépendance attribuée au nucléus vV, qui assure qu'une position est bien extraite, n'est plus une dépendance syntaxique, mais une quasi-dépendance, comme le lien *sujet* dans les structures des verbes à montée ou copule. La valeur du trait quasi de cette quasi-dépendance est prioritaire, puisqu'elle devra s'unifier avec une vraie dépendance et rester une quasi-dépendance après l'unification. On peut faire pointer cette quasi-dépendance sur QU- ou sur l'antécédent ; les deux analyses sont possibles. Nous garderons la deuxième : dans ce cas, QU- n'est plus du tout un pronom et il joue uniquement le rôle de complémenteur dans la subordonnée.

Dans cette représentation, le placement du nucléus vV par rapport au pronom relatif QU- est effectué, comme dans le cas de la conjonction de subordination QU-, par la dépendance *conj(onctive)* entre QU- et lui.

Les deux structures proposées pour le "pronom" relatif QU- (la version du §0 et la version présente) requiert exactement les mêmes opérations de combinaison ; en particulier, dans les deux cas, l'attribution d'une dépendance au nucléus vV est assurée par l'opération R3. D'ailleurs dans les deux cas, les graphes de dérivation sont quasiment les mêmes ; ils seraient même exactement les mêmes si la quasi-dépendance de la structure de QU- pointait sur le nœud QU-.

EXTRACTIONS DANS UNE GRAMMAIRE DE DÉPENDANCE

Il nous reste à indiquer le fonctionnement du trait *suj*. Celui-ci possède, par défaut, la valeur *suj:+* pour un verbe fini. Lorsque le sujet est extrait (par l'opération R3) et qu'il devient une quasi-dépendance sur le nucléus *vV*, le trait *suj* doit prendre la valeur *suj:-* ; l'opération R3 doit être modifiée en conséquence.

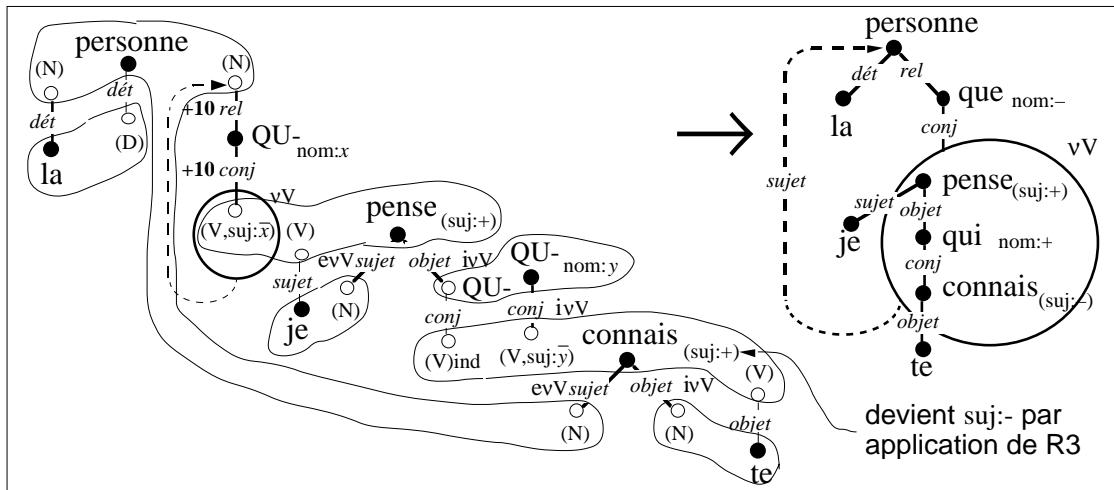


Figure 31. Dérivation de (8) : *la personne que je pense qui te connais*

Nous allons maintenant étendre notre analyse du pronom relatif comme sommet de la relative aux vrais pronoms relatifs. Pour cela, nous sommes obligés d'enrichir notre formalisme. Comme le pronom relatif doit aussi agir comme pronom et occuper une position dans la relative (en plus de sa position de sommet), nous allons donc associer au pronom relatif une **double position**. Chacune de ces positions sera représentée par un **demi-nœud** (Figure 32).

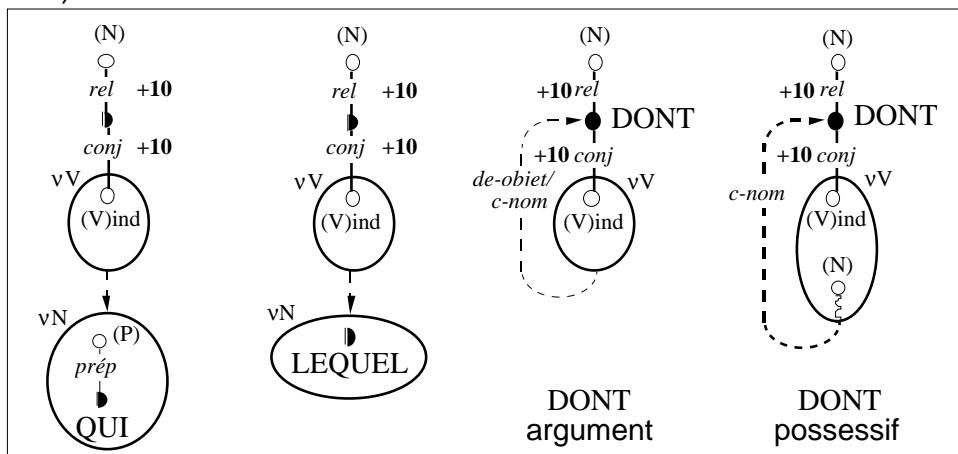


Figure 32. Structures élémentaires pour les pronoms relatifs QUI, LEQUEL et DONT

Remarquons également que la dépendance entre les nucléus *vV* et *vN* n'est pas utile pour la linéarisation, puisque le placement antéposé du pronom relatif est assuré par les liens du pronom avec l'antécédent et le nucléus *vV* et le placement du nucléus *vN* par la présence du pronom en son sein. Cette

dépendance peut donc être remplacée par une quasi-dépendance, ce qui nous ramène à une structure très proche de celle du “pronome” relatif QU-. Il faut néanmoins remarquer que cette structure n'est plus un arbre (à bulles). Une étude complémentaire serait nécessaire pour définir plus précisément la nature de telles structures. Le traitement des pronoms relatifs modifieurs, comme OÙ et DONT possessif (*la personne dont la voiture est mal garée*), ne semble pas pouvoir être effectué sans un nouvel enrichissement du formalisme. En effet, comme ces pronoms doivent s'ajointre sur le dernier nœud du nucléus vV, il nous faut un moyen de relier la dépendance promue sur le nucléus avec un nœud du nucléus, ce que nous indiquons par un lien en tortillon.

2.3.6 Les autres types d'extraction

Nous allons montrer comment étendre notre description de la relativisation à d'autres types d'extraction : l'interrogation indirecte (*Pierre veut savoir sur le mari de quelle dame je pense que tu peux compter*), l'interrogation directe (*Sur le mari de quelle dame penses-tu que je peux compter ?*), le clivage (*C'est sur le mari de cette dame que je pense que tu peux compter*) et la topicalisation (*Sur le mari de cette dame, je pense que tu peux compter*).

A l'instar de notre analyse des relatives, nous allons 1) assurer l'introduction des nucléus par l'élément qui permet l'extraction (le pronom interrogatif, le QU- du clivage ou la topicalisation) et 2) chercher à obtenir un graphe de dérivation proche d'un graphe sémantique. Nous allons commenter les structures élémentaires que nous proposons Figure 33.

Pour le clivage *c'est...que/qui*, nous considérons qu'il s'agit encore de la conjonction QU-. La différence essentielle avec la relativisation est que le complément extrait devient un argument du verbe ÊTRE et qu'il n'est pas contraint par la partie du discours (ce qu'indique l'étiquette vide ()). Pour une clivée telle que *C'est Pierre qui dort*, le complément *Pierre* est ainsi attr(ibut) de *est* et sujet de *dort* ; néanmoins cette dépendance sujet entre *Pierre* et *dort* doit être promue par l'opération R3 sur le nucléus vV où elle devient une quasi-dépendance ; *dort* prendra alors la valeur *suj:-* et par conséquent QU- aura la valeur *nom:+* et se réalisera par *qui*.

Les pronoms interrogatifs indirects recevront des structures élémentaires assez similaires aux pronoms relatifs, à la différence que ceux-ci ne s'ajointent pas, mais se substituent. Un étiquetage adéquat des pronoms interrogatifs devra assurer qu'ils soient utilisés seulement comme argument des verbes qui sous-catégorisent une proposition interrogative.

D'autre part, pour permettre l'analyse d'une double extraction du type *les pommes que je sais à qui donner ou le livre dont Pierre sait quelle page il a arrachée* (dans la structure desquelles on aura deux nucléus verbaux emboîtés l'un dans l'autre : un déclenché par le pronom relatif et l'autre par le pronom interrogatif), il est nécessaire qu'un pronom interrogatif puisse se trouver à l'intérieur d'un nucléus vV. Pour ce faire, il faut que la dépendance conj(onctive) entre le pronom interrogatif et le nucléus verbal qu'il régit soit étiquetée ivV.

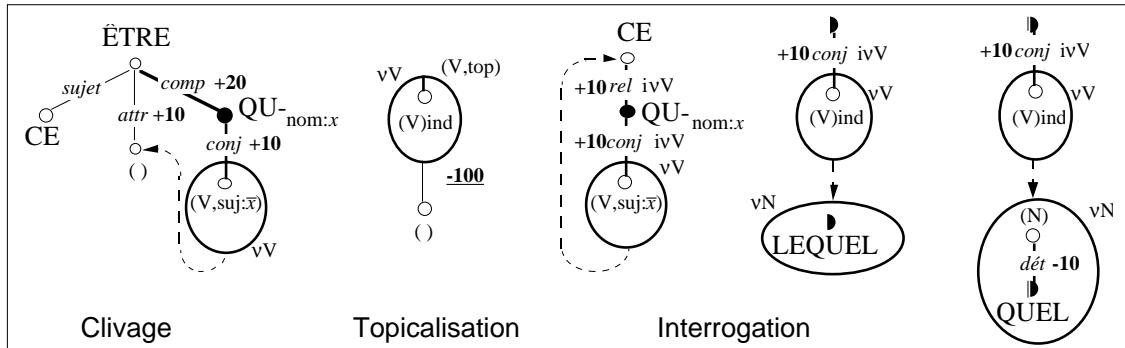


Figure 33. Structures élémentaires pour le clivage, la topicalisation et les pronoms interrogatifs *ce qui/que*, *lequel* et *quel*

Pour la topicalisation, on utilisera une structure élémentaire sans ancre lexicale, c'est-à-dire sans nœud noir. Cette structure introduit simplement une bulle vV et permet donc, par le biais de l'opération R3, à un complément d'un des verbes du nucléus vV d'être topicalisé.

Un nucléus vV dont un complément est topicalisé reçoit l'étiquette (V,top) = [cat:V, top:+, ...] et se comporte comme un nœud verbal ordinaire. Néanmoins, par le biais du trait top, on peut empêcher qu'un verbe dont un complément est topicalisé vienne occuper n'importe quel nœud verbal (par ex., il n'est pas très heureux qu'il y ait une topicalisation dans une relative : *?le livre qu'hier j'ai lu*).

2.3.7 Inversion du sujet

Les extractions favorisent l'inversion du sujet en français. Néanmoins, il faut remarquer que l'inversion doit se faire par rapport au nucléus entier :

- (9) a. *le livre que Pierre veut lire*
 b. *le livre que veut lire Pierre*
 c. **le livre que veut Pierre lire*

Nous proposons donc pour l'inversion du sujet une structure élémentaire non ancrée comportant un nucléus vV sur lequel est promu le sujet du verbe sommet de ce nucléus (Figure 34).¹⁸ Cette analyse ressemble à celle de la

¹⁸ L'inversion du sujet est plus contrainte que l'extraction. D'une part, elle est impossible si l'un des verbes de la chaîne nucléaire possède un complément d'objet nominal :

- (i) a. **la personne à qui a donné un cadeau Pierre*
 b. **la personne à qui menace Jean de parler Pierre*

mais elle possible si cet objet est extrait ou même cliticisé :

- (ii) a. *le cadeau qu'a reçu Pierre*
 b. *Le paquet a été rendu à la personne à qui l'avait offert Pierre.*

D'autre part, l'inversion du sujet est impossible si le nucléus contient une complétive, à moins que le sujet de la complétive soit clitique :

- (iii) a. **la personne à qui souhaite que Marie parle Pierre*
 b. ?*la personne à qui souhaite qu'on parle Pierre*

topicalisation, à la différence que le nucléus vV de la structure élémentaire pour l'inversion du sujet est blanc et doit s'unifier avec un nucléus déjà existant, bloquant ainsi l'inversion en dehors d'une extraction.

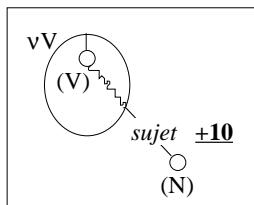


Figure 34. Structure élémentaire pour l'inversion du sujet

Pour terminer, il faut noter que le placement du sujet inversé est libre par rapport aux autres complément à droite du nucléus :

- (10) a. *le livre que Pierre veut donner à Marie*
 b. *le livre que veut donner Pierre à sa sœur*
 c. *le livre que veut donner à Marie son frère*

Nous résolvons ce problème en notant que la condition (3) de la projectivité ne s'applique pas au sujet inversé ; autrement dit celui-ci peut se placer librement par rapport aux compléments du dernier verbe du nucléus. Le fait de typer différemment les différentes dépendances d'un arbre à bulles peut être rapproché des CG multimodales (Hepple 1995) où les principes structuraux s'appliquent différemment aux différentes catégories.

CONCLUSION

Bien que notre grammaire se situe à la croisée de différents courants en linguistique formelle (notamment la TST, TAG, les CG et HPSG), nous pensons qu'elle permet une description originale des extractions. Cette description présente différents avantages. Premièrement, la représentation par bulles donne à l'outil utilisé pour traiter les dépendances une véritable existence géométrique (à la différence, par ex., des grammaires utilisant la propagation d'un trait non local). Deuxièmement, notre analyse peut être intégrée à une grammaire lexicalisée (bien que la lexicalisation ne soit pas une nécessité, cf. 1.2.5) sans provoquer la multiplication des structures élémentaires. Enfin, le nucléus (et les opérations de combinaison qui vont avec) n'est pas une notion ad hoc, puisqu'il intervient dans la description de nombreux autres phénomènes syntaxiques, dont la coordination, la négation et la comparaison. L'extension de la grammaire à ces phénomènes est en cours.

En résumé, il nous faut bloquer l'inversion pour les nucléus dont l'un des nœuds porte un sujet ou un objet direct nominal non clitique. Cette condition autorise l'inversion du sujet en cas d'extraction du sujet de la complétive :

- (iv) *Je me demande ce que pense qui peut nous être utile le propriétaire du chalet.*

Il semble difficile de vérifier la condition précédente autrement que par l'ajout d'un trait sur le nucléus vV dont la valeur est recalculée à chaque fois qu'un nœud de la chaîne nucléaire réalise un sujet ou un objet direct.

RÉFÉRENCES

- ABEILLÉ Anne (1991) : *Une grammaire lexicalisée d'Arbres Adjoints pour le français*, Thèse de Doctorat, Univ. Paris 7.
- ABEILLÉ Anne & GODARD Danièle (1998) : "French Word Order and Lexical Weight", in Borsley R. (ed.), *Syntactic Categories, Syntax and Semantics*, NY : Acad. Press.
- AHO Alfred V. & ULLMAN Jeffrey D. (1972) : *The Theory of Parsing, Translation and Compiling*, Vol. 1, London : Prentice-Hall.
- BLINKENBERG A. (1928) : "L'ordre des mots en français moderne", *Bianco Lunos Bogtrykkeri, Historik-filologiske Meddelelser XVII:1 et XX:1*, Copenhague.
- CANDITO Marie-Hélène (1996) : "A Principled-based Hierarchical Representation of LTAG", *COLING'96*, Copenhague, pp. 194-99.
- CANDITO Marie-Hélène (1999) : *Organisation modulaire et paramétrable de grammaires électroniques lexicalisées. Application au français et à l'italien*, Thèse de Doctorat, Univ. Paris 7.
- CANDITO Marie-Hélène & KAHANE Sylvain (1998)a : "Une grammaire TAG vue comme une grammaire Sens-Texte précompilée", *TALN'98*, Paris, pp. 24-34.
- CANDITO Marie-Hélène & KAHANE Sylvain (1998)b : "Defining DTG Derivations to get Semantic Graphs", *TAG+4 Workshop*, Philadelphie, pp. 25-28.
- CHOMSKY Noam (1977) : "On Wh-movement", in P. Culicover et al. (eds), *Formal Syntax*, Dordrecht : Reidel.
- COURTIN Jacques & GENTHIAL Damien (1998) : "Parsing with Dependency Relations and Robust Parsing", in Kahane S. & Polguère A. (eds), *Workshop on dependency-based grammars*, *COLING-ACL'98*, Montréal, pp. 95-101.
- DANLOS Laurence (1998) : "G-TAG : un formalisme lexicalisé pour la génération de textes inspiré de TAG", *T.A.L.*, 39 : 2, pp. 7-33.
- DYMETMAN Marc & COPPERMAN Max (1996) : "Extended Dependency Structures and their Formal Interpretation", *COLING'96*, Copenhague, pp. 255-61.
- GAZDAR Gerald, KLEIN Ewan, PULLUM Geoffrey K. & SAG Ivan A. (1985) : *Generalized Phrase Structure Grammar*, Cambridge : Harvard Univ. Press.
- GLADKIJ Aleksej V. (1968) : "On describing the syntactic structure of a sentence" (en russe avec résumé en angl.), *Computational Linguistics*, 7, Budapest, pp. 21-44.
- HEPPLE Mark (1995) : "Hybrid Categorial Logics", *Bulletin of the IGPL*, 3, pp. 343-56.
- HUDSON Richard A. (2000) : "Discontinuity", *T.A.L.*, ce volume.
- JOSHI Aravind K. (1987) : "Introduction to Tree Adjoining Grammar", in Manaster Ramer (ed.), *The Mathematics of Language*, Amsterdam:Benjamins, pp. 87-114.
- JACKENDOFF Ray S. (1977) : *X' Syntax : A Study of Phrase Structure*, Linguistic Inquiry Monograph, MIT Press.
- KAHANE Sylvain (1997) : "Bubble Trees and Syntactic Representations", in Becker T. & Krieger H.-U. (eds), *Proc. MOL'5*, Saarbrücken : DFKI, pp. 70-76.
- KAHANE Sylvain (2000) : "How to solve some failures of LTAG", *TAG+5 Workshop*, Paris, pp. 123-28.
- KAHANE Sylvain, CANDITO Marie-Hélène, DE KERCADIO Yannick, "An alternative description of extractions in TAG", *TAG+5 Workshop*, Paris, pp. 115-22.
- KAHANE Sylvain & MEL' CUKIgor (1999) : "La synthèse des phrases à extraction en français contemporain. Du réseau sémantique à l'arbre syntaxique", *T.A.L.*, 40 : 2, pp. 25-85.

- KAHANE Sylvain, NASR Alexis & RAMBOW Owen (1998) : "Pseudo-Projectivity: a Polynomically Parsable Non-Projective Dependency Grammar", *COLING-ACL'98*, Montréal, pp. 646-52.
- KAYNE Richard (1975) : *French Syntax : the Transformational Cycle*, Cambridge : MIT Press.
- LECEF Yves (1961) : "Une représentation algébrique de la structure des phrases dans diverses langues naturelles", *C. R. Acad. Sc. Paris*, 252, pp. 232-234.
- KROCH Anthony (1987) : "Unbounded Dependencies and Subjacency in a Tree Adjoining Grammar", in Manaster Ramer A. (ed.), *The Mathematics of Language*, Amsterdam : Benjamins, pp. 143-71.
- LECOMTE Alain (1992) : "Connection Grammars : a Graph-oriented Interpretation", in A. Lecomte (ed), *L'ordre des mots dans les grammaires catégorielles*, Clermont-Ferrand : Adosa, pp. 129-48.
- LOMBARDO Vincenzo & LESMO Leonardo (1998) : "Formal Aspects and Parsing Issues of Dependency Theory", *COLING-ACL'98*, Montréal, pp. 787-93.
- MEL'CUK Igor (1967) : "Ordre des mots en synthèse automatique des textes russes", *T.A. Information*, 2, pp. 65-84.
- MEL'CUK Igor (1988) : *Dependency Syntax: Theory and Practice*, NY : State Univ. of NY Press.
- MEL'CUK Igor (1992) : "Paraphrase et lexique: la Théorie Sens-Texte et le Dictionnaire Explicatif et Combinatoire", in Mel'čuk et al., *Dictionnaire Explicatif et Combinatoire*, vol. 3, pp. 10-57.
- MEL'CUK Igor (à paraître) : *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*, Amsterdam : Benjamins.
- MEL'CUK Igor & PERTSOV Nicolas (1987) : *Surface syntax of English – A FormaModel within the Meaning-Text Framework*, Amsterdam : Benjamins.
- MOORTGAT Michael (1989) : *Categorial Investigations. Logical and Linguistic Aspects of the Lambek Calculus*, Dordrecht : Foris.
- NASR Alexis (1995) : "A Formalism and a Parser for Lexicalised Dependency Grammars", *4th Int. Workshop on Parsing Technologies*, State Univ. of NY Press.
- NASR Alexis (1996) : *Un modèle de reformulation automatique fondé sur la Théorie Sens-Texte – Application aux langues contrôlées*, Thèse de Doctorat, Univ. Paris 7.
- POLGUÈRE Alain (1990) : *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*, Thèse de Doctorat, Univ. de Montréal.
- POLLARD Carl & SAG Ivan A. (1994) : *Head-Driven Phrase Structure Grammar*, CSLI series, Chicago : Univ. of Chicago Press.
- ROSS John (1967) : *Constraints on Variables in Syntax*, PhD Thesis, MIT (publié sous le titre *Infinite syntax I*, Dordrecht : Reidel (1985).
- STEEDMAN Mark (1985) : "Dependency and Coordination in the Grammar of Dutch and English", *Language*, 61, pp. 523-568.
- STEEDMAN Mark (1997) : *Surface Structure and Interpretation*, Cambridge : MIT Press.
- TESNIÈRE Lucien (1959, 2^e ed. 1967)) : *Eléments de syntaxe structurale*, Paris : Klincksieck.
- XTAG Research Group (1995) : *A Lexicalized TAG for English*, Technical Report IRCS 95-03, Univ. of Pennsylvania, version mise à jour sur le web.