

Le rôle des structures et représentations dans l'évolution des théories syntaxiques

Sylvain Kahane

Sciences du langage

Modyco, Université Paris Ouest Nanterre La Défense

Résumé

Cet article compare deux modes de représentation de la structure syntaxique des énoncés en langue naturelle : les structures de constituants, promues par les travaux du linguiste américain Noam Chomsky (1957), et les structures de dépendance, définies par le linguiste français Lucien Tesnière (1959). Nous intéresserons aux conséquences que peut avoir le choix d'un mode de représentation plutôt que l'autre sur l'évolution des théories linguistiques. Nous montrerons l'avantage qu'il peut y avoir à utiliser les structures de dépendance pour modéliser l'analyse comme la production d'énoncés et nous nous interrogerons sur les raisons pour lesquelles les structures de constituants leur ont été préférées.

1. Introduction

Il existe deux modes de représentation de l'organisation syntaxique d'un énoncé : la structure de dépendance et la structure de constituants. Ces deux modes de représentation sont à peu près équivalents notamment pour des énoncés simples, mais ils se prêtent à des interprétations différentes sur le fonctionnement de la syntaxe des langues et offrent des extensions de nature assez différentes. Dans cet article, nous souhaitons montrer que le choix de la structure de constituants, au détriment de la structure de dépendance, comme structure de représentation syntaxique n'a pas été sans conséquence sur l'évolution des théories syntaxiques.

Cet article aborde donc la question de l'évolution sous un angle très particulier. Il ne s'agit pas d'étudier comment les langues et leur syntaxe évoluent, question par ailleurs passionnante, mais comment le choix d'un type de représentation peut influencer de manière significative l'évolution des théories et donc de tous les travaux d'une discipline, ici la linguistique.

Il est à noter que cette présentation est tout sauf objective. L'auteur est un spécialiste de la syntaxe de dépendance, alors que les théories dominantes depuis 50 ans utilisent des représentations syntaxiques en constituants. On observe aujourd'hui un retour important de la syntaxe de dépendance, notamment à travers le traitement automatique des langues où plusieurs analyseurs automatiques, pour le français ou l'anglais, ont montré des performances supérieures à des systèmes basés sur la syntaxe de constituants. Cet article souhaite montrer ce qui différencie vraiment les deux modes de représentation et comment les choix de représentation actuels d'un grand nombre de linguistes sont davantage dus à des habitudes transmises qu'à une décision raisonnée et scientifique.

2. Représenter la structure d'un calcul

Nous allons commencer cet exposé par l'étude de la structure d'un calcul algébrique élémentaire. Ceci nous permettra de montrer que le problème du choix de la représentation syntaxique n'est pas une question de linguistique et d'apprécier, hors des débats internes à cette discipline, les qualités des différentes représentations possibles.

Considérons la formule $(7 \times 19) + (5 \times 31)$. Pour effectuer ce calcul, il faut nécessairement effectuer les deux multiplications avant de les sommer, ce que l'on peut représenter par le diagramme suivant :

$$\begin{array}{r} (7 \times 19) + (5 \times 31) \\ \hline 133 \quad 155 \\ \hline 288 \end{array}$$

Figure 1. Diagramme associé à un calcul

On voit que chaque calcul intermédiaire met en jeu deux nombres et un opérateur. La structure du calcul peut donc être représentée de la manière suivante :

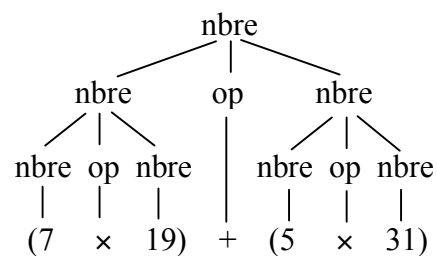


Figure 2. Structure de constituants d'un calcul

Cette structure est appelée un *arbre de constituants*. Les éléments de la formule (nous omettons les parenthèses) sont les feuilles de cet arbre et les nœuds internes représentent les constituants de cet formule. Chaque nœud interne reçoit une étiquette indiquant la nature de ce constituant, selon qu'il s'agit d'un nombre ou d'un opérateur ; à l'exception des opérateurs, chaque constituant correspond à un nombre, soit directement, soit comme résultat d'un calcul. L'arbre de constituants de la Fig. 2 montre bien comment s'effectue le calcul, comme le montre sa similarité avec diagramme de la Fig. 1 : on passe de l'arbre de constituants à ce diagramme en remplaçant les étiquettes nombre par leur valeur et les nœuds opérateur par les traits horizontaux modélisant les calculs.

On peut représenter de la même façon un calcul où intervient un opérateur unaire comme la racine carrée dans $1 + \sqrt{5}$:

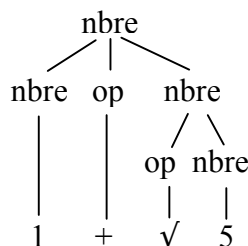


Figure 3. Structure de constituants d'un calcul

En considérant ce calcul, on voit qu'il y a des constituants nombre qui se décomposent respectivement en deux ou trois sous-constituants. L'élément qui décide de la structure du constituant est l'opérateur, selon qu'il est unaire, comme $\sqrt{\quad}$, ou binaire, comme $+$ et \times .

Nous appelons *tête* l'élément qui contrôle la structure interne d'un constituant. On dit qu'un élément X *gouverne* les sous-constituants du constituant dont X est la tête et inversement qu'un constituant Y *dépend* d'un élément X si Y est un sous-constituant du constituant dont X est la tête.

On peut alors adopter une autre représentation arborescente des calculs en remplaçant chaque constituant par sa tête et en plaçant ainsi chaque élément sous son gouverneur :

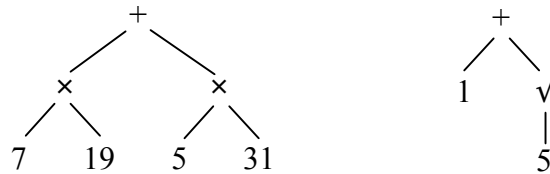


Figure 4. Structure de dépendance des deux calculs

Cette structure est appelée un *arbre de dépendance*.

On peut reconstruire facilement l'arbre de constituants à partir de l'arbre de dépendance. En effet, les constituants sont toujours apparents dans la structure de dépendance : ce sont les projections des sous-arbres de l'arbre de dépendance. Un *sous-arbre* est l'arbre formé par tous les nœuds qui sont sous un nœud donné ; la *projection* d'un sous-arbre est le segment de formule formé par les éléments qui étiquettent ce sous-arbre. La figure suivante montre comment, en étiquetant par leur catégorie les nœuds et les sous-arbres de l'arbre de dépendance, on retrouve l'arbre de constituants, les relations hiérarchiques entre constituants correspondant simplement à l'inclusion des constituants (cf. Kahane 1997 pour une formalisation).

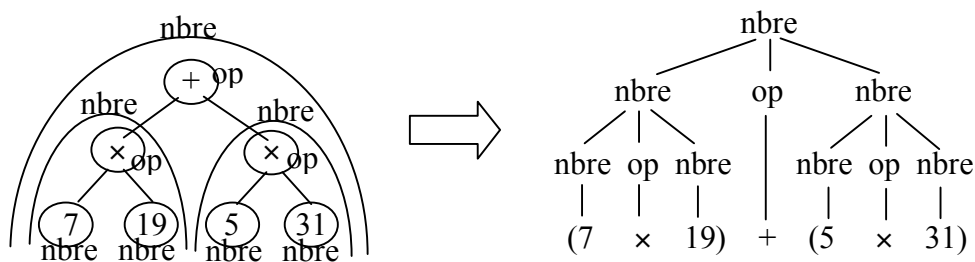


Figure 5. Passage d'une structure de dépendance à une structure de constituants

Par contre, comme nous l'avons vu, pour passer d'un arbre de constituants à un arbre de dépendance, il est nécessaire d'ajouter l'information concernant la tête de chaque constituant. Ainsi un arbre de dépendance, tout en étant plus simple (le nombre de nœuds est deux fois moindre), contient-il plus d'information qu'un arbre de constituants¹.

¹ Certains argueront que l'on peut considérer des arbres de constituants plus riches en augmentant le nombre de constituants intermédiaires et que de telles structures ne sont plus équivalentes à un arbre de dépendance. Mais on peut tout aussi bien enrichir l'arbre de dépendance en ordonnant en fonction de leur saillance les différents dépendants d'un nœud et considérer ainsi des projections intermédiaires contenant seulement les nœuds les plus saillants.

Nous nous sommes intéressés jusque-là uniquement à la structure hiérarchique et, bien que nous ne l'ayons pas spécifié, les arbres que nous considérons étaient non ordonnés². Un arbre est dit *ordonné* si les dépendants de chaque nœud sont linéairement ordonnés entre eux, ce qui induit un ordre total sur les feuilles de l'arbre. Pour imaginer un arbre non ordonné, il faut d'ailleurs se le représenter dans l'espace comme un mobile suspendu par la racine, les dépendants pouvant tourner autour de leur gouverneur. Les arbres de constituants sont généralement considérés comme plus intéressants pour le traitement de l'ordre linéaire, car ils sont naturellement ordonnés par la représentation planaire. Mais, comme nous allons le voir, il est d'une certaine façon plus avantageux de traiter les questions d'ordre linéaire à partir d'un arbre de dépendance que d'un arbre de constituants, notamment lorsque l'ordre linéaire ne découle pas immédiatement de la structure hiérarchique.

Avant de passer aux énoncés en langue naturelle, rappelons que les structures de dépendance pour les formules ont été introduites à la suite des travaux du logicien polonais Jan Łukasiewicz (1930), qui a montré qu'il y avait plusieurs façons d'encoder linéairement un tel arbre. Une première façon est celle que nous utilisons quand nous écrivons la formule $(7 \times 19) + (5 \times 31)$, où à chaque fois l'opérateur binaire a été placé entre ses deux dépendants. Cette écriture nécessite des parenthèses pour ne pas être ambiguë. Une autre écriture, dite *écriture polonaise* ou *préfixée*, consiste à lire l'arbre en ramassant toujours le gouverneur avant ses dépendants et les dépendants de gauche à droite : $+ \times 7 19 \times 5 31$. Cette formule n'est pas ambiguë, malgré l'absence de parenthèses, dès qu'on connaît l'arité de chaque symbole. L'*arité* d'un symbole est le nombre d'arguments et donc de dépendants qu'il a. La *lecture postfixée* ou *polonaise inverse*, où le gouverneur est ramassé après ses dépendants, est particulièrement appropriée au calcul et utilisée par les calculateurs automatiques, y compris certaines machines à calculer d'usage courant : en effet, pour effectuer le calcul $7 19 \times 5 31 + \times$, il suffit d'empiler les nombres au fur et à mesure de la lecture et à la lecture de chaque opérateur binaire d'effectuer l'opération sur les nombres des deux lignes qui précèdent en les supprimant :

7			
19			
×	133	133	
	5		
	31		
	×	155	
		+	288

Figure 6. Calcul à partir d'une écriture postfixée

Ces différentes stratégies pour passer d'une structure hiérarchique à un ordre linéaire se retrouvent dans les langues naturelles et l'on peut considérer que la grammaire des langues naît justement du besoin d'encoder linéairement (car le signal sonore que nous produisons en parlant est linéaire) une information qui est par essence non linéaire.

² Traditionnellement, les arbres de dépendance sont considérés comme non ordonnés, comme c'est le cas chez Tesnière, tandis que les arbres de constituants sont ordonnés, comme dans les travaux de Chomsky (1957, 1965, 1993). Mais depuis, le début des années 1980, des linguistes comme Gerald Gazdar ont montré l'avantage qu'il y avait, même en grammaire de constituants, à distinguer la hiérarchie syntaxique ou constituance immédiate de l'ordre linéaire. Les modèles syntagmatiques actuels, comme HPSG (Pollard & Sag 1994), manipulent séparément les règles de constituance immédiate et les règles de précédence linéaire.

3. Structure syntaxique d'un énoncé

Considérons maintenant un énoncé en langue naturelle :

Quelques étudiants motivés organisent une conférence internationale.

La structure syntaxique³ de cet énoncé est très similaire à celle du premier calcul que nous avons étudié. Commençons par présenter sa structure en constituants, qu'on appelle plus communément, en linguistique, *structure syntagmatique*. Cet énoncé est constitué d'une forme verbale — *organisent* — accompagné de deux groupes nominaux qui sont respectivement son sujet — *quelques étudiants motivés* — et son complément d'objet direct — *une conférence internationale*. Chacun de ces deux groupes nominaux est constitué d'un déterminant, un nom et un adjectif. On obtient donc l'arbre de constituants⁴ suivant :

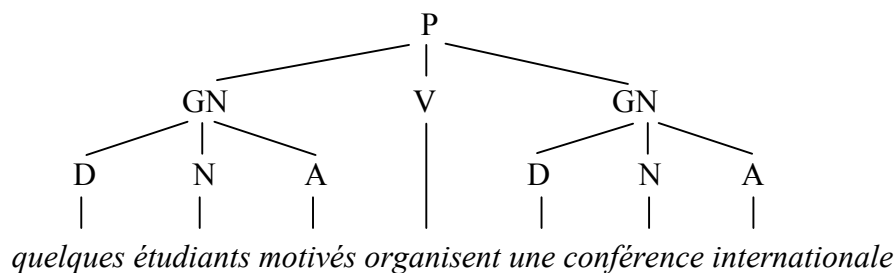


Figure 7. Arbre de constituants d'un énoncé

On peut maintenant introduire l'arbre de dépendance de cet énoncé. La définition de la tête est ici plus complexe. On peut en envisager deux définitions qui, bien que différentes, coïncident la plupart du temps. On appelle *tête interne* d'un constituant le mot qui contrôle la structure interne du constituant et *tête externe* d'un constituant le mot qui contrôle la distribution du syntagme, c'est-à-dire les environnements dans lesquels le constituant peut être enchâssé. Il y a de bons arguments pour montrer que la forme verbale⁵ est à la fois la tête interne et externe d'une proposition (Kahane 2001). D'une part, c'est la forme verbale qui contrôle la présence d'un sujet et d'un objet direct : ainsi certaines formes verbales, comme l'impératif ou

³ Définir précisément ce que nous appelons *structure syntaxique* nous amènerait trop loin ici. Ce que l'on peut dire, c'est que les énoncés possèdent, comme nous le verrons plus loin, une structure hiérarchique, dans le sens où les caractéristiques syntaxiques d'un mot (partie du discours, marque de rection, marque d'accord et place linéaire) sont généralement contrôlées par un autre mot et ceci jusqu'à remonter à l'élément principal de la phrase, qui est, en français, la plupart du temps un verbe à l'indicatif. Nous appelons *structure syntaxique de surface* cette structure hiérarchique sur les mots et c'est ce qui nous intéresse ici. Nous ne nous intéressons pas à l'ordre des mots pour l'instant et nous considérons qu'il n'est pas encodé dans la structure syntaxique, qui est donc un arbre non ordonné. L'ordre des mots, comme la prosodie et le découpage en constituants prosodiques, sont encodés dans une autre structure, la *structure topologique*, et les règles d'ordre des mots sont traitées comme des règles de correspondance entre ces deux niveaux structurels (Gerdes & Kahane 2006).

⁴ Les arbres de constituants considérés par les linguistes qui travaillent en grammaire syntagmatique sont généralement moins plats. Elles comprennent notamment une bipartition de la phrase en un GN sujet et un groupe verbal qui inclut le verbe et ses compléments. Nous ne discuterons pas ici des motivations d'une telle bipartition, qui a été de nombreuses fois contestée, notamment pour le français (Tesnière 1959, Abeillé 1996-97, Kahane 2001).

⁵ Nous prenons bien soin de distinguer la forme verbale du verbe lui-même : la *forme verbale* est formée d'un verbe ou *lexème verbal* et d'une *flexion*. Les deux sont indissociables et forment ensemble un mot : il n'y a donc pas lieu de les séparer pour une analyse grossière comme nous faisons ici. Si l'on souhaite néanmoins les séparer, on doit alors considérer que c'est plutôt la flexion et non le verbe qui est la tête externe de la proposition.

l'infinitif ne se combine pas avec un sujet, tandis que de nombreux verbes, comme *dormir* ou *parler*, n'ont pas de complément d'objet direct. D'autre part, c'est la forme verbale qui fait qu'une proposition est une proposition et peut par exemple s'enchâsser dans l'environnement *je crois que* ___ :

Je crois que [quelques étudiants motivés organisent une conférence internationale].

C'est également le verbe qui hérite des marques morphosyntaxiques imposées par le gouverneur de la proposition, par exemple lorsque la proposition est gouvernée par un verbe qui exige le subjonctif : *Je crois que [Pierre vient]* vs *Il faut que [Pierre vienne]*. On peut de même considérer que le nom est la tête du groupe nominal, même s'il existe de bons arguments pour considérer que le déterminant contrôle une partie de la distribution du groupe nominal⁶. On obtient alors l'arbre de dépendance suivant :

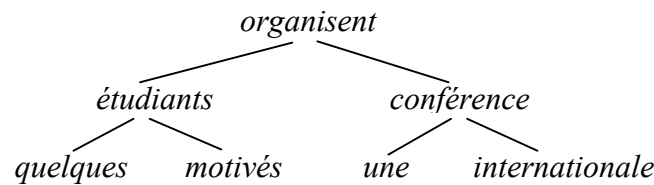


Figure 6. Arbre de dépendance d'un énoncé

On ajoute généralement à cette représentation un étiquetage des dépendances indiquant la fonction de chaque nœud par rapport à son gouverneur : *sujet, cod, déterminant, épithète*, etc. Comme précédemment l'arbre de constituants et l'arbre de dépendance sont à peu près équivalents : on passe de l'arbre de constituants à l'arbre de dépendance en décidant pour chaque constituant quelle est sa tête et en agrégeant chaque mot avec tous les constituants dont il est la tête :

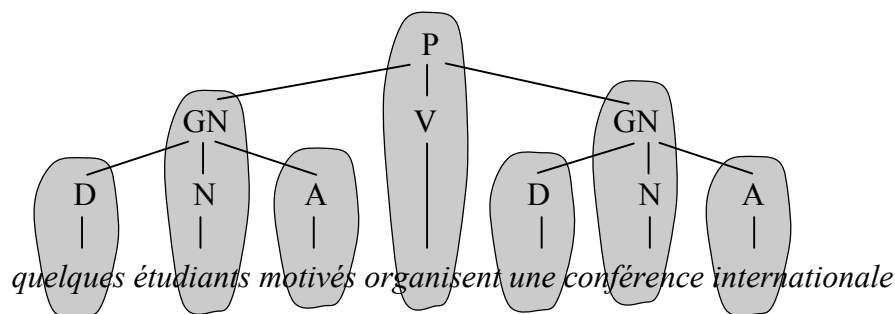


Figure 7. Passage d'un arbre de constituants à un arbre de dépendance

Inversement, on passe de l'arbre de dépendance à un arbre de constituants non ordonné en considérant la hiérarchie d'inclusion entre les projections de tous les sous-arbres de l'arbre de dépendance (voir Fig. 5 pour une illustration avec une formule algébrique).

L'information portée par les deux types de structure — l'arbre de dépendance et l'arbre de constituants avec têtes — étant équivalente⁷ (à l'étiquetage près)⁸, voyons quels sont les fac-

⁶ Cf. par exemple l'alternance d'articles selon que le groupe nominal complément d'objet direct est dans une assertion ou une négation : *Je veux [du poulet]* vs *Je ne veux pas [de poulet]*.

⁷ Nous parlons ici d'équivalence au sens mathématique du terme : deux familles de structures sont équivalentes s'il existe une correspondance bi-univoque entre les deux familles qui peut être définie algorithmiquement. C'est le cas ici : nous avons montré comment associer à chaque arbre de constituants plat avec têtes un arbre de dépendance et inversement comment associer à chaque arbre de dépendance un arbre de constituants plat avec

teurs qui ont présidé au choix d'un mode de représentation plutôt que l'autre. Cet article n'est pas le premier à aborder cette question. Plusieurs linguistes, et en tout premier lieu Richard Hudson (1980a,b) et Igor Mel'čuk (1988, 2003), ont défendu l'utilisation d'arbres de dépendance pour la représentation syntaxique. Cet article s'en distingue notamment en posant comme préalable que les deux modes de représentation sont en partie équivalents et que ce n'est que lorsqu'on manipule les structures (en écrivant une grammaire) ou que l'on étend les modèles (en proposant des structures plus complexes) que l'on fait apparaître des différences dans les deux modes de représentation.

4. Choisir une représentation

L'analyse en constituants apparaît dans l'ouvrage de 1933 de Leonard Bloomfield, *Language*, considéré à juste titre comme l'ouvrage fondateur de la linguistique structurale américaine. Mais ce sont surtout les travaux du linguiste américain Noam Chomsky, de son premier ouvrage *Syntactic Structures* en 1957 jusqu'au *Minimalist Program* de 1993 en passant par les travaux sur la syntaxe X-barre des années 1970, qui vont imposer les structures de constituants. La grande force de ces travaux est d'associer dès le départ un modèle mathématique très simple au calcul des structures de constituants. Il s'agit des grammaires de réécriture, que vont également s'approprier les « informaticiens » qui vont développer les premiers langages de programmation. Les grammaires de réécriture vont être à la base de tous les travaux en théorie des langages formels et en analyse syntaxique automatique, dans le domaine du traitement automatique comme dans celui des compilateurs de langages de programmation. On peut ainsi décrire l'arbre de constituants de la Fig. 7 par les deux règles syntagmatiques suivantes, $P \rightarrow GN V GN$ et $GN \rightarrow D A N$, indiquant comment se décompose les constituants P et GN.

Même si on peut probablement faire remonter l'apparition des structures de dépendance aux travaux des grammairiens arabes au 8^{ième} siècle et si on en trouve par exemple dans des grammaires scolaires anglaises au 19^{ième}, c'est au linguiste français Lucien Tesnière que l'on doit la première théorie linguistique entièrement basée sur des structures de dépendance. Cette théorie est présentée dans un premier article de 1934, puis dans un ouvrage posthume publié en 1959, *Éléments de syntaxe structurale*.

Que se passe-t-il alors ? Les travaux du jeune prodige américain Noam Chomsky vont avoir immédiatement un retentissement international. Celui-ci va en plus bénéficier des moyens du prestigieux MIT (*Massachusetts Institute of Technology*), où il va former de nombreux et brillants étudiants qui sont aujourd'hui encore les ténors de la linguistique formelle internationale, même si certains comme Ivan Sag, fondateur du modèle HPSG (*Head-driven Phrase Structure Grammar* ; Pollard & Sag 1994), ou Joan Bresnan, fondatrice du modèle LFG (*Lexical-Functional Grammar* ; Bresnan 2001), ont pris leur distance avec le maître⁹. De son

têtes. Il s'en suit que d'un strict point de vue mathématique, on peut considérer que arbre de dépendance et arbre de constituants sont deux façons de *représenter* ou de *définir* la même structure mathématique sous-jacente.

⁸ Ce point n'est d'ailleurs pas si secondaire. Car bien que les structures non étiquetées soient équivalentes (c'est-à-dire véhiculent la même information), elles n'en mettent pas moins en évidence des objets différents, à savoir les constituants pour les arbres de constituants et les dépendances pour les arbres de dépendance. Il s'en suit non seulement que les structures étiquetées ne sont pas équivalentes, mais que les règles symboliques qui vont manipuler ces structures vont aussi manipuler des objets étiquetés différents.

⁹ Les modèles HPSG et LFG, qui sont aujourd'hui considérés comme les principales alternatives aux modèles chomskyens, font un usage important de la notion de tête. Le H (pour *head-driven* 'guidé par les têtes') de HPSG renvoie directement au fait que le modèle prend fortement en compte le fait que la description d'un constituant

côté, Lucien Tesnière, connu comme militant communiste, n'a jamais pu accéder aux prestigieux ors de la Sorbonne et restera professeur à l'École d'Institutrices de Montpellier¹⁰. Il mourra en 1954 sans avoir pu achever son grand ouvrage suite à la guerre et à une santé fragilisée. Celui-ci, publié en 1959, connaît un certain succès, mais il paraît deux ans après *Syntactic Structure* de Noam Chomsky, et les jeunes linguistes français s'emploient à le ringardiser. Citons ainsi la conclusion d'un article de Michel Arrivé de 1969 consacré à l'ouvrage de Tesnière, très axé sur la comparaison avec la grammaire transformationnelle de Chomsky (1965) :

« Pour le linguiste d'aujourd'hui, la théorie syntaxique de Tesnière n'a plus qu'un intérêt historique. A cet égard, son ambiguïté même est le signe de sa richesse. Elaborés dans une relative solitude, les concepts essentiels de la *Syntaxe Structurale* sont restés pour une bonne part à l'écart des tendances linguistiques de leur temps. Il leur arrive aujourd'hui de rencontrer, par deux voies différentes, quelques-uns des thèmes centraux de la réflexion linguistique contemporaine. »

Malgré ce contexte, quelques chercheurs vont s'intéresser à la syntaxe de dépendance. On peut citer notamment le travail du mathématicien français Yves Lecerf (1961), qui le premier montrera l'équivalence entre arbres de constituants et arbres de dépendance et comparera l'expression de l'ordre linéaire pour les deux types de structure en montrant l'équivalence de la connexité des constituants avec la projectivité des arbres de dépendance (voir plus loin). Les linguistes américains, notamment Hays (1964) et Gaifman (1965), proposeront un modèle formel pour la syntaxe de dépendance et étudieront l'équivalence avec les grammaires de réécriture de Chomsky¹¹. Mais c'est à l'Est, à Prague avec Petr Sgall (1965) et le modèle FGD (*Functional Generative Description*) et à Moscou avec Igor Mel'čuk (1963, 1988, 1997) et la Théorie Sens-Texte, que la syntaxe de dépendance connaîtra un véritable essor et c'est par les chercheurs formés à l'Est, notamment Petr Sgall et Eva Hajičová de Prague (Sgall et al. 1986) et Igor Mel'čuk exilé à Montréal en 1978, que la syntaxe de dépendance reviendra en force à l'Ouest, lorsque les relations scientifiques entre les deux blocs redeviendront possibles.

Comme on le voit dans ce bref exposé historique, les principales raisons qui ont présidé à la suprématie d'un choix de représentation plutôt que de l'autre sont davantage politiques que

est très largement imposée par celle de sa tête, ce qui fait qu'on peut assez naturellement interpréter une grammaire HPSG comme une implémentation particulière de la grammaire de dépendance (Kahane 2008b). LFG, quant à lui, utilise explicitement une représentation en termes de dépendance couplée à une représentation plus traditionnelle en termes de constituants. Le courant chomskyen a lui-même mis en avant la notion de tête depuis l'introduction de la syntaxe X-barre, où chaque constituant est nommé de la même façon que sa tête en ajoutant à chaque fois une « barre ».

¹⁰ Citons à ce propos la conclusion de la préface de Jean Fourquet à la première édition des *Éléments de syntaxe structurale* : « Ce n'est pas sans une certaine amertume que l'on pense aux 25 ans d'avance qu'aurait pu avoir la France, si elle avait donné à Tesnière en 1934 [année de la parution du premier article présentant ses théories de la dépendance et de la translation] la direction d'un institut de linguistique appliquée comme celui d'Edimbourg [attribué à cette époque au dialectologue et structuraliste Catford]. »

¹¹ Jean Fourquet souligne dans sa préface à la deuxième édition de 1967 que « la diffusion de la syntaxe de Tesnière ne ferait peut-être aujourd'hui que commencer, s'il n'y avait eu un fait nouveau que je n'avais pas prévu : c'est l'apparition de la traduction et de la documentation automatiques, à l'aide des ordinateurs électroniques. Des milieux de formation mathématicienne, libres des habitudes acquises dont la résistance avait retardé la publication de l'ouvrage de Tesnière, ont eu à établir des programmes d'analyse linguistique rigoureux, et ont trouvé chez Tesnière des suggestions immédiatement utiles. Son livre a été lu à l'Euratrom, aux C.E.T.A. [Centre d'Etude de la Traduction Automatique] de Paris et de Grenoble. Quand on discute dans ces milieux des graphes de Tesnière-Hays, comparés à ceux de Chomsky-Harris, il va de soi que les *Éléments de syntaxe structurale* sont parmi les ouvrages universellement connus, au moins quant aux thèses essentielles. »

scientifiques. Nous reviendrons sur ce point, mais nous allons d'abord étudier les arguments scientifiques qui expliquent le retour de la syntaxe de dépendance (en rappelant encore une fois que cet exposé est clairement partisan).

5. Connexion entre mots ou groupement de mots

Commençons par une citation de la première page de l'ouvrage posthume de Lucien Tesnière :

« Tout mot qui fait partie d'une phrase cesse par lui-même d'être isolé comme dans le dictionnaire. Entre lui et ses voisins, l'esprit aperçoit des connexions, dont l'ensemble forme la charpente de la phrase. [...] Comprendre une phrase, c'est saisir l'ensemble des connexions qui en unissent les différents mots. »

La syntaxe de dépendance repose sur l'idée que la phrase se construit en connectant ensemble les unités minimales de la syntaxe, que nous considérerons ici, à la suite de Tesnière, comme étant les mots¹². Cela induit que l'interprétation d'un énoncé se fait au fur à mesure que les mots sont reconnus et connectés entre eux. Prenons un exemple :

Je connais un type qui peut résoudre ce problème.

Le complément d'objet direct de la forme verbale *connais* est le groupe nominal *un type qui peut résoudre ce problème*. Il n'est pas nécessaire d'avoir analysé l'ensemble de ce groupe nominal pour commencer à interpréter cet énoncé et pour situer le complément d'objet direct du verbe *connais*. Le segment *Je connais un type qui* pourra être entièrement interprété sans qu'il soit nécessaire d'avoir analysé la suite de cet énoncé (qui pourrait d'ailleurs rester facilement incomplet, comme le sont souvent de tels énoncés à l'oral). Cela signifie qu'il n'est pas nécessaire d'avoir analysé complètement un constituant avant de le rattacher à son gouverneur. Il est assez clair que les énoncés sont analysés *incrémentalement*, c'est-à-dire au fur et à mesure de leur réception (on dit également de gauche à droite lorsqu'il s'agit de lecture). La syntaxe de dépendance rend assez bien compte de cela : on peut considérer qu'à chaque nouveau mot analysé, le locuteur va construire les liens syntaxiques possibles entre ce mot et le segment qui le précède. Comme le segment *Je connais un type qui* est connexe pour la dépendance, il n'y a aucune difficulté à l'analyser complètement :

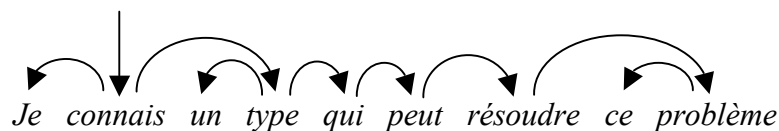


Figure 8. Arbre de dépendance linéarisé

Lorsque la forme verbale *connais* est analysée (au moment où elle est lue ou entendue), elle peut immédiatement être connectée au pronom sujet *je* qui la précède. Lorsque le nom *type* est analysé, il pourra être immédiatement connecté à l'article *un* et à la forme verbale *connais* qui le précède. Et ainsi de suite. On notera que le pronom relatif *qui* est ici analysé comme la tête de la proposition relative, ce qui est justifié par le fait que c'est bien lui qui marque la proposition comme pouvant modifier un nom, même s'il possède également un rôle pronominal et

¹² Voir Kahane 2008a pour une discussion consacrée aux unités minimales de la syntaxe où il est défendu que les mots ne sont qu'un niveau de cohésion particulier de combinaison des unités minimales, lesquelles sont plutôt les lexèmes et les morphèmes flexionnels.

sature la position sujet de la forme verbale qui suit. Il sera ainsi connecté à son antécédent, c'est-à-dire ici le nom *type*.

L'analyse en syntaxe de dépendance repose donc sur l'idée, défendue notamment par Richard Hudson (1993), que les mots se connectent deux par deux, pour former une structure arborescente, au fur et à mesure de leur réception par le destinataire de l'énoncé. Cette idée a été exploitée avec succès en traitement automatique des langues, notamment en analyse syntaxique automatique. On appelle *analyse syntaxique automatique* le calcul automatique d'une (ou plusieurs) représentation(s) syntaxique(s) pour une phrase donnée. Les deux campagnes d'évaluation des analyseurs syntaxiques ont été remportées à chaque fois par un analyseur basé sur les dépendances : en 2000 avec l'analyseur de Jacques Vergne (2000), puis en 2007 avec l'analyseur de Didier Bourigault (2007) (Campagne EASY — Evaluation des Analyseurs SYntaxiques). Ce dernier repose sur une architecture relativement simple : un module est dédié à chaque type de dépendance. Ainsi un premier module va-t-il chercher tous les liens entre déterminant et nom, sachant que ce type de lien est nécessairement court (seuls certains adjectifs éventuellement modifiés par un adverbe peuvent s'interposer : *un autre très petit livre*). Puis on cherche à lier auxiliaire et participe et ainsi de suite en étudiant des types de liens de plus en plus complexes. Trois niveaux de complexité sont considérés : les dépendances « courtes » (déterminant ← nom, clitique ← verbe, auxiliaire → participe, etc.), les dépendances sujet ← verbe et verbe → objet_direct, puis les autres. Les groupes prépositionnels sont étudiés en dernier en raison de la grande ambiguïté qu'ils provoquent puisqu'ils peuvent à priori dépendre d'un verbe (*parler à tout le monde de partir, compter sur quelqu'un*) comme d'un nom (*une part de gâteau, un article sur la syntaxe, un problème à résoudre*) ou d'un adjectif (*loin de Paris, difficile à traiter*). La désambiguïsation est effectuée grâce à des données statistiques obtenues sur de très gros corpus : par exemple si l'on analyse *donner une glace à la fraise*, on observera que *glace→à→fraise* est plus courant que *donner→à→fraise* et on rattachera donc *à la fraise* à *glace* plutôt qu'à *donner* (le résultat sera inverse avec *donner une glace à la dame*). Comme on le voit, un tel analyseur exploite à différents niveaux le fait que la structure syntaxique s'exprime par des relations de dépendance (plutôt que des regroupements en constituants).

6. Syntaxe et prosodie

Un autre domaine où la syntaxe de dépendance semble particulièrement avantageuse est celui de l'interface syntaxe-prosodie. Considérons le texte suivant extrait d'un article de Jacqueline Vaissière (1997). Il s'agit d'un texte lu à haute voix pour lequel le découpage en constituants prosodiques a été indiqué :

[On a depuis longtemps / essayé de faire parler des machines.]GS-
[Ce courant / a été inauguré / à la fin / du dix-huitième siècle]GS+
[par les essais / de Van Kempelen]GS+ [et de l'Abbé Rousselot]GS-
[Par des mécanismes / souvent astucieux,]GS-
[il fut possible / de reproduire / certains sons,]GS+
[puis certains enchaînements / de sons.]GS-

Les constituants prosodiques majeurs sont les *groupes de souffle*, indiqué ici entre crochets : tout locuteur doit régulièrement reprendre son souffle, ce qui se traduit par une pause perceptible (pause qui peut éventuellement être simulée, même quand le locuteur n'a pas besoin de reprendre son souffle). Deux types de groupes de souffle — GS+ et GS- — sont considérés, selon le contour terminatif : les GS+ se terminent sur une intonation montante indiquant que l'énoncé va se poursuivre, tandis que les GS- se terminent sur une intonation descendante. Les

GS sont eux-mêmes subdivisés en *groupes accentuels* séparés par des frontières prosodiques mineures, symbolisées ici par des slashes.

Il a été noté depuis longtemps que les constituants prosodiques ne correspondent pas aux constituants syntaxiques. Par exemple, ni *on a depuis longtemps*, ni *a été inauguré*, ni *à la fin*, ni *par les essais*, ... ne sont des constituants syntaxiques, au sens de l'analyse en constituants immédiats. Les frontières prosodiques sont généralement des frontières de constituants, mais il est rare que la frontière gauche et droite d'un constituant prosodique soient les frontières d'un même constituant syntaxique. De plus, les frontières prosodiques majeures ne sont pas nécessairement des frontières de constituants majeurs : tel est le cas par exemple de la frontière prosodique majeure GS+ entre *de Van Kempelen* et *et de l'abbé Rousselot*, qui forment pourtant ensemble un constituant très enchâssé.

Les constituants prosodiques présentent par contre une propriété très caractéristique et fort simple à énoncer du point de vue de la syntaxe de dépendance, proposée par Piet Mertens (1987) : les groupes prosodiques sont connexes du point de vue de la syntaxe de dépendance, c'est-à-dire que deux mots d'un même constituant prosodique sont toujours reliés par une chaîne de dépendances à l'intérieur de ce constituant. On peut constater la validité de cette propriété sur notre petit exemple dans la Fig. 9 où nous avons ajouté les relations de dépendance.

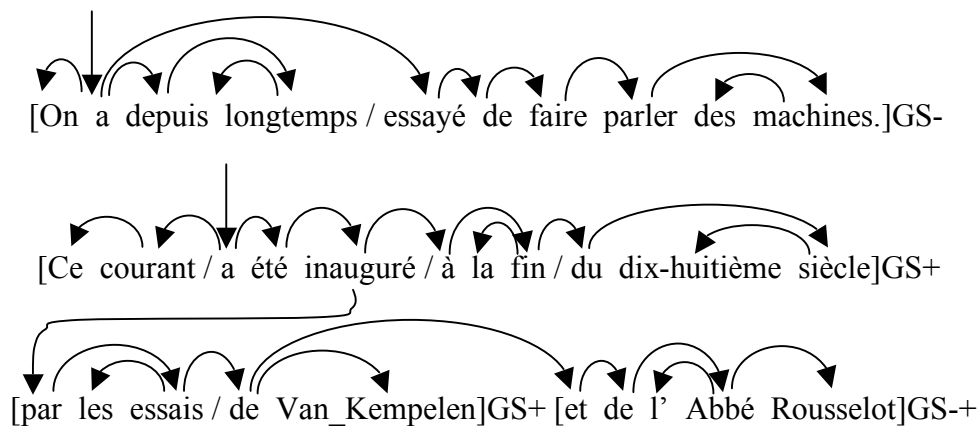


Figure 9. Dépendances syntaxiques et constituants prosodiques

On remarquera de plus que, même si les règles sont en fait plus complexes que cela, les frontières prosodiques majeures GS+ tendent à rompre des dépendances longues : tel est le cas par exemple de la dépendance entre *a* et *essayé*, qui est d'ailleurs une dépendance théoriquement « courte », mais qui est ici allongée par l'insertion du complément circonstanciel *depuis longtemps*. C'est encore le cas de la dépendance entre *inauguré* et *par les essais*. Les frontières prosodiques majeures GS- correspondent quant à elles aux frontières de phrases, qui ne sont pas couvertes par une dépendance. Autrement dit, plus la dépendance entre deux éléments est courte plus la cohésion prosodique entre eux est forte. On peut ainsi à partir de l'arbre de dépendance linéarisé, sachant que les groupes accentuels font autour de 6-7 syllabes en débit normal, prédire quels sont les découpages possibles, en groupes accentuels comme en groupes de souffle (Mertens 1987).

7. Syntaxe et sémantique

Nous avons étudié l'analyse, nous allons maintenant nous intéresser à la production d'énoncés.

Considérons un locuteur du français qui veut exprimer un certain sens que nous allons essayer de décrire sans le verbaliser. Ce sens concerne une personne appelée Pierre et un événement concernant Pierre. Cet événement est une maladie et la durée de cette maladie est de deux semaines. Nous pouvons représenter ce sens par la « formule » suivante :

x : 'Pierre'
 e : 'malade'(x)
 'durée'(e, '2 semaines')

Dans cette représentation, il y a quatre éléments de sens considérés : 'Pierre', 'malade', 'durée' et '2 semaines'. Ce dernier élément est en fait la combinaison de deux sens : l'unité de temps 'semaine' et le prédicat 'deux' qui la quantifie. Certains sens possèdent des *arguments* : le sens 'malade' possède toujours un argument (*quelqu'un est malade*), le sens 'durée' possède deux arguments (*quelque chose dure quelque temps*). Le sens 'Pierre', quant à lui, renvoie à une entité du monde et n'a pas d'argument. Dans la formule ci-dessus, les variables x et e nous ont servi à indiquer que certains sens étaient arguments d'autres : x désigne le sens 'Pierre' et 'malade'(x) signifie que l'élément désigné par x est l'argument de 'malade'. On peut représenter la structure du sens que nous considérons par un graphe étiqueté, que nous appelons le *graphe sémantique*, dont les nœuds sont les sens 'Pierre', 'malade', 'durée', 'semaine' et 'deux' et dont les arcs représentent la relation entre un sens et un de ses arguments. Nous numérotons les arcs avec les chiffres 1 et 2 pour distinguer premier et deuxième argument.

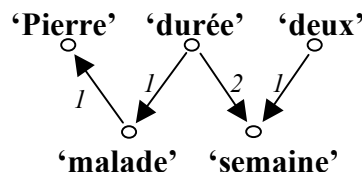


Figure 10. Graphe sémantique

Il y a plusieurs façons d'exprimer ce sens en français. Nous allons nous intéresser à deux formulations possibles de ce sens, suffisamment différentes pour illustrer notre propos.

- (1) *La maladie de Pierre a duré deux semaines.*
- (2) *Pierre a été malade pendant deux semaines.*

Pour produire un énoncé à partir d'une représentation sémantique, nous devons lexicaliser chacun des sens. Un sens peut s'exprimer de façons variées : par exemple, le sens 'durée' peut aussi bien être exprimé par le nom DURÉE que le verbe DURER ou encore les prépositions DURANT ou PENDANT. La lexicalisation obéit à des contraintes syntaxiques imposées par la structure hiérarchique. Tout se passe comme si on avait suspendu le graphe sémantique par l'un de ses nœuds dont on décide de faire le verbe principal de l'énoncé et qu'on parcourait le graphe à partir de ce nœud pour les autres lexicalisations. Nous pouvons illustrer cela par les schémas suivants : à gauche nous représentons le graphe sémantique suspendu par un de ses nœuds (visé ici par une grosse flèche grise) et à droite nous avons la structure hiérarchique correspondante où les sens ont été lexicalisés.

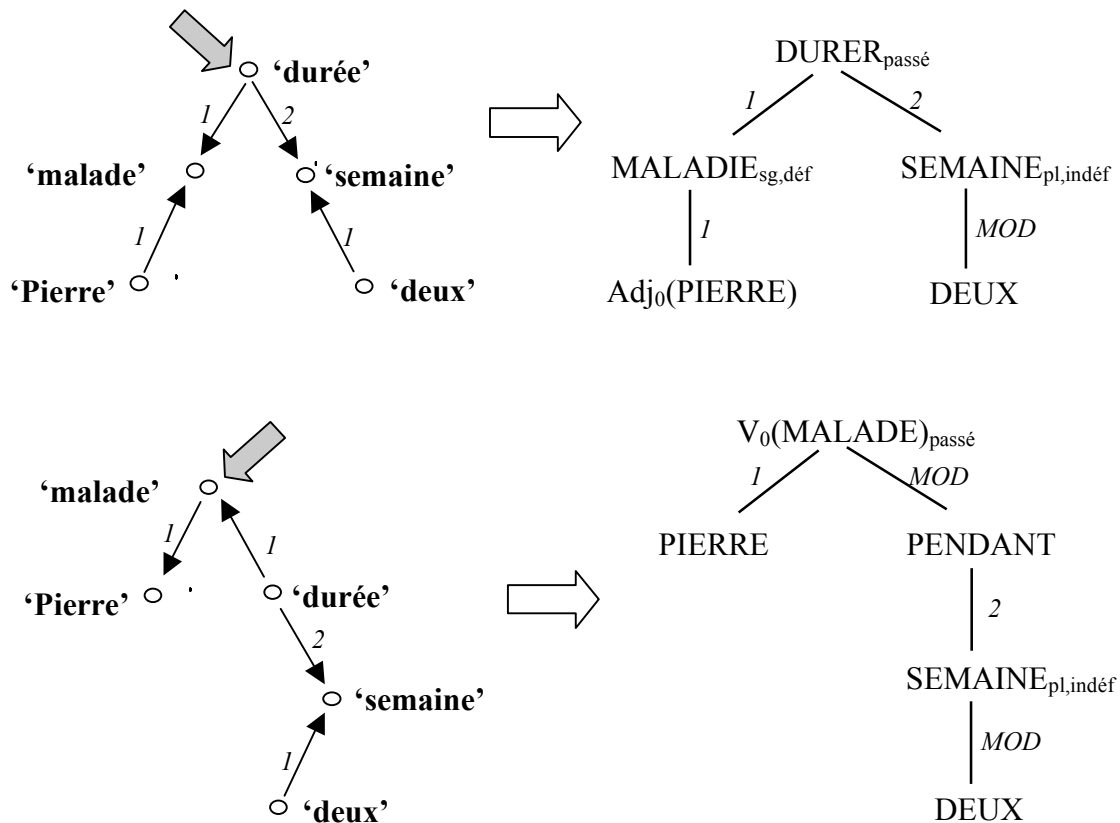


Figure 11. Passage d'un graphe sémantique à un arbre syntaxique profond

La structure que nous obtenons s'appelle un *arbre de dépendance syntaxique profond*. Chaque nœud de l'arbre est occupé par une unité lexicale, éventuellement accompagnée d'éléments grammaticaux : temps pour les verbes (présent, passé, etc.), nombre (singulier, pluriel) et définitude (défini, indéfini) pour les noms. Chaque dépendance est étiquetée en fonction de son parcours : les relations sémantiques qui ont été parcourues dans le sens de la flèche (du prédicat vers l'argument) donne une dépendance syntaxique *actancielle* de même numéro, tandis que les relations sémantiques qui ont été parcourues à contresens donne une dépendance *modificative* (étiquetée *MOD*). Chaque unité lexicale se voit imposer sa *partie du discours* par son gouverneur, la racine étant un verbe à l'indicatif. Ainsi dans la phrase (1), PIERRE qui dépend du nom MALADIE devrait être réalisé par un adjectif, ce que nous indiquons par la notation $Adj_0(PIERRE)$. C'est la préposition DE qui assurera cette *translation* (*la maladie de Pierre*) et permettra à PIERRE d'être complément du nom. De la même façon, la notation $V_0(MALADE)$ indique que MALADE occupe une position où un verbe est attendu et c'est la copule ÊTRE qui assurera la translation d'adjectif en verbe. La lexicalisation obéit ainsi à des contraintes générales, dégagées par Tesnière (1959), imposées par la structure hiérarchique : 1) la racine de l'arbre syntaxique d'une phrase ordinaire est un verbe ; 2) les dépendants actanciels d'un verbe sont des noms ; 3) les modificateurs d'un verbe sont des adverbes ; 4) les dépendants d'un nom sont des adjectifs ; 5) les dépendants d'un adjectif ou d'un adverbe sont des adverbes. Par ailleurs, des éléments de n'importe quelle catégorie peuvent occuper n'importe quelle position à condition d'être traduits par un élément adéquat. Ainsi un verbe peut être actant d'un autre verbe ; les principaux translatifs de verbes en nom sont l'infinitif (*Pierre veut manger*) et la conjonction de subordination *que* (*Pierre veut que Marie mange*). Dans l'exemple (2), la relation sémantique entre 'durée' et 'malade' est parcourue à l'envers : 'durée' est donc lexicalisé comme un modificateur d'une tournure verbale et c'est ainsi que la préposition PENDANT peut être choisie.

La procédure que nous venons de décrire, ainsi que la notion de structure syntaxique profonde, est au cœur de la théorie Sens-Texte (cf. Mel'čuk 1988, 1997, 2003 ou Kahane 2001 ; elle a été formalisée par Lida Iordanskaja et Alain Polguère (Iordanskaja & Polguère 1988, Polguère 1990) et implémentée dans plusieurs systèmes de génération de textes. Cette procédure montre que le passage d'une représentation sémantique à une représentation syntaxique est assez élégant à modéliser dans le cadre d'une syntaxe de dépendance. Pour assurer la production complète de l'énoncé, il faut encore passer de l'arbre syntaxique profond à l'arbre de dépendance de surface, ce qui consiste essentiellement en l'expansion de certains nœuds (par exemple de V_0 (MALADE) en ÊTRE → MALADE). L'étape suivante est la linéarisation de l'arbre de dépendance et le calcul de la prosodie. C'est là qu'on voit tout l'avantage à pouvoir considérer les relations hiérarchiques indépendamment de l'ordre linéaire : c'est une fois établies les relations de dépendance que l'on peut calculer l'ordre des mots, ordre des mots qui est très variable d'une langue à l'autre avec notamment des langues où les dépendants précèdent normalement leur tête (comme le coréen ou le turc), des langues où les dépendants suivent normalement leur tête (comme le gallois ou l'arabe classique) et des langues où l'ordre n'est pas contrôlé par la structure hiérarchique, mais plutôt par l'importance communicative des éléments (comme le russe). Le français est une langue où l'ordre des mots est fortement contrôlé par la structure de dépendance, mais où la tête tend à se situer au milieu de ses dépendants (tel le verbe entre sujet et compléments).

8. Chaînes verbales

Nous allons terminer notre présentation par l'étude de phénomènes syntaxiques complexes, que l'on regroupe traditionnellement sous le terme de *dépendances non bornées* et qui ont constitué la pierre d'achoppement de tous les modèles formels développés dans la deuxième moitié du 20^{ième} siècle. Nous allons plus précisément nous intéresser aux contraintes syntaxiques qui pèsent sur les relatives en français. Considérons un premier exemple de relative (indiquée entre crochets) :

Voilà la personne [à laquelle Félix veut parler].

Dans une relative, le pronom relatif (ici *laquelle*) joue un double rôle : d'une part, c'est un subordonnant qui permet à la proposition relative de venir modifier un nom (ici *personne*) ; d'autre part, c'est un pronom qui remplit dans la relative le même rôle que *cette personne* dans la proposition *Félix veut parler à cette personne*, à savoir celui de complément d'objet indirect. On passe de cette proposition à la proposition relative en remplaçant *cette personne* par le pronom relatif *laquelle*. Mais à la différence d'un groupe nominal ordinaire, le pronom relatif doit nécessairement être au début de la proposition (entraînant ici avec lui la préposition *à* qui marque son rôle syntaxique de complément d'objet indirect). Il s'en suit des contraintes qui font que n'importe quelle position syntaxique ne peut pas être relativisée.

Il y a plusieurs façons de modéliser ces contraintes. La plus connue a été proposée par John Ross (1967) dans le cadre des grammaires syntagmatiques. Il s'agit de regarder quelles frontières de constituants doivent être « traversées » pour aller de la position canonique qu'occuperait un groupe nominal ordinaire à la position antéposée du pronom relatif. On constate par exemple qu'on peut traverser la frontière d'une subordonnée complétive objet :

Voilà la personne à laquelle je veux [que tu parles __] .

Mais on ne peut traverser la frontière d'une subordonnée complétive sujet, ni celle d'une subordonnée circonstancielle ou d'une relative :

**Voilà la personne à laquelle [que tu parles __] me déplaît.*

*Voilà la personne à laquelle je viendrai [quand tu parleras __].

*Voilà la personne à laquelle je connais la fille [qui a parlé __].

On peut donner dans le cadre de la syntaxe de dépendance une tout autre modélisation, qui, bien que formellement équivalente, jette un regard différent sur les raisons de ces contraintes. Plutôt que de décrire ce qu'on ne peut pas relativiser, concentrons-nous sur ce qui peut l'être. On remarque qu'il est possible dans une relative de remplacer le verbe principal de la relative par n'importe quelle chaîne de verbes :

Voilà la personne à laquelle Félix / parle.
veut parler.
devrait penser à parler.
aurait aimé que Luce puisse accepter de parler.

Si l'on prend le dernier exemple, il y figure la chaîne *aurait* → *aimé* → *que* → *accepte* → *de* → *parler* qui est composée de maillons de différents types : Verbe_auxiliaire → Verbe_participe (*avoir* → *aimé*), Verbe → Verbe_infinitif (*pouvoir* → *accepter*), Verbe → Préposition → Verbe_infinitif (*accepter* → *de* → *parler*) ou Verbe → Complémenteur → Verbe (*aimer* → *que* → *puisse*). On peut encore ajouter d'autres types de maillons avec des tournures verbales contenant un nom (*à laquelle Félix a* → *l'intention de parler* ; *à laquelle Félix fait* → *mine de parler*) ou un adjectif (*à laquelle il est* → *impossible de parler* ; *à laquelle Félix semble* → *heureux de parler*).

De telles chaînes de verbes et de tournures verbales ont la particularité de se comporter, du point de vue de la relativisation, comme un verbe simple : nous les appellerons des *nucléus verbaux* (Kahane 1997). On peut alors formuler la contrainte sur l'extraction de manière très simple : seul peut être relativisé un dépendant immédiat du verbe principal, sachant que le verbe principal peut être un verbe simple, mais aussi un nucléus verbal.

On peut aller plus loin dans cette modélisation. Remarquons avant que les structures de dépendance linéarisées des relatives avec dépendance non bornnée violent une propriété que l'on appelle la *projectivité*. Un arbre de dépendance linéarisé est dit *projectif* si, une fois les nœuds alignés sur une droite, on peut placer les dépendances dans le même demi-plan sans qu'elles se coupent. En particulier, la racine de l'arbre, sur lequel nous plaçons une dépendance verticale infinie ne doit pas être couverte. Comme l'a montré Yves Lecerf (1961), un arbre de dépendance linéarisé est projectif si et seulement si les constituants de l'arbre de constituants correspondant sont tous *continus* (c'est-à-dire connexes pour l'ordre linéaire). Les arbres de dépendances des Fig. 8 et 9 sont projectifs, mais celui de la Fig. 12 ne l'est pas : en effet, la dépendance entre *parler* et *à laquelle* coupe la dépendance entre le gouverneur de la relative et le verbe principal (par soucis de simplification nous n'avons pas pris en compte le rôle de complémenteur du pronom relatif qui aurait dû nous amener à en faire la tête de la relative) ; de manière équivalente, on constate que le constituant [*à laquelle ... parler*] est discontinu.

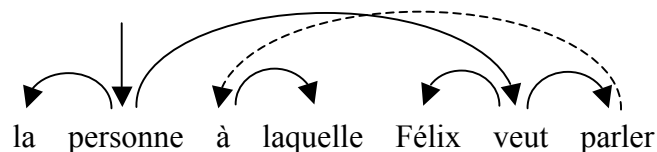


Figure 12. Un arbre de dépendance linéarisé non projectif

La non-projectivité de la relative de la Fig. 12 résulte du fait que le pronom relatif ne dépend pas directement du verbe principal de la relative. Si l'on considère maintenant que la chaîne de verbes qui va du verbe principal au verbe qui gouverne le pronom relatif se comporte comme un tout — que nous avons appelé le nucléus verbal — et qu'on attribue au nucléus verbal une position structurale, on obtient la structure suivante, que nous appelons un *arbre à bulles* :

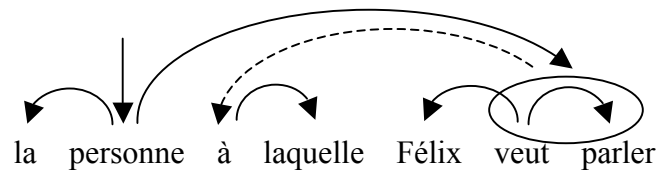


Figure 12. Un arbre à bulles

Contrairement à l'arbre de dépendance auquel il correspond, cet arbre à bulles est projectif, dans le sens où les dépendances ne se coupent pas (cf. Kahane 1997 pour une définition formelle) : pour cela nous avons attribué à la bulle représentant le nucléus verbal les dépendances venant du gouverneur de la relative et allant vers le pronom relatif.

Pour conclure sur ce point, nous pouvons dire que, à notre avis, la description de la relativisation doit prendre en compte le fait que les chaînes de verbes se comportent pour ce phénomène comme un tout et qu'une modélisation satisfaisante de la relativisation doit pouvoir permettre de considérer facilement les nucléus. Ce n'est pas le cas des arbres de constituants où les chaînes verbales sont difficiles à isoler. Par contre, les nucléus sont faciles à décrire dans le cadre d'une syntaxe de dépendance et ils peuvent même être intégrés dans la structure de dépendance à condition d'enrichir un peu le formalisme en introduisant des bulles. Les grammaires syntagmatiques sont, elles, obligées, pour saisir les phénomènes de dépendances non bornées, d'introduire des mécanismes supplémentaires complexes, comme les transformations dans les modèles chomskyens (Chomsky 1957, 1965) ou les traits non locaux des grammaires d'unification comme HPSG (cf. Gerdes 2005 pour une critique approfondie).

9. Conclusion

Pour les raisons historiques que nous avons exposées au début de cette présentation (notamment le très fort impact des travaux de Noam Chomsky), les représentations syntaxiques en termes de constituants se sont imposées au détriment des représentations en termes de dépendance. Le choix d'une représentation en constituants est rarement justifié par les linguistes qui en font usage et ceux-ci donnent souvent l'impression de n'avoir jamais envisagé que la structure syntaxique puisse être représentée autrement que par un arbre de constituants¹³. Nous pensons pour notre part que la structure syntaxique est de toute façon plus complexe que ne laisse apparaître les représentations utilisées actuellement en linguistique et que, en particulier, elle n'est pas strictement hiérarchique, comme le laisse par exemple penser des phénomènes tels que la relativisation. Nous pensons que l'idée, formulée par Lucien Tesnière, qu'un énoncé se construit en connectant ensemble des unités syntaxiques minimales en formant des mots puis en assemblant ces mots est une constatation incontournable sur le fonctionnement des langues que toute modélisation se doit de refléter. C'est pourquoi nous pen-

¹³ Signalons que nous considérons qu'une représentation en termes d'arbre de constituants est utile pour représenter la structure topologique et prosodique, où la notion de tête ne joue plus de rôle, alors que l'ordre des mots est central (cf. Gerdes & Kahane 2006).

sons, au-delà de tous les autres arguments possibles, qu'une structure de dépendance est mieux adaptée qu'une structure de constituants à la représentation de la structure syntaxique : elle met en évidence la connexion des éléments deux à deux¹⁴ et permet de considérer facilement des configurations d'éléments connectés entre eux qui ne sont pas nécessairement des constituants au sens des grammaires syntagmatiques. Nous avons vu l'intérêt de telles configurations lorsqu'il a été question de constituants prosodiques, de chaînes verbales dans la relativisation ou des segments initiaux dans une analyse incrémentale.

Les arbres de dépendance s'illustrent de plus par le fait que, tout en étant plus simples que les arbres de constituants (ne serait-ce qu'en nombre de nœuds), elles n'en sont pas moins plus informatives, puisque, sans éliminer la notion de constituant, elles mettent en évidence la notion de tête. Comme l'ont montré de nombreuses expériences pédagogiques, à commencer par celles menées par Tesnière dans les écoles de Montpellier, elles sont très faciles à enseigner et rapidement acquises par des étudiants de tout âge. On peut donc regretter que la syntaxe de dépendance ne soit pas aujourd'hui enseignée en France, ni à l'école, ni dans les premières années des cursus universitaires en sciences du langage et reste peu utilisée en dehors des courants de la linguistique issus du traitement automatique des langues. On peut s'étonner (ou pas) du fait qu'un grand linguiste comme Lucien Tesnière n'ait pas été davantage défendu par la communauté scientifique française et que ses idées novatrices n'aient pas été exploitées plus tôt en France, les linguistes français préférant alors, comme certainement leurs condisciples de d'autres disciplines, promouvoir des théories venant des Etats-Unis d'Amérique. C'est encore le cas aujourd'hui où il est plus facile d'accéder à la reconnaissance nationale et internationale en adaptant au français une théorie développée en Amérique que de défendre des idées réellement originales.

Remerciements

Ce texte a été relu par Igor Mel'čuk qui, comme à chaque fois, m'a signalé de nombreuses erreurs, que je n'ai qu'imparfaitement corrigées.

Références

- Abeillé A. (1996-97), Fonction objet ou position objet (1^{ère} et 2^{nde} parties), *Gré des langues*, 11, 8-29 ; 12, 8-33.
- Arrivé M. (1969), Les Eléments de syntaxe structurale de Lucien Tesnière, *Langue Française*, 1, 36-40.
- Bloomfield L. (1933), *Language*, New York.
- Bresnan J. (2001), *Lexical-Functional Syntax*, Blackwell, Oxford.
- Bourigault D. (2007), *Un analyseur syntaxique opérationnel : SYNTAX*, Thèse d'HDR, Université Toulouse-Le Mirail.

¹⁴ Profitons pour souligner qu'on peut penser que la connexion met souvent en jeu plus de deux éléments en même temps. Par exemple, dans *Marie parle à Pierre*, on peut considérer qu'il y a une connexion marquée par *à* qui se fait entre *parler* et *Pierre*, et que donc on doit traiter les trois éléments simultanément, plutôt que de considérer deux connexions *parler* → *à* et *à* → *Pierre*. Cette idée est déjà présente chez Tesnière qui donne à des éléments tels que *à* une position structurale bien spéciale et en introduit des structures du type arbre à bulles pour représenter des configurations d'éléments fonctionnant simultanément. Dans un même ordre d'idée, attribuer à la forme verbale, plutôt qu'au lexème verbal seul, une position structurale, c'est considérer que les deux éléments — le verbe et sa flexion — agissent de pair dans la connexion avec d'autres éléments.

- Chomsky N. (1957), *Syntactic Structure*, MIT Press, Cambridge.
- Chomsky N. (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge.
- Chomsky N. (1993), *The Minimalist Program*, MIT Press, Cambridge.
- Gaifman H. (1965), Dependency systems and phrase-structure systems, *Information and Control*, 18, 304-337 ; Rand Corporation, 1961, RM-2315.
- Gerdes K. (2005), Sur la non-équivalence des représentations syntaxiques : comment la représentation en X-barre nous amène au concept du mouvement, *Les Cahiers de Grammaire*, 30, 17 p.
- Gerdes K., Kahane S. (2006), L'amas verbal au cœur d'une modélisation topologique de l'ordre des mots, in K. Gerdes & C. Muller (éds), *Ordre des mots et topologie de la phrase française*, *Linguisticae Investigationes*, 29:1, 75-80.
- Hays D. (1964), Dependency theory: A formalism and some observations, *Language*, 40:4, 511-525 ; Rand Corporation, 1960, RM-2646 (Grouping and dependency theories).
- Hudson R. (1980), Constituency and Dependency, *Linguistics*, 18:3/4, 179-198; A Second Attack on Constituency: A Reply to Dahl, *Linguistics*, 18:5/6, 489-504.
- Hudson R. (1993), Do We Have Heads in Our Minds?, in Corbett G., Fraser N., McGlashan S. (éds), *Heads in Grammatical Theory*, Cambridge Univ. Press, 266-291.
- Iordanskaja L., Polguère, A. (1988), Semantic processing for text generation, Technical report, ORA, Canada.
- Kahane S. (1997), Bubble trees and syntactic representations, in T. Becker & U. Krieger (éds), *Proc. 5th Meeting of the Mathematics of Language (MOL5)*, DFKI, Saarbrücken, 70-76.
- Kahane S. (2001), Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel, *Actes TALN 2001*, vol. 2, 17-76.
- Kahane S. (2008a), Les unités minimales de la syntaxe et de la sémantique : le cas du français, *Congrès mondial de linguistique française*, Paris, 2519-2538.
- Kahane S. (2008b), On the Status of Phrases in Head-driven Phrase Structure Grammar - Illustration by a Totally Lexical Treatment of Extraction, in A. Polguère (ed.), Benjamins, 34 p.
- Łukasiewicz J. (1930), Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls, *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 23, 51-77. Trad. angl. de H. Weber : Philosophical Remarks on Many-Valued Systems of Propositional Logics, in S. McCall (ed), *Polish Logic 1920-1939*, Clarendon Press, Oxford (1967).
- Mel'čuk I. (1988), *Dependency Syntax: Theory and Practice*, The SUNY Press, Albany, N.Y.
- Mel'čuk I. (1997), *Vers une Linguistique Sens-Texte*, Leçon inaugurale au Collège de France, Collège de France, Paris, 78p.
- Mel'čuk I. (2003), Levels of Dependency in Linguistic Description: Concepts and Problems. In V. Agel, L. Eichinger, H.-W. Eroms, P. Hellwig, H. J. Heringer, H. Lobin (eds): *Dependency and Valency. An International Handbook of Contemporary Research*, vol. 1, Berlin - New York, W. de Gruyter, 188-229.
- Mertens P. (1987), *L'intonation du français. De la description linguistique à la reconnaissance automatique*, Thèse de doctorat, Université de Leuven, Belgique.
- Polguère A. (1990), *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*, Thèse de doctorat, Université de Montréal.
- Pollard C., Sag I. (1994), *Head-driven Phrase Structure Grammar*, CSLI, Stanford.

- Ross, J. R. (1967), *Constraints on Variables in Syntax*, Thèse de doctorat, MIT, Cambridge ; publié sous le titre *Infinite Syntax!*, Reidel, Dordrecht, 1985.
- Sgall P. (1967), Functional Sentence Perspective in a Generative Description, *Prague Studies in Mathematical Linguistics*, 2, 203-225.
- Sgall P., Hajičová E., Panevová J. (1986), *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*, Reidel, Dordrecht.
- Tesnière L. (1934), Comment construire une syntaxe, *Bulletin de la Faculté des Lettres de Strasbourg*, 7, 12^{ème} année, 219-229.
- Tesnière L. (1959), *Éléments de syntaxe structurale*, Kincksieck, Paris.
- Vaissière J. (1997), Langues, prosodie et syntaxe, *TAL*, 38:1, numéro spécial *Prosodie et syntaxe*, 53-82.
- Vergne J. (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.