

L'amas verbal au cœur d'une modélisation topologique du français

Kim Gerdes & Sylvain Kahane

1. Introduction

Cette communication poursuit un double objectif. Premièrement, nous souhaitons montrer que les principaux phénomènes d'ordre des mots (montée des clitiques, inversion du sujet, extraction, dislocation, ...) peuvent être modélisés dans une approche topologique où la linéarisation est traitée par la mise en correspondance d'une structure syntaxique de dépendance (sans ordonnancement) avec une structure ordonnée de type structure de constituants topologiques.

Deuxièmement, nous voulons mettre en évidence un constituant particulier, très rarement considéré dans les modélisations de l'ordre des mots en français et qui s'impose naturellement dans une modélisation topologique du français : ce constituant, que nous appelons *amas verbal*, est constitué d'un amas d'un ou plusieurs verbes, des éléments grammaticaux nécessaires à la formation de cet amas (prépositions et complémenteurs) et d'éléments lexicaux très contraints, c'est-à-dire en dehors des verbes eux-mêmes (ou d'éléments prédicatifs tels que adjectifs ou noms prédicatifs appartenant à des tournures verbales) des clitiques et des adverbes.

2. Une première approche de l'amas verbal

On peut donner deux justifications à l'existence des amas verbaux. D'un point de vue onomasiologique, il y a une nécessité dans la langue à pouvoir exprimer des prédications complexes formées de différents éléments décrivant l'événement, y compris des éléments de sens exprimant modalités, aspects, temps, causalité, mise en évidence de tel ou tel participant, etc. Dans les langues qui, tel le français, expriment ces sens par des verbes, il faut s'attendre à la formation de prédications complexes formées de chaînes de verbes et d'autres éléments prédicatifs tels que adjectifs et noms prédicatifs : *devoir partir*, *faire mourir*, *chercher à réussir*, *avoir l'intention de réussir*, *avoir beau être capable de réussir*, etc. Ces prédicats complexes vont alors être la source d'une ambivalence fondamentale : d'un côté, ils sont complexes et formés de l'agrégation de plusieurs sens et de plusieurs mots, mais d'un autre, ils forment à eux tous une seule et unique prédication (la description d'un événement par exemple) et vont de ce point de vue se comporter comme une prédication simple. La langue se doit donc d'offrir dans son dispositif syntaxique la possibilité de gérer de telles unités complexes (le terme d'*unité complexe* résumant à lui seul la nature de ces objets).

D'un point de vue sémasiologique, on peut voir plusieurs indices dans la syntaxe du français¹ de la présence d'amas verbaux. L'un des indices les plus nets est l'existence de phénomènes de non projectivité mettant en jeu des compléments verbaux (sujet compris). Sans rentrer dans

¹ Nous concentrons notre étude sur le français, mais il va de soi que l'amas verbal n'est en rien une particularité du français. On peut par exemple en montrer facilement la présence dans les langues germaniques, où la chaîne verbale se regroupe en fin de phrase (et forme clairement une unité prosodique) laissant les différents compléments des différents verbes de cette chaîne se « mélanger » comme il le veulent pour l'allemand (phénomène connu sous le nom de *scrambling* ; cf. Gerdes 2002) ou dans le même ordre que les verbes dont ils dépendent comme en néerlandais provoquant de multiples croisements de dépendances (*cross serial dependencies*).

une définition formelle de la non projectivité d'un arbre linéairement ordonné (cf. pour cela Lecerf 1961, Gladkij 1966 ou Kahane 2000), on peut définir la notion de la manière suivante : une construction est dite non projective quand un élément X ne se place pas rapport à son gouverneur² syntaxique Y, mais plutôt par rapport à un autre élément, comme le gouverneur de Y ou encore le gouverneur de celui-ci et ainsi de suite.

Le phénomène de non projectivité le plus clair et le plus simple en français (simple s'entend ici d'un strict point de vue des structures mathématiques et non de la grammaire qui les décrit) est certainement la montée des clitiques : en effet, comme chacun sait, en français (ainsi que dans les autres langues romanes), certains compléments pronominaux se cliticisent et leur hôte n'est pas nécessairement le verbe qui les sous-catégorise. Ainsi, en français, les compléments clitiques d'une forme verbale composée ne se placent-ils pas sur le participe (qui est l'élément prédicatif qui contrôle la sous-catégorisation)³, mais sur le verbe auxiliaire, qui comme son nom l'indique n'est qu'un auxiliaire dont la fonction est purement grammaticale.

Or l'auxiliaire et le participe peuvent être séparés par d'autres mots (*Pierre ne lui a pas encore vraiment répondu*) et à ce titre doivent être considérés comme deux mots à part entière (même si leur sort est lié du point de vue de l'interface sémantique-syntaxe). Néanmoins, les éléments qui peuvent s'interposer entre l'auxiliaire et le participe sont très contraints ; il s'agit seulement de certains adverbes : *Pierre a déjà répondu*, *Pierre a tout de suite répondu*, *Pierre est rapidement venu*, **Pierre a hier répondu*, **Pierre est ici venu*, etc. Il ne peut par exemple jamais s'agir de groupe prépositionnel⁴ : **Pierre a en 3 secondes répondu*, **Pierre est à toute vitesse venu*, etc.

Il nous semble que les deux propriétés sont corrélées, à savoir la possibilité pour un clitique de « monter » sur l'auxiliaire et l'impossibilité pour des éléments trop « complexes » de s'intercaler entre l'auxiliaire et le participe. En effet, l'une et l'autre sont pour nous la trace d'un même phénomène, la formation d'un « nucléus verbal » auxiliaire-participe, c'est-à-dire l'existence d'une prédication verbale complexe qui d'un certain point de vue est bien complexe (c'est-à-dire formée de plusieurs mots), mais d'un autre continue à se comporter comme un tout et garde certaines caractéristiques d'une forme simple. La modélisation que nous proposons de la « montée » des clitiques rend justement compte de cette corrélation. En effet, nous rendons possible la « montée » des clitiques par la formation d'un constituant privilégié autour de l'auxiliaire et du participe et c'est cette même formation qui rend naturelle la possibilité de contraindre les éléments qui vont apparaître dans cette formation.

² Nous considérons la notion de gouverneur syntaxique comme acquise. Rappelons grossièrement qu'un élément X est gouverné par Y s'il est un argument sous-catégorisé de Y ou s'il modifie Y. La notion peut être sujet à controverses pour des relations telles que celles qu'entretiennent le nom et son déterminant, mais elle nous semble relativement consensuelle en ce qui concerne les compléments (actanciels et circonstanciels) du verbe. Notons que nous considérons qu'un sujet est toujours gouverné par le verbe fini qui s'accorde avec lui. En particulier, le sujet d'une forme verbale composée dépend de l'auxiliaire et non du participe.

³ Certaines approches résolvent en fait la question de la « montée » des clitiques par un véritable mécanisme de montée de ces clitiques, lesquels vont figurer dans la sous-catégorisation de l'auxiliaire, qui devient alors variable (cf. Miller & Sag 1997).

⁴ Nous laissons de côté les incises qui obéissent clairement à un autre mécanisme de placement et qui peuvent venir « couper » l'amas verbal : *Pierre a déjà, comme tu le sais, répondu à cette question*. Nous pensons que les incises sont des éléments qui n'obéissent pas stricto sensu au modèle topologique : contrairement aux autres éléments, elles ne viennent pas se placer dans des champs prédéfinis, mais viennent au contraire s'insérer entre deux champs. Tout au plus, le modèle topologique contrôle-t-il le placement des incises en interdisant l'accès à certaines portions (par exemple entre les clitiques et le verbe).

On notera au passage que les adverbes qui peuvent s'intercaler entre l'auxiliaire et le participe sont des adverbes modalisants (*Pierre est **souvent** venu ici*) et jamais des adverbes locatifs (**Pierre est **ici** venu souvent*). Hors ces derniers fixent des cadres externes à l'action, tandis que les premiers visent directement les caractéristiques aspectuelles de l'action. Ces mêmes éléments sémantiques pourront d'ailleurs, dans les langues à morphologie riche, être intégrés à la flexion du verbe (cf. par exemple Mel'cuk 1994 pour une typologies des significations flexionnelles).

Le deuxième phénomène qui va nous intéresser est l'inversion du sujet. Là aussi nous pensons que ce phénomène met en évidence un amas verbal. Premièrement, l'« inversion » du sujet ne se fait pas par rapport au seul verbe principal, mais par rapport à quelque chose que délimite justement notre amas verbal :

*le livre que **Pierre** veut lire*
le livre que veut **Pierre lire*
*le livre que veut lire **Pierre***

Deuxièmement, l'ordre respectif du sujet inversé et des compléments des autres verbes de la chaîne verbale est libre :

*le livre que veut donner **Pierre** à sa sœur*
*le livre que veut donner à Marie **son frère***

Bien qu'il soit clair d'un point de vue syntaxique que le sujet dépende du verbe fini (accord) et que l'objet indirect dépende de l'infinitif (rection), les deux verbes mettent en commun dans un même pot leurs dépendants pour la linéarisation. Ceci est à notre avis la trace de l'existence du constituant *amas verbal*.

3. Le modèle topologique du verbe en français

Dans une approche topologique, le calcul de l'ordre des mots est géré au niveau de la correspondance entre une structure de dépendance syntaxique (non ordonnée) et une structure de constituants ordonnée. Les nœuds de la structure syntaxique sont les mots. Lors de la linéarisation de l'arbre, chaque mot va ouvrir un *constituant topologique*. Un tel constituant est comme une boîte dont les cases ordonnées sont appelées les *champs*. L'élément qui a ouvert un constituant en occupe un certain champ et les éléments qui dépendent de lui vont venir occuper les autres champs. On doit spécifier pour chaque champ s'il peut rester vide ou non et s'il peut accueillir ou non plus d'un élément. Un élément peut par ailleurs *s'émanciper* du constituant ouvert par son gouverneur syntaxique et venir se placer dans un constituant ouvert par un élément plus haut⁵ dans l'arbre syntaxique. Dans un tel formalisme, une règle d'ordre est donc une règle qui spécifie pour un dépendant donné quel champ il peut occuper et s'il peut ou non s'émanciper et jusqu'où en fonction :

1. de sa catégorie
2. et/ou de sa fonction syntaxique⁶.

⁵ Remarquons que nous n'autorisons pas un élément à venir se placer plus bas, c'est-à-dire à pénétrer dans le constituant ouvert par un de ses dépendants. Ceci rend compte d'une propriété notoire des langues et que d'autres formalismes saisissent par des moyens similaires. HPSG par exemple, reprenant un mécanisme introduit dans le cadre des grammaires catégorielles (Ades & Steedman 1982), permet à un élément d'être « slashé » et de remonter ainsi dans la structure.

⁶ Ceci dépend des langues. Les langues non casuelles comme le français ou l'anglais ont tendance à placer les éléments en fonction de leur fonction syntaxique (par exemple le sujet se distingue de l'objet par sa place par rapport au verbe), alors que dans les langues casuelles la fonction est indiquée par des moyens morphologiques. Dans ces langues où l'ordre des mots n'est pas contrôlé par la fonction syntaxique, il dépend principalement d'un autre facteur : la structure communicative (ou

A ces règles s'ajoutent les règles qui disent quel type de constituant pourra ouvrir un mot placé dans un champ donné et pour chaque constituant, quelle est la liste des champs qui le composent et pour chaque champ, combien de constituants peuvent ou doivent l'occuper (cf. Gerdes & Kahane 2001 pour une présentation formelle d'une grammaire topologique).

Nous avons dans un premier temps dit que chaque mot ouvrait un constituant. La structure résultante sera alors dit *plate*, le nombre de constituants topologiques étant exactement égal au nombre de mots. Mais on peut également, à l'instar des structures X-barre⁷, autoriser chaque mot à ouvrir un certain enchâssement de constituants plutôt qu'un seul constituant. Formellement les deux solutions sont équivalentes. La structure *plate*, en présentant simultanément l'ensemble des champs ouverts par un mot, permet d'exprimer plus facilement des contraintes liant des champs qui dans une autre modélisation apparaîtraient dans des constituants différents. C'est le cas par exemple dans la modélisation de l'inversion du sujet où nous devons indiquer que, quand le sujet n'est pas clitique, l'un des champ du domaine micro précédent le verbe doit être rempli (cf. *infra*). Ceci peut nous amener à préférer une structure *plate* dans un formalisme relativement contraint comme celui proposé par Gerdes & Kahane 2001. Néanmoins, la deuxième solution permet deux choses. Premièrement, elle permet une certaine économie dans l'écriture de la grammaire en permettant de factoriser des regroupements qui apparaissent dans différentes règles : par exemple, une proposition relative diffère d'une proposition principale par la présence d'un groupe extrait et l'absence d'éléments détachés à gauche, mais obéit néanmoins aux mêmes règles d'ordre pour ce qui concerne la majorité des compléments. On pourra donc séparer la description d'une proposition en un premier domaine où sont placés les éléments détachés et extraits et un domaine enchâssé où sont placés les éléments non détachés (cf. domaines macro- et micro-syntaxique ci *infra*). Deuxièmement, elle permet de mettre en évidence des regroupements plus privilégiés, lesquels se traduiront par exemple dans la structure phonologique par la formation obligatoire d'un constituant prosodique. Ainsi, les clitics entretiennent-ils avec le verbe un rapport plus étroit que les autres compléments nominaux et peuvent être regroupés au sein d'un constituant intermédiaire, qui est justement notre *amas verbal*.

Venons-en à la modélisation de l'ordre des dépendants du verbe et donc à la description du ou des constituants topologiques ouverts par le verbe. Nous considérons trois niveaux d'enchâssement ici. Nous commençons par présenter les constituants ouverts par le verbe principal d'une phrase. Nous discuterons ensuite des constituants ouverts par des verbes enchâssés pour lesquels des modifications peuvent être nécessaires.

- Le premier niveau est constitué par l'*amas verbal* : celui-ci comprend les champs pour les clitics avant le verbe et deux champs après le verbe. Les champs clitics sont au nombre de 7 : *il < ne < me < le < lui < y < en*. Les deux autres champs accueillent successivement des adverbes et un dépendant verbal du verbe. La présence de ce dernier champ dans l'*amas verbal* constitue très certainement la grande originalité de notre approche⁸ et nous aurons l'occasion d'y revenir tout au long de cet article. A

informationnelle) qui vient se superposer à la structure de dépendance (cf. Mel'cuk 2001, Yoo & Gerdes 2003).

⁷ A noter que contrairement à la syntaxe X-barre, notre structure est purement surfacique et dédiée au calcul de l'ordre des mots et au regroupement des mots en surface. Notre structure de constituant n'encode nullement relations de sous-catégorisation ou autres relations syntaxiques, lesquelles sont encodées dans la structure syntaxique proprement dite.

⁸ De nombreuses approches actuelles en analyse syntaxique automatique effectuent un « tronçonnage » préalable de la phrase en tronçons appelés *chunks* (Abney 1991, Vergne 2000). Le chunk verbal

noter que l'amas verbal offre également une position pour l'enclitique en cas de construction interrogative (*Pierre lui donnera-t-il la réponse ?*). Le verbe à l'impératif nécessite un amas verbal spécifique où tous les pronoms sont enclitiques (*Donne-le-lui ! *Donne-lui-le ! Donne-le-moi ! Donne-moi-le ! Donne-m'en ! Donne-moi-z-en ! ? Donne-z-en-moi !*).

- Le deuxième niveau est constitué par le *domaine micro-syntaxique*. Ce domaine inclut tous les éléments sous-catégorisés et les modifieurs de verbe (par opposition aux « modifieurs de phrase »). Nous proposons d'appeler ces éléments les *valents*⁹ du verbe, le terme *actant* s'appliquant aux seuls éléments sous-catégorisés et *dépendant* incluant également les éléments détachés. Nous reprenons ici la distinction entre micro- et macro-syntaxe faite par Blanche-Benveniste 1990 et Berrendoner 1991 à qui nous renvoyons pour une définition plus précise, ainsi qu'à Deulofeu 2003. Le domaine micro-syntaxique comprend au moins un champ pour le sujet, un champ pour l'amas verbal et un champ pour les compléments. Tous les compléments sont placés dans le même champ car leur ordre respectif est libre, ou du moins, il n'est pas guidé par les mêmes paramètres que ceux que nous modélisons par la structure topologique.
- Le troisième niveau est constitué par le *domaine macro-syntaxique*. Celui-ci comprend les éléments détachés à gauche ou à droite : *Moi, ma sœur, il ne veut plus lui parler, Pierre ; ce qui est sûr, c'est que Pierre ne viendra pas*. Nous postulons deux champs à gauche et un champ à droite.

Nous reviendrons brièvement dans la section 7 sur les champs des domaines micro et macro.

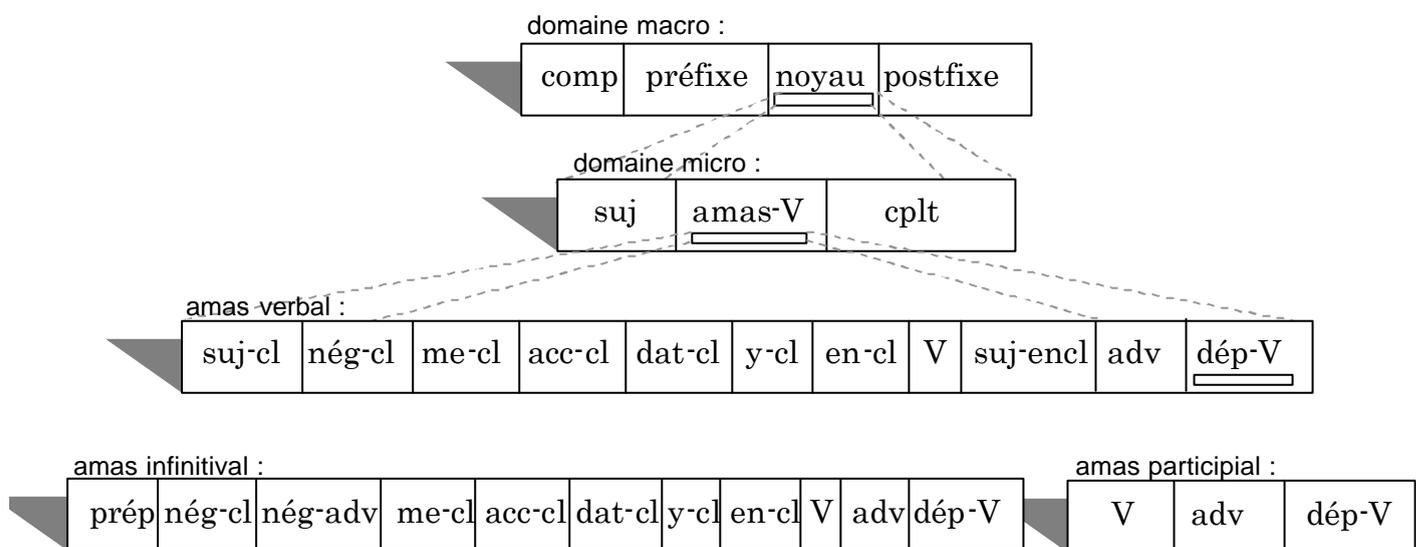
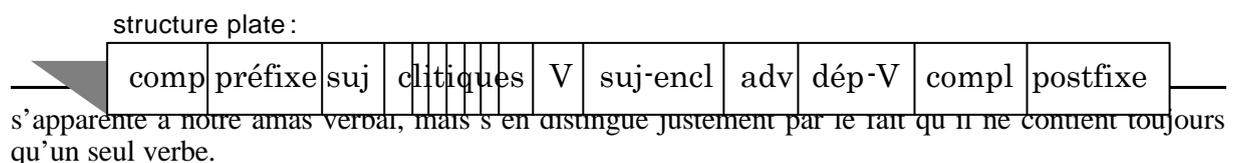


Figure 1. Les trois constituants du verbe principal à l'indicatif (enchâssés l'un dans l'autre) et les constituants des verbes non finis (qui peuvent apparaître dans le champ *dep-V* de l'amas verbal).



⁹ De *valence*, lui-même construit sur lat. *valere* 'valoir'. Les valents sont les éléments qui valent suffisamment pour que le verbe les garde auprès de lui.

Figure 2. La topologie du verbe principal à l'indicatif en version plate.

Figure 2. La structure plate (où les trois constituants du verbe fini apparaissent ensemble, combinés dans un seul gabarit de places)

Passons maintenant aux verbes enchâssés.

Notre premier exemple est le verbe au participe passé d'un passé composé. Le participe passé dépend du verbe auxiliaire et doit nécessairement aller dans l'amas verbal ouvert par l'auxiliaire, où il ouvrira lui-même un amas verbal. Le participe passé possède la propriété particulière de ne pas pouvoir accueillir de clitiques, ce qui peut être modélisé en bloquant les positions pour les clitiques dans l'amas verbal ouvert par un participe passé.

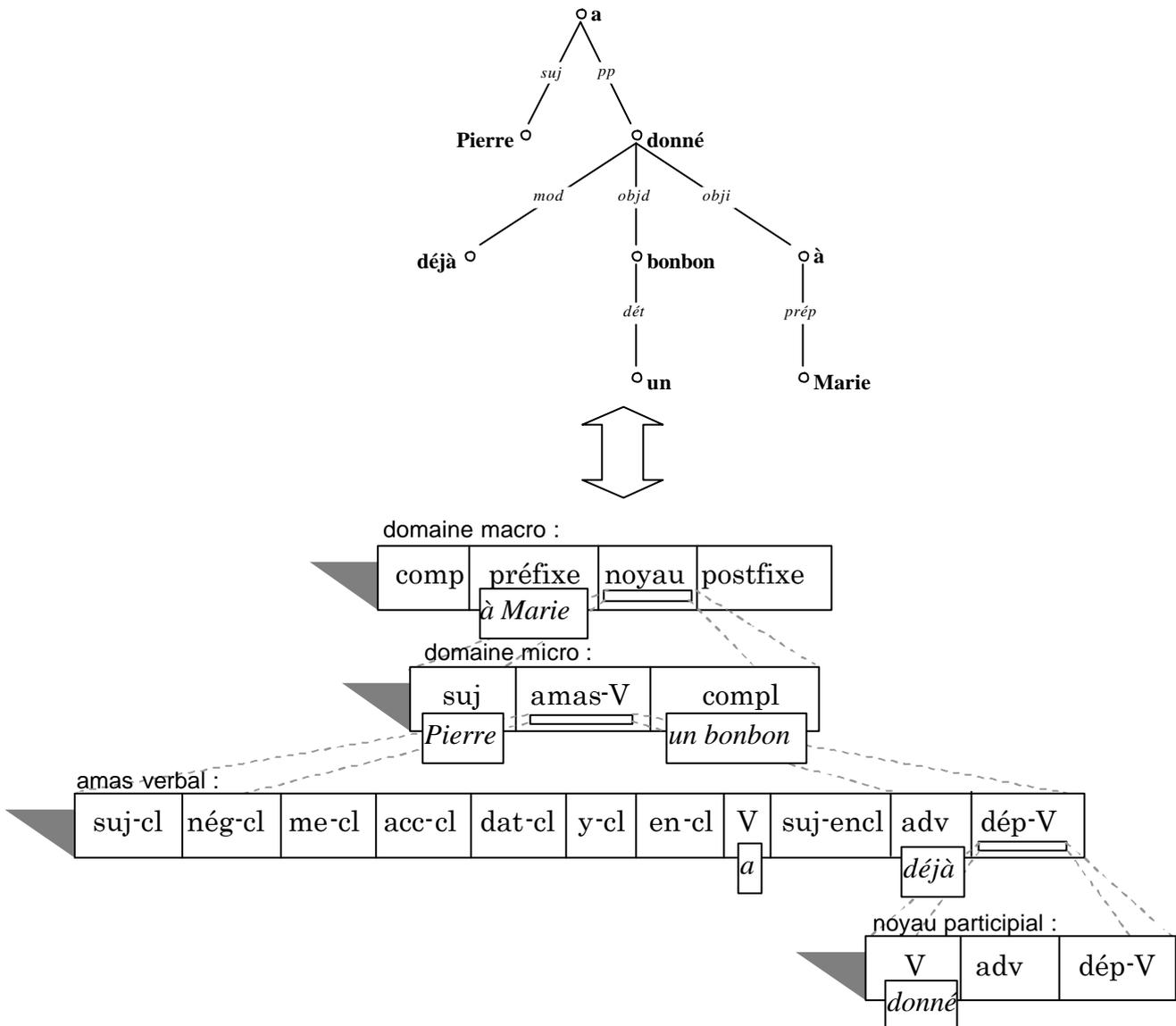


Figure 3. Exemple : arbre de dépendance et structure topologique de la phrase *A Marie, Pierre a déjà donné un bonbon.*

Montrons comment est établie la correspondance entre l'arbre de dépendance syntaxique de cette phrase et sa structure topologique¹⁰. Plaçons nous dans le sens de la synthèse (le locuteur parle et transforme un sens en une phrase en passant par une représentation syntaxique). Le verbe auxiliaire *a*, qui est la racine de l'arbre de dépendance, ouvre d'abord trois constituants enchâssés : un domaine macrosyntaxique, un domaine microsyntaxique et enfin un amas verbal à l'intérieur de ce dernier. Ses dépendants vont pouvoir se placer dans les différents champs ouverts : le sujet *Pierre* ira dans le champ sujet du domaine micro, tandis que le participe passé *donné* devra occuper le champ du dépendant verbal dans l'amas verbal, où il ouvrira un amas verbal enchâssé (pour un éventuel dépendant verbal comme dans *Pierre a voulu venir*). L'adverbe *déjà* restera préférentiellement dans l'amas, où il précèdera le participe¹¹. Le complément d'objet direct *un bonbon*, qui dépend de *donné* ne pourra pas aller dans le constituant topologique ouvert par son gouverneur (un amas dans l'amas) et devra s'émanciper pour atteindre le champ complément du domaine micro ouvert par l'auxiliaire. Le complément d'objet direct *à Marie* pourrait suivre le même chemin, mais il peut aussi comme ici, en raison d'un marquage communicatif particulier, rejoindre le champ préfixe du domaine macro (c'est la topicalisation).

Contrairement à un participe passé, un infinitif a deux possibilités : il peut

- soit, comme le participe passé, aller dans le champ dépendant de l'amas verbal de son gouverneur, où il ouvrira lui-même un amas verbal (enchâssé) ;
- soit, comme les autres compléments, rejoindre le champ complément (ou même un champ du domaine macro) et ouvrir un domaine micro qui pourra accueillir tous ses dépendants.

Dans les deux cas, le verbe infinitif ouvre un amas verbal : soit directement s'il forme un amas avec son gouverneur, soit à l'intérieur de son domaine micro s'il prend un peu d'indépendance. L'amas verbal ouvert par un infinitif (le même dans les deux cas) est un peu différent de celui d'un verbe fini en raison du placement particulier du champ adverbe entre le « clitique » *ne* et les autres clitiques : *Pierre pense [ne pas le lui dire]*, *Pierre a peur [de ne vraiment rien y comprendre]*, etc.

Un verbe fini d'une complétive ou d'une subordonnée circonstancielle (*Pierre sait [qu'hier Jean n'est pas venu]* ; [*Quand tu l'auras fini, le livre*], *rapporte-le-moi*) ouvre un domaine

¹⁰ Par soucis de simplification, nous masquons totalement le traitement de la morphologie. Nous manipulons donc directement des formes fléchies sans montrer comment parallèlement à la construction de la phrase la formation des mots peut être assurée.

¹¹ Le placement des adverbes dans l'amas verbal mérite une étude plus poussée. Comment rend-on compte de l'ordre entre plusieurs adverbes quand celui-ci est contraint :

Pierre a déjà tout donné
 **Pierre a tout déjà donné.*

Une possibilité serait d'avoir un champ adverbial devant le verbe dans l'amas ouvert par un participe (le même qu'on a devant un infinitif dans *Pierre espère [ne pas tout perdre]*). L'« adverbe » *tout* resterait alors dans l'amas enchâssé ouvert par le participe et ne pourrait donc pas précéder l'adverbe *déjà*, qui dépend de l'auxiliaire et qui, s'il dépendait du participe, se placerait après celui-ci (*Pierre a tout donné déjà*). Ceci permettrait également de rendre compte du contraste dans les participiales entre participe passé et participe présent dans le placement des adverbes :

un livre souvent lu par les étudiants
 **des étudiants souvent lisant ce livre*
des étudiants lisant souvent ce livre.

macro comme le verbe principal d'une phrase. Ce domaine macro comporte en plus des champs pour les éléments détachés, un champ pour le complémenteur (*que, si, quand, etc.*).

Le cas des relatives est tout à fait intéressant. Elles se distinguent des complétives, du point de vue topologique, par deux propriétés :

1. L'impossibilité d'avoir un détachement à gauche :
J'ai retrouvé le livre que je voulais prêter à Pierre
**J'ai retrouvé le livre que, Pierre, je voulais lui prêter*
Tu pensais que, Pierre, je voulais lui prêter le livre
**la personne à qui, le livre, je l'ai donné*
la personne à qui je l'ai donné, le livre, (elle) doit passer demain
2. La possibilité d'une inversion du sujet :
le livre que Pierre lit
le livre que lit Pierre
**Marie croit que dort Pierre*
*?*Marie veut que dorme Pierre*

Ces deux propriétés nous semblent résulter d'une même caractéristique, qui distingue fondamentalement la relative de la complétive : le fait que le complémenteur d'une relative est aussi un pronom et qu'il joue ainsi un rôle valentiel dans la proposition qu'il introduit¹². Pourquoi cela aurait-il une influence sur l'ordre des mots dans cette proposition? Comme nous l'avons vu pour la complétive, le complémenteur occupe la position la plus à gauche du domaine macro. Par contre, les éléments qui jouent réellement un rôle dans la proposition reste dans le domaine micro. Le pronom relatif, qui est les deux à la fois, va donc occuper le champ du complémenteur, mais en même temps, il va attirer ce champ dans le domaine micro, avec deux conséquences :

1. le champ préfixe est écrasé, bloquant ainsi dislocation gauche et topicalisation ;
2. un des champs du domaine micro précédant le verbe est maintenant occupé, validant ainsi l'inversion du sujet, c'est-à-dire permettant au sujet d'aller dans le champ complément.

Nous allons maintenant revenir sur les différents phénomènes qui nous intéressent en montrant comment la description permet de les modéliser.

4. La montée des clitiques

Comme nous l'avons vu, l'amas verbal est au centre de notre modélisation de la montée des clitiques. Pour traiter la montée de *lui* dans *Pierre lui a parlé*, nous considérons que le participe passé rejoint l'amas verbal ouvert par l'auxiliaire, où le participe ouvre lui-même un constituant de type amas verbal sans position pour les clitiques ; les compléments clitiques du participe vont donc s'émanciper pour se placer dans les champs clitiques de l'amas verbal ouvert par l'auxiliaire. Notre modélisation repose finalement sur trois propriétés :

1. la nécessité pour le participe de former un amas verbal avec l'auxiliaire ;
2. l'impossibilité pour le participe d'être l'hôte des clitiques ;

¹² Les complémenteurs de relative *que* et *dont* sont particuliers du fait qu'il n'accepte pas le pied-piping, c'est-à-dire d'entraîner avec eux un élément de la proposition, comme le fait le *qui* humain (*la personne à la mère de qui je parle* ; **la personne à la mère dont je parle*). Ceci peut amener à ne pas les traiter comme des pronoms, au même titre que le *qui* humain. Quoi qu'il en soit, tous ces complémenteurs viennent saturer une position syntaxique dans la proposition.

3. la possibilité pour les clitiques de s'émanciper de l'amas verbal du participe.

A propos de cette dernière propriété, notons que les clitiques ont un faible pouvoir d'émancipation. Ils peuvent seulement s'échapper d'un amas verbal enchâssé. On peut d'ailleurs modéliser le fait que certains verbes ne peuvent réaliser un argument que comme clitique (*il nous faut repenser notre stratégie* ~ **il faut à Pierre repenser sa stratégie* ~ *il lui faut repenser sa stratégie, à Pierre* ; *il me semble que c'est possible* ~ **il semble à Pierre que c'est possible* ~ *il lui semble que c'est possible, à Pierre*) en considérant que cet argument ne peut s'échapper de l'amas verbal.

Notons que notre modélisation s'applique également aux emplois du participe passé comme modifieur de nom ; ici aucun champ clitique n'est accessible et les clitiques ne peuvent pas se réaliser : *le livre donné à Pierre* ~ **le livre lui donné* ~ *le livre lui a été donné par Marie*.

Notre modélisation s'étend facilement aux autres cas de montée de clitique :

Pierre en est content [de sa nouvelle voiture] ~ **Pierre est en content*

Pierre le fait manger aux invités [le poulet] ~ **Pierre fait le manger aux invités*

Pierre l'a entendu chanter [la Marseillaise] ~ **Pierre a entendu la chanter*

De même qu'un participe passé, un adjectif gouverné par une copule ou un infinitif gouverné par *faire* devra nécessairement rejoindre l'amas verbal de son gouverneur, où il ouvrira un amas enchâssé sans position pour les clitiques. Notons qu'un adjectif ne peut jamais ouvrir un constituant avec des places pour les clitiques (*un type content de sa voiture* ~ **un type en content*), alors que pour un infinitif c'est la rection particulière du causatif *faire* qui va déclencher l'ouverture d'un amas enchâssé sans position pour les clitiques.

5. L'inversion du sujet

Venons-en maintenant à notre modélisation de l'inversion du sujet. Nous laissons de côté l'inversion dite élaborative (Marandin 2003), qui n'est possible qu'avec une classe particulière de verbes à sujets peu ou non agentifs et pour laquelle le sujet doit occuper la dernière position du domaine micro :

Sont possibles de rares inversions avec verbe initial

Arrivèrent alors des hordes de barbares

Rendront un devoir supplémentaire les élèves qui ont raté l'examen de chimie.

Nous allons nous intéresser à l'inversion dite stylistique qui obéit à des contraintes syntaxiques : il ne doit pas y avoir de complément d'objet direct dans le champ complément (**la personne à qui donne le livre Pierre*)¹³ et la proposition doit comprendre un extraction : interrogation en *qu*, relative, coda de clivée ou topicalisation de groupe prépositionnel (Marandin 2003). La deuxième condition se ramène, selon nous, à la présence d'un élément (non clitique) devant le verbe dans le domaine micro, soit que l'élément topicalisé occupe le champ sujet (voir section 6), soit, comme nous l'avons déjà dit pour les relatives, que le champ complémentateur est aspiré dans le domaine micro lorsque le complémentateur est en même temps un élément valent.

Nous n'étudierons pas davantage les raisons de l'inversion du sujet. Nous souhaitons juste ici montrer rapidement comment notre modélisation topologique assure le bon placement du sujet inversé. En voici le principe. Le sujet inversé va dans le champ complément avec les autres compléments des verbes de l'amas verbal. Nous assurons ainsi que le sujet inversé puisse se placer de façon quelconque vis-à-vis des autres compléments des verbes de l'amas verbal, car dans un même champ, l'ordre des éléments qui s'y trouvent réunis est par définition libre :

¹³ Le complément d'objet direct peut par contre être extrait (*le livre que lit Pierre*) ou cliticisé (*la personne à qui le donne Pierre*).

le livre que veut donner Pierre à sa sœur

le livre que veut donner à Marie son frère.

Ces dernières données, facilement traitées ici par la formation d'un amas verbal, sont fort problématiques pour les grammaires syntagmatiques, qui y voit un cas de «scrambling», le sujet inversé (ici *Pierre*) venant s'insérer dans l'infinitive (ici *donner à sa sœur*) et se « mélanger » avec les compléments du verbe infinitif.

Le fait que, dans *le livre que Marie veut lire*, l'inversion soit seulement possible par rapport à *veut lire* (*le livre que veut lire Marie* ; **le livre que veut Marie lire*) est traité en obligeant *lire* à rejoindre l'amas verbal ouvert par *veut*. Nous discuterons dans la section suivante des conditions qui vont forcer la création de l'amas verbal.

Voyons maintenant quels sont les éléments qui peuvent former un amas verbal. Nous avons dit que les participes et les infinitifs nus devaient aller dans l'amas ouvert par leur gouverneur¹⁴. D'autres éléments peuvent y aller : les infinitifs précédés d'un marqueur (*le livre que rêve de lire Marie* ; ?*le livre que rêve Marie de lire*), les adjectifs attribués (*le livre qu'a été heureuse de lire Marie* ; **le livre qu'a été Marie heureuse de lire*) ou les noms prédicatifs (*le livre qu'avait l'intention de lire Marie* ; **le livre qu'avait Marie l'intention de lire*). Des verbes subordonnés finis peuvent également venir dans l'amas verbal de leur gouverneur lorsque leur sujet est clitique :

le livre que veut qu'on lise Marie

**le livre que veut que Pierre lise Marie.*

Il semble qu'un verbe avec un sujet non clitique puisse quand même aller dans l'amas verbal. Auquel cas, son sujet ne peut aller dans l'amas et devra rejoindre le champ complément : ?*le livre que Marie veut que lise Pierre*. Notons également la construction suivante : *Ce qu'a dit que ferait Luc à sa petite amie est inquiétant* (Claude Muller, communication orale). Ici le syntagme *Luc* sature en même temps la position sujet des deux verbes finis de l'amas – *a* et *ferait*. Ce phénomène ne met donc pas simplement en cause l'ordre des mots, mais met en évidence un lien, plus étroit que ne le suggère notre traitement, entre l'ordre des mots et la structure syntaxique, puisque seule l'inversion du sujet valide le partage de sujet : **Ce que Luc a dit que ferait à sa petite amie est inquiétant*.

Une autre construction peut être rapproché de l'inversion du sujet — l'inversion du sujet logique dans les phrases telles que :

J'ai vu ma sœur sortir des toilettes.

Le syntagme *ma sœur* est le sujet logique de l'infinitif *sortir*, mais il est important de remarquer qu'il n'en est pas le sujet syntaxique¹⁵ : d'une part, un infinitif n'a de sujet syntaxique dans aucune autre construction et d'autre part, ce syntagme se cliticise comme un accusatif sur *voir* (*Je l'ai vu sortir des toilettes*) ou s'interroge comme un objet (*Qu'as-tu vu sortir des toilettes ?*), ce qui montre qu'il s'agit du complément d'objet direct de *voir*. Néanmoins, ce complément peut venir s'insérer entre *sortir* et ses compléments :

J'ai vu sortir ma sœur des toilettes.

Ce cas de « scrambling » rappelle évidemment celui de l'inversion du sujet. Nous proposons de le résoudre de la même façon, par la formation d'un amas verbal : le verbe *voir* peut former avec l'infinitif qu'il gouverne un amas verbal. Les compléments de l'infinitif devront donc

¹⁴ Les infinitifs nus peuvent néanmoins ouvrir un domaine comme dans *Pierre regarde Marie lire un livre*. Ils le feront également en cas de dislocation : *Lire ce livre, j'aurais dû le faire il y a longtemps*. Mais dans ce cas, ce n'est pas l'infinitif nu qui est régi par le verbe introducteur (ici *devoir*), mais le pro-verbe *le faire* avec lequel il coréfère. La dislocation droite nécessite « bizarrement » l'introduction du marqueur *de* : *J'aurais dû le faire, *ø/de lire ce livre*.

¹⁵ Il est à noter que la grammaire générative a toujours traité cette construction (ou son équivalent anglais) comme une proposition enchâssée : *J'ai vu* [[*ma sœur*]_{GN} [*sortir des toilettes*]_{GV}]_s.

s'émanciper et se placer dans le champ complément ouvert par *voir*, où ils se placeront librement par rapport au complément d'objet direct de *voir* qui s'y trouve. Cette analyse est encore corroborée par le contraste entre les deux constructions en ce qui concerne les clitiques :

J'ai vu ma sœur en sortir [des toilettes]

**J'en ai vu ma sœur sortir*

??*J'ai vu en sortir ma sœur*¹⁶

J'en ai vu sortir ma sœur

Comme on le voit, dans l'ordre « standard », le clitique va sur l'infinitif et ne peut monter sur *voir* (ou même plus haut, sur l'auxiliaire de *voir* comme dans l'exemple), alors qu'en cas d'« inversion », le clitique se place plus difficilement sur l'infinitif, mais il peut par contre monter sur *voir*. Ceci rapproche la construction en *voir* + infinitif avec « inversion » de la construction causative en *faire* + infinitif, où les clitiques doivent monter sur *faire*. Remarquons un contraste curieux entre les différents clitiques : alors que le clitique *en* peut monter sur *voir*, le clitique *lui* ne le peut pas :

J'ai vu le directeur lui parler [à Jean]

??*J'ai vu lui parler le directeur*

**Je lui ai vu parler le directeur.*

Notre formalisme permet facilement d'exprimer cette contrainte, sans pour autant pouvoir l'expliquer.

Les formes sans sujet logique (*J'ai entendu crier dans la cour ; J'ai vu jouer cette pièce deux fois*)¹⁷ montrent une plus grande cohésion encore de l'amas verbal, avec l'obligation de montée du clitique :

Je l'ai vu jouer deux fois [cette pièce]

**J'ai vu la jouer deux fois.*

Notons encore que l'inversion du sujet logique obéit aux mêmes contraintes que l'inversion du sujet concernant la non-cooccurrence avec un complément d'objet :

J'ai vu ma sœur manger le gâteau

**J'ai vu manger ma sœur le gâteau.*

Dans les deux constructions, cette non-cooccurrence est bloquée en interdisant deux compléments non régis à se trouver dans le champ complément.

Un dernier argument en faveur de l'amas verbal en cas d'inversion est illustré par le contraste suivant :

A la télé, j'ai vu parler quelqu'un sans bouger les lèvres

A la télé, j'ai vu quelqu'un parler sans bouger les lèvres

?**J'ai vu à la télé parler quelqu'un sans bouger les lèvres*

J'ai vu à la télé quelqu'un parler sans bouger les lèvres.

L'inacceptabilité en cas d'inversion de cet exemple découle dans notre modélisation de l'impossibilité de former un amas verbal en cas d'insertion d'un complément comme *à la télé*.

Notre traitement de l'inversion du sujet logique repose donc sur la formation d'un amas verbal. La formation d'un amas verbal est bloquée quand le verbe gouverneur possède un autre complément que l'infinitif car dans ce cas le « scrambling » est inacceptable :

J'ai dit à Pierre d'aller au même endroit que l'autre fois

??*J'ai dit d'aller à Pierre au même endroit que l'autre fois.*

¹⁶ Remarquons que *Je l'ai vu en sortir* est meilleur, ce qui est normal puisqu'il s'agit d'une instance de l'ordre « standard » (sans amas verbal).

¹⁷ Les conditions qui permettent la non réalisation du sujet logique reste à déterminer : **J'ai vu sortir des toilettes*, **J'ai entendu tomber*.

Les verbes du type *voir* ou *entendre*, de même que le causatif *faire*, sont donc considérés comme des verbes particuliers pouvant à la fois gouverner un complément nominal et former un amas verbal avec leur complément verbal.

Notons pour finir ce chapitre le contraste entre les deux phrases suivantes

Voilà les copies qu'a dit de corriger Marie à son assistant

Voilà le poly qu'a proposé de distribuer Marie à ses étudiants.

Dans la première phrase, le complément d'objet indirect *à son assistant* dépend du verbe recteur rendant difficile la formation de l'amas verbal et donc l'inversion du sujet (même s'il n'y a pas ici de « scrambling »), tandis que, dans la deuxième, le complément d'objet indirect *à ses étudiants* dépend de l'infinitif, ce qui ne bloque ni la formation de l'amas verbal, ni l'inversion du sujet et le « scrambling ».

6. Extraction et amas verbal

Nous avons vu que notre modélisation de l'inversion du sujet nécessitait la création d'un amas verbal et que cette même inversion était déclenchée par une extraction. Nous allons maintenant voir quels sont les liens entre l'extraction et la formation d'un amas verbal.

La première chose que nous pouvons remarquer, c'est que l'extraction ne nécessite pas la formation d'un amas verbal et que l'extraction peut être possible dans des situations où l'inversion ne l'est pas, comme par exemple avec une complétive à sujet non clitique ou avec un verbe recteur avec un complément nominal :

*une réunion que **mon collègue Pierre** veut absolument que le directeur organise*

*?*une réunion que veut absolument que le directeur organise mon collègue Pierre*

*une réunion à laquelle **mon collègue Pierre** menace le directeur de ne pas participer*

*?*une réunion à laquelle menace le directeur de ne pas participer mon collègue Pierre*

L'existence d'un groupe extrait qui contient un complémenteur et joue en même temps un rôle valentiel dans la proposition qu'il introduit concerne l'interface sémantique-syntaxe. Il existe de très nombreux travaux sur la modélisation de l'extraction, qui reste la pierre d'achoppement de la quasi-totalité des modèles proposés depuis les années 1970. Comme on le sait, l'extraction obéit à diverses contraintes, appelées traditionnellement *contraintes d'îlots* (Ross 1967). Kahane 2000 propose une version de ces contraintes en termes de chaînes de verbes et de tournures verbales, appelées *nucléus verbaux* : l'élément extrait dépend d'un nucléus verbal qui contient le verbe principal de la proposition qu'il introduit (cette chaîne est soulignée dans les exemples précédents). Les nucléus verbaux, qui s'apparentent aux amas verbaux, sont des objets de l'interface sémantique-syntaxe. Le nucléus verbal ne correspond pas nécessairement à une portion continue de la chaîne parlée (un syntagme), mais il aura néanmoins tendance à être le plus compact possible et tant que faire se peut à former un amas verbal (même si cela n'est pas toujours possible).

Reprenons le problème sous l'angle de l'ordre des mots (l'interface syntaxe-topologie) pour mieux expliquer pourquoi le nucléus verbal tend à former un amas verbal. Le placement du groupe extrait nécessite généralement une émancipation : en effet, le groupe extrait se place dans le champ complémenteur de la proposition qui est ouvert par le verbe principal de la proposition et ce verbe n'est pas nécessairement son gouverneur. Pour rejoindre le champ complémenteur, le groupe extrait devra donc « traverser » les frontières de tous les constituants ouverts par tous les éléments du nucléus verbal, c'est-à-dire par tous les verbes et éléments prédicatifs qui appartiennent à la chaîne de dépendance entre le verbe principal et le gouverneur du groupe extrait. Si ces éléments ont pu se regrouper au sein d'un amas verbal, les frontières à traverser seront donc uniquement des frontières d'amas verbaux enchâssés, c'est-à-dire des frontières relativement perméables (puisqu'elle peuvent même être traversées par des compléments ordinaires) et l'émancipation/extraction ne sera donc pas trop « coûteuse ». Par contre, si l'un des verbes du nucléus verbal n'a pas rejoint l'amas verbal de

son gouverneur et a ouvert son propre domaine, l'extraction, bien que toujours possible, devient beaucoup plus délicate. Ce deuxième type d'extraction se rencontre d'ailleurs très peu et essentiellement à l'écrit. On voit donc que l'extraction va favoriser (voire même forcer) la formation d'un amas verbal contenant l'ensemble du nucléus verbal et c'est cet amas verbal qui servira ensuite de pivot à une éventuelle inversion du sujet.

7. Les champs des domaines micro- et macro-syntaxique

Bien que la question des champs hors de l'amas verbal dépasse le sujet de notre étude, il nous a semblé difficile de ne pas placer la question de l'amas verbal dans une description complète de la topologie du verbe, qui restera néanmoins à l'état d'ébauche ici. Nous allons donc revenir sur les champs du domaine macro, puis sur ceux du domaine micro.

Les champs préfixe et postfixe du domaine macro accueillent avant tout les éléments disloqués. La dislocation fait apparaître un élément détaché qui n'est en relation avec le noyau de la phrase que par la coréférence avec un pronom, la présence d'un tel pronom n'étant d'ailleurs pas toujours avérée¹⁸ (ce pronom est parfois appelé résomptif, mais nous éviterons au maximum ce terme qui laisse à penser que le pronom fait écho à l'élément détaché, alors qu'il ne nous semble pas clair que la relation soit orientée de cette façon, notamment dans la dislocation droite, où c'est plutôt l'élément détaché qui semble faire écho à un pronom du noyau). Il est important, à la suite de Cinque 1990, de distinguer deux types de dislocation :

- la *dislocation liée*, où l'élément détaché est substituable au pronom du noyau avec lequel il coréfère (*De ça, Pierre en parle dans sa thèse*) ; en quelque sorte, l'élément détaché répète le régime porté par le pronom et la phrase fait apparaître un double marquage rectionnel ;
- la *dislocation libre*, où l'élément détaché ne porte aucune marque de rection (*Ca, Pierre en parle dans sa thèse*).

Nous préférons les termes lié/libre aux termes clitic/hanging topic proposés par Cinque 1990 ou construit/non construit proposés par Blasco-Dulbecco 1999. Blasco-Dulbecco parle également d'adjonction sur le pronom pour la dislocation libre et d'adjonction sur le noyau pour la dislocation liée. Cette « métaphore » nous semble assez pertinente et une analyse sémantico-syntaxique des disloquées (que nous ne ferons pas) devrait en tenir compte.

Les dislocations du sujet et de l'objet direct (**l'agneau, il l'a mangé, le loup**) pose problème pour cette classification : en l'absence d'un marquage rectionnel de ces positions en français, ces constructions restent ambiguës entre dislocation liée et dislocation libre. Sont également ambiguës les constructions détachées avec un pronom régi par une préposition (*Pierre, tu peux pas compter sur lui*). Seules ne sont pas ambiguës les dislocations mettant en jeu un clitique datif *lui/leur* clitique, le clitique génitif *en* et le locatif *y*.

¹⁸ Blasco-Dulbecco 1999 donne quelques exemples attestés de dislocation sans pronom « résomptif ». En voici 3 parmi 7 relevés sur un corpus de 1500 dislocations à l'oral (corpus du Gars, Université d'Aix-en-Provence) :

L1 *donc je pense que pour le reste il n'y a pas vraiment de problèmes*

L2 *et les ustensiles quand même parce que mettre une poêle au four il faut quand même un four particulier une poêle particulière* (Ricci 9, 15)

les radios euh je les ai vues euh le même le lendemain quoi et elles montraient un tassement de vertèbres – de quatre centimè- de qu- de quatre vertèbres pardon – et le nez je m'en suis tiré avec la déviation – de la cloison nasale (Vallier 10, 1)

et les lycées techniques euh tu arrives avec un bac de technicien (Charnay 4, 14).

La distinction entre dislocation liée et libre est importante pour la modélisation topologique. Premièrement, la dislocation droite est toujours une dislocation liée¹⁹ (à noter que cette formulation nous oblige à considérer les détachés sans marquage rectionnel apparent comme pouvant être des disloqués liés) :

Il en a souvent parlé, de ce voyage

**Il en a souvent parlé, ce voyage*

Je ne lui parle pas souvent, à Marie

**Je ne lui parle pas souvent, Marie.*

Deuxièmement, la dislocation liée est sensible aux contraintes d'îlot :

Marie, le type qui lui a parlé, il est suisse

??A Marie, le type qui lui a parlé, il est suisse

Marie, le type qu'elle veut épouser est suisse

**Le type qu'elle veut épouser est suisse, Marie²⁰*

Ces deux propriétés entraînent que les règles de placement des disloquées liées et libres seront très différentes. Les disloquées libres pourront accéder au seul champ préfixe et les règles permettront n'importe quelle émancipation. Les disloquées liées pourront accéder aux champs préfixe et postfixe, mais les règles d'émancipation devront être contraintes afin de vérifier les contraintes d'îlots, à moins que ces contraintes ne soient assumées par l'interface sémantique-syntaxe, ce qui serait plus satisfaisant puisque c'est dans l'interface sémantique-syntaxe que les régimes sont imposés et que c'est justement le marquage rectionnel qui distingue disloquées liées et libres. Quoiqu'il en soit, ces questions restent ouvertes, comme le sont la question des représentations syntaxique et sémantique de ces constructions (cf. Deulofeu 2003).

Certains auteurs, comme Larsson 1979, Cinque 1990, Delais-Roussarie et al. à paraître, mentionnent également des contraintes d'ordre entre les éléments détachés, les disloqués libres devant précéder les disloqués liés. On a ainsi les contrastes suivants :

Mais les adultes, son secret, elle le leur cache

**Mais son secret, les adultes, elle le leur cache*

Mais son secret, aux adultes, elle le leur cache

Ces contraintes d'ordre sont faciles à exprimer dans le cadre du modèle topologique : il suffit de considérer deux champs distincts pour les deux types de disloqués. Néanmoins, ces contraintes nous semblent trop fragiles pour être prise en compte, d'autant que l'opposition libre/lié n'est certainement pas le seul facteur à jouer ici.

difficiles à évaluer, si elles étaient confirmées, nous amèneraient à considérer deux champs différents pour les détachés à gauche.

¹⁹ On observe quelques rares cas d'éléments libres détachés à droite. Néanmoins sur un corpus de 1500 disloquées orales, Blasco-Dulbecco n'en relève que 7 et toutes concernent des pronoms *moi* et *toi* (*non quand on t'emmerde ça te plaît pas toi* ; *si on me demandait de faire partie d'un comité moi pour juger d'autres chanteurs je le ferais volontiers*). On peut en effet un contraste net entre les pronoms de 1^{ère} et 2^{ème} personne et les autres éléments :

Elle me parle encore moi

**Elle lui parle encore Pierre/lui*

On peut donc estimer qu'il s'agit encore d'un autre cas d'éléments détachés, des éléments d'interpellation destinés au locuteur et à ses interlocuteurs.

²⁰ Etant admis que seule la dislocation liée est possible à droite, nous considérons ici, malgré l'impossibilité de contrôler le marquage rectionnel, qu'il s'agit bien d'une dislocation liée.

La topicalisation pose également des questions intéressantes. La topicalisation se distingue de la dislocation par l'absence de pronom résomptif. Elle se caractérise par l'antéposition d'un élément valent et obéit, comme les extractions avec mots *qu* et comme la dislocation liée, aux contraintes d'îlots :

A Marie, je ne parle plus

**A Marie, je ne connais personne qui ne parle plus*

**A Marie, les gens qui ne parlent plus ne sont pas nombreux*

Venons-en au point qui nous intéresse. Il existe probablement deux placements différents de l'élément topicalisé. Comparons les énoncés suivants :

Sur la place se dresse une cathédrale

Sur la place, une cathédrale se dresse.

Dans les deux cas *sur la place* est topicalisé, mais on observe une différence de prosodie importante : dans le deuxième cas, l'élément topicalisé est nettement détaché, ce qu'enregistre d'ailleurs l'écrit avec l'usage de la virgule. Dans le premier cas, l'élément topicalisé pourrait tout simplement occuper le champ sujet. La confusion avec le sujet est d'ailleurs assez nette dans une construction figée comme *Peu importe (que tu ne sois pas là)*, où le sujet véritable devient même optionnel, ce qui est tout à fait inhabituel en français. Dans le deuxième cas, l'élément topicalisé est détaché et il rejoindrait donc le champ préfixe du domaine macro. Ceci est confirmé par une relative liberté d'ordre entre disloqué et topicalisé (dans la mesure où la coexistence des deux est acceptable) :

A Jeanne, Pierre, il ne parle plus

Pierre, à Jeanne, il ne parle plus.

Comme on le voit, il est probable que l'élément topicalisé puisse choisir entre domaine micro et macro et qu'il ne soit pas nécessaire de considérer un champ propre pour la topicalisation.

Pour revenir au domaine micro, nous aurions donc bien un unique champ devant le verbe pouvant accueillir soit le sujet, soit un élément topicalisé, le sujet devant alors être inversé. La question de savoir s'il y a plusieurs champs après l'amas verbal est plus délicate. Pour rendre compte d'un ordre relativement rigide comme dans

Marie a forcé Pierre à venir

**Marie a forcé à venir Pierre*

il faudrait probablement considérer deux champs compléments. Il en va de même pour les inversions de sujet élaboratives (que nous n'avons pas étudiées ici) pour lesquelles le sujet doit toujours être à la frontière droite du domaine micro (Marandin 2003) :

Iront en salle C tous les élèves punis

**Iront tous les élèves punis en salle C.*

Nous n'irons pas plus loin dans la description des domaines micro et macro.

8. Conclusion

Ce travail est une première ébauche d'une modélisation de l'ordre des mots en français dans le cadre du modèle topologique, couvrant, de manière parfois un peu grossière, les principaux problèmes connus : le placement des clitiques, l'inversion du sujet, l'extraction, la dislocation, etc. La présente description est restée quelque peu informelle, mais elle s'appuie sur un formalisme mathématique complet (Gerdes & Kahane 2001) implanté par K. Gerdes (<http://talana.linguist.jussieu.fr/~kim/deplin>). Nous n'avons pas non plus cherché à justifier l'approche topologique, ni à mener des comparaisons avec d'autres approches. A ce sujet, on pourra consulter Gerdes & Kahane, sous presse.

Notre étude s'est centrée sur l'ordre des dépendants directs des verbes, c'est-à-dire sur la topologie du verbe, mettant en évidence la nécessité de considérer un constituant verbal particulier que nous avons appelé l'amas verbal. Une étude similaire sur la topologie du nom

mettraient en évidence des parallèles intéressants, notamment sur la constitution d'amas nominaux du type «N de N».

Notre étude s'est également limitée à l'étude de l'interface syntaxe-topologie (la correspondance entre arbres de dépendance syntaxiques et structures de constituants topologiques). Ce module s'inscrit dans un modèle complet de la langue comprenant également une interface sémantique-syntaxe et une interface topologie-phonologie. Cela pour dire que les structures syntaxiques et topologiques que nous considérons ne peuvent se justifier du seul point de vue de l'interface syntaxe-topologie, mais doivent être envisagées comme des structures intermédiaires entre une représentation sémantique de la phrase (incluant une structuration informationnelle/communicative) et une représentation phonologique (incluant une représentation de la prosodie et une structuration en constituants phonologiques) (cf. Gerdes & Yoo 2003, Yoo 2003 sur ce dernier point). Une validation nécessaire de notre proposition d'amas verbal doit être justement de montrer que celui-ci tend à être un constituant phonologique²¹.

Enfin, notre étude s'inscrit dans une modélisation de l'ordre des mots dans la diversité des langues. Le modèle topologique vient de l'étude de l'ordre des mots dans les langues germaniques, qui pose des problèmes importants en raison d'une tension entre une grande liberté des mots (scrambling, extraposition, ...) et des contraintes syntaxiques fortes (verbe final ou en deuxième position). Le modèle a également été appliqué à la description de langues à ordre plus libre comme le grec (Yoo 2003) ou moins libre comme l'arabe (El Kassas & Kahane 2004). L'application à des langues à ordre assez rigide comme le français restait à faire.

Références

- Abney Steven, 1991, Parsing by chunks, in R. Berwick, S. Abney and C. Tenny (eds.), *Principle-Based Parsing*, Kluwer.
- Ades, A. & Steedman, M. J., 1982, On the Order of Words. *Linguistics and Philosophy*, 44, 517-518.
- Berrendoner Alain, 1991, Pour une macro-syntaxe, in Dominique Willens (éd), *Données orales et théories linguistiques*, Duculot, Paris/Louvain, pp. 25-31.
- Blanche-Benveniste Claire, 1990, *Le français parlé : études grammaticales*, avec Mireille Bilger, Christine Rouget et Karel van den Eynde, Editions du CNRS, Paris.
- Blasco-Dulbeco Mylène, 1999, *Les dislocations en français contemporain. Etude syntaxique*, Honoré Champion, Paris.
- Bonami Olivier, Gilles Boyé & Jesse Tseng, 2004, An Integrated Analysis of French Liaison, in G. Jaeger (éd.), *Formal Grammar 2004 Preproceedings*, Nancy, France.
- Cinque G., 1990, *Types of A-bar dependencies*, Linguistic Inquiry Monograph, MIT Press, Cambridge.
- Delais-Roussarie E., J. Doetjes & P. Sleeman, à paraître, Dislocations in French. In *Handbook of French semantics*, F. Corblin & H. de Swart (éds). CSLI, 447-479.

²¹ Les phénomènes de liaison (*Pierre a bien_aimé ça ; Pierre ne fera pas_écouter le début ; Pierre ne cherche plus_à comprendre*) pourraient également servir de justification à la formation d'un amas verbal, même si clairement la liaison ne met pas en jeu la seule interface topologie-phonologie (Bonami *et al.* 2004).

- Deulofeu José, 2003, L'approche macrosyntaxique en syntaxe : un nouveau rasoir d'Occam contre les notions inutiles, in Gagean N. (éd.), *Rencontres linguistiques en Pays Rhénan -12, Scolia n° 16*, Publications de l'Université de Strasbourg 2, Strasbourg.
- El Kassas Dina & Sylvain Kahane, 2004, Modélisation de l'ordre des mots en arabe standard, Atelier sur le Traitement Automatique de la Langue Arabe Ecrite et Parlée, *JEP-TALN*, Fès, 259-264.
- Gerdes Kim, 2002, *Topologie et grammaires formelles de l'allemand*, Thèse de doctorat, Université Paris 7.
- Gerdes Kim & Sylvain Kahane, 2001, Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy, *ACL*, Toulouse, 8 p.
- Gerdes Kim & Sylvain Kahane, sous presse, "Phrasing It Differently", in L. Wanner (ed.), *Papers in Meaning-Text Theory in honour of Igor Mel'cuk*, Benjamins, 34 p.
- Gerdes Kim & Hi-Yon Yoo, 2003, *La topologie comme interface entre syntaxe et prosodie : un système de génération appliqué au grec moderne Actes de TALN 2003*, Batz-sur-Mer.
- Gladkij Aleksej V., 1966, *Leckii po matematicheskoj linguistike dlja studentov NGU*, Novosibirsk (trad. : *Leçons de linguistique mathématique*, fasc. 1, 1970, Dunod).
- Kahane Sylvain, 2000, "Extractions dans une grammaire de dépendance lexicalisée à bulles", *T.A.L.*, 41:1, 211-243.
- Larsson Eva, 1979, *La dislocation en français. Etude de syntaxe générative*, Gleerup, Lund.
- Lecerf Yves, 1961, "Une représentation algébrique de la structure des phrases dans diverses langues naturelles", *C. R. Acad. Sc. Paris*, 252, 232-34.
- Marandin Jean-Marie. 2003. «Inversion du sujet et structure de l'information dans les langues romanes», in Danièle Godard (éd), *Langues romanes. Problèmes de la phrase simple*, Paris: Editions du CNRS. Version longue à <http://www.llf.cnrs.fr/fr/Marandin>.
- Miller, P. and Sag, I. 1997 "French Clitic Mouvement without Clitics or Movement", *Natural Language and Linguistic Theory* 15.3, pp. 573-639.
- Ross John, 1967, *Constraints on Variables in Syntax*, PhD Thesis, MIT (publié sous le titre *Infinite Syntax*, Reidel, Dordrecht, 1985).
- Vergne Jacques, 2000, *Étude et modélisation de la syntaxe des langues à l'aide de l'ordinateur - Analyse syntaxique automatique non combinatoire*, Thèse d'HDR, Université de Caen.
- Yoo 2003, *Ordre des mots et prosodie : Essai de description et de formalisation pour le grec moderne*, thèse de doctorat, Université Paris 7.