

UNE GRAMMAIRE DE DEPENDANCE LEXICALISEE AVEC BULLES POUR TRAITER L'EXTRACTION

Sylvain KAHANE

TALaNa (Univ. Paris 7) & Univ. Paris 10 - Nanterre
sk@ccr.jussieu.fr

Abstract

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

Introduction

Cette étude présente une grammaire de dépendance lexicalisée traitant les extractions (= relativisation, interrogation, ...) par un enrichissement de la structure syntaxique et des opérations de combinaison des structures élémentaires.¹

Dans la Section 1, nous présenterons une grammaire de dépendance lexicalisée basée sur celle de Nasr 95, 96. Nous verrons quelques enrichissements de ce formalisme permettant d'éviter la multiplication des structures élémentaires. Dans la Section 2, nous présenterons, suivant Kahane 97, les arbres à bulles, une extension des arbres de dépendance permettant une représentation élégante des extractions. Dans la Section 3, nous proposerons une grammaire lexicalisée avec bulles traitant l'extraction. Une comparaison sera effectuée avec le traitement d'autres formalismes (TAG, CG, HPSG, Syntaxe X' et des grammaires de dépendance).

Pour éviter toute ambiguïté par la suite nous distinguerons deux termes *grammaire* : une grammaire₁ est un système formel permettant de décrire une langue (par ex., une grammaire₁ générative); dans une grammaire₁, on distingue généralement deux composantes qu'on appelle la grammaire₂ et le lexique.

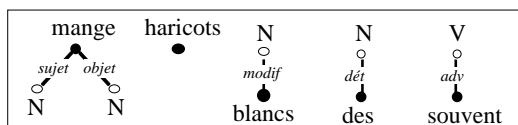
¹ La mise en forme de ce formalisme est conséquent à un début de collaboration avec A. Nasr. Ce travail a également bénéficié de discussions avec A. Abeillé, M.-H. Candito, G.-J. Kruijff, I. Mel'cuk et A. Polguère.

1. Une grammaire₁ de dépendance sans bulles

La grammaire₁ de base

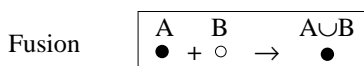
Notre grammaire₁ de base est la grammaire₁ proposée par Nasr 95 et 96, elle-même inspirée de la grammaire₁ TAG (Joshi 87). Il s'agit d'une **grammaire₁ lexicalisée**, c'est-à-dire que la grammaire₁ est réduite à un lexique où à chaque mot (= forme fléchée simple) est associé une ou plusieurs structures élémentaire (où sont également encodées les informations grammaticales₂). A la différence de TAG, une **structure élémentaire** est un **arbre de dépendance** du type des arbres syntaxiques de surface de Mel'cuk 88, eux-mêmes largement inspirés des stemmas de Tesnière 59. Les **nœuds** de la structure sont étiquetés par une structure de traits simples, comme par ex. [lemme: MANGER, cat: V, mode: ind, temps: présent, pers: 3, nbre: sg]. (Dans les figures, une telle structure de traits sera simplement résumée par *mange*.) Les **branches** ou dépendances sont étiquetées par des structures de traits indiquant une **fonction syntaxique**, ainsi que des informations relatives à l'ordre linéaire, sur lesquels nous reviendrons. L'un des nœuds, appelé **l'ancore** et dessiné en noir, est étiqueté par le mot associé à la structure élémentaire². Les autres nœuds, correspondant aux **requêtes** de l'ancore, sont en blanc.

² Notre grammaire se limite à l'interface morphologie-syntaxe, c'est-à-dire à la correspondance entre une chaîne morphologique et une représentation syntaxique. Pour cette raison, les structures élémentaires correspondent à des mots (ou plus exactement à des lemmes, la correspondance entre lemme+grammèmes et mot étant laissée à un module morphologique). Certaines grammaires lexicalisées tentent de réaliser directement l'interface morphologie-sémantique (notamment les grammaires TAG; Abeillé 91, Candito & Kahane 98a) et associent pour cette raison les structures élémentaires à des unités sémantiques pouvant correspondre à plus d'un mot. Nous pensons qu'il est préférable, surtout si l'on souhaite décrire élégamment des phénomènes syntaxiques complexes comme la coordination, de séparer les modules morphologie-syntaxe et syntaxe-sémantique pour la description linguistique, même si ces deux modules peuvent être ensuite combinés pour un traitement automatique. Le module syntaxe-sémantique ne sera pas considéré ici. Indiquons simplement qu'il s'agit essentiellement de la correspondance entre une configuration de nœuds syntaxique et un prédicat sémantique.

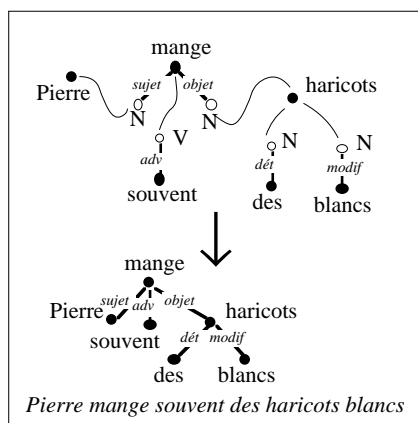


Des exemples de structures élémentaires
(sans indication pour la linéarisation)

Les structures élémentaires se combinent ensuite par une seule opération, la **fusion** d'un nœud noir (= ancré) et d'un nœud blanc (= non ancré), qui donne un nœud noir. Deux nœuds peuvent être fusionnés si leurs étiquettes A et B sont unifiables; le nœud résultant reçoit l'étiquette $A \cup B$, obtenue par unification des étiquettes A et B.



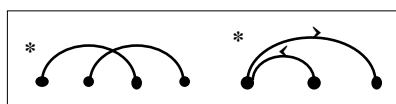
Une phrase est reconnue si on peut associer à chaque mot de la phrase une structure élémentaire et si ces structures peuvent être combinées afin de former un arbre dont tous les nœuds sont noirs.



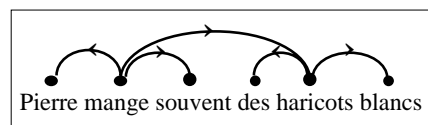
Une combinaison de structures élémentaires
formant une phrase

Ordre linéaire et projectivité

Il nous reste à voir comment est pris en compte l'ordre linéaire. Nous nous restreignons pour l'instant à des arbres de dépendance projectifs. Un arbre de dépendance assorti d'un ordre linéaire sur les nœuds est dit **projectif** si et seulement si, en plaçant les nœuds sur une ligne horizontale et tous les arcs dans le demi-plan supérieur, on peut assurer que deux arcs ne se coupent jamais et qu'aucun arc ne couvre un de ses ancêtres. Cette propriété est équivalente au fait que la **projection** de tout nœud x de l'arbre (c'est-à-dire l'ensemble des nœuds dominés par x , x compris) forme un segment continu de la phrase (Lecerf 61, Iordanskaja 63, Gladkij 66).

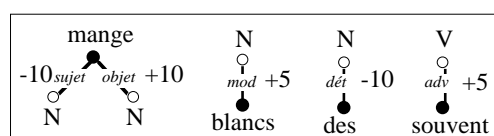


Les cas de non projectivité



Un exemple d'arbre de dépendance projectif

La projectivité présente un intérêt immédiat : il suffit pour ordonner un arbre projectif de spécifier la position de chaque nœud par rapport à son gouverneur, ainsi que vis-à-vis de ses frères. Pour ce faire, nous suivrons la proposition de Courtin & Genthial 98 consistant à assigner à chaque dépendance un poids³ : si le poids est négatif le dépendant doit se trouver à gauche de son gouverneur, s'il est positif, le dépendant doit se trouver à droite. Quant à deux frères, celui portant le poids le plus important (en valeur absolue) sera le plus éloigné du gouverneur.⁴



Des exemples de structures élémentaires
avec poids pour la linéarisation

Un analyseur à pile associé à la grammaire₁

Nous présentons ici de façon schématique l'analyseur décrit dans Nasr 95, 96. L'analyse d'une phrase requiert deux opérations : l'empilement et la réduction. L'**empilement** consiste simplement à mettre dans la pile une structure élémentaire à chaque fois qu'un mot est lu. La **réduction** consiste à fusionner deux cases adjacentes en combinant les arbres qu'elles contiennent. La combinaison de deux arbres n'est possible que si la projectivité est respectée. L'analyse réussit si, à la fin, seule une case de la pile est non vide et si celle-ci contient un arbre dont tous les nœuds sont noirs. Cet "analyseur" s'apparente à une grammaire₁ catégorielle (Bar-Hillel 53). La grammaire₁ de Nasr présente néanmoins à notre avis deux avantages sur ces dernières : 1) elle construit directement une véritable représentation géométrique et

³ D'autres solutions sont possibles. On peut par exemple écrire des règles de précedence linéaire comme en GPSG (Gazdar *et al.* 85) ordonnant un nœud et son gouverneur ou bien deux dépendants. On peut également, comme Nasr 96, en s'inspirant de Mel'cuk & Pertsov 87, indiquer pour chaque nœud la position par rapport à son gouverneur (droite ou gauche) et la liste des autres dépendants autorisés à se trouver entre eux.

⁴ Nous ne discuterons pas davantage la question des poids par la suite. Notons simplement que le poids dépend de différents facteurs : la fonction syntaxique (en français par ex., les syntagmes possèdent une légère tendance à apparaître dans l'ordre d'oblicité croissante), la taille et la nature du syntagme (un syntagme de taille importante a tendance à être plus éloigné de son gouverneur qu'un syntagme de petite taille; par ailleurs certaines positions n'autorisent que des syntagmes "lexicaux" : *un très haut mur* vs **un haut de 3m mur* (Blinkenberg 28, Abeillé & Godard 98) et la saillance communicative (l'importance ou le rôle donné par le locuteur au syntagme). Le poids devrait donc être une fonction de ces différents paramètres.

2) elle n'est pas précisément orientée vers l'analyse (et peut donc être facilement utilisée pour la synthèse).

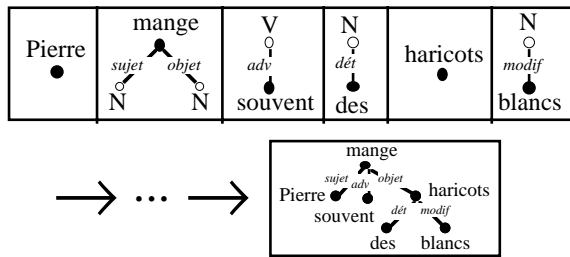


Schéma général de réduction de la pile

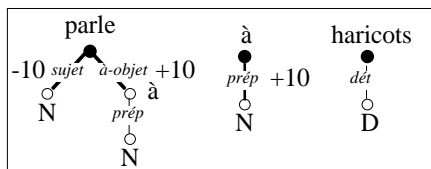
Nous ne discuterons pas ici de la mise en place d'une procédure réelle. Notons simplement que pour une analyse donnée l'ordre dans lequel les réductions sont faites n'a pas d'importance. On peut donc réaliser un analyseur incrémental qui privilégiera toujours la réduction avant l'empilement ou bien un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

Une grammaire₁ plus souple: liens blancs et traits prioritaires

Nous allons proposer une première extension au formalisme de Nasr 95, 96 visant à éviter la multiplication des structures élémentaires. De même que nous avons considéré des nœuds noirs et blancs, nous allons considérer des arcs noirs et blancs (représentés respectivement par des traits gras et fins). Un arc blanc peut être fusionné avec un autre arc (noir ou blanc), de même qu'un nœud blanc peut être fusionné avec un autre nœud (noir ou blanc). Comme précédemment, la structure résultante doit être composée entièrement de nœuds et arcs noirs.

L'utilisation d'arcs blancs permet d'assurer une contrainte sur un nœud adjacent sans pour autant construire la dépendance avec ce nœud. Ainsi, comme on le voit dans la figure ci-dessus, l'arbre de *parle* indique que la préposition *à* doit gouverner un nom, mais le nœud *à*, comme le lien entre *à* et son dépendant sont réalisés par l'arbre de *à*. Le fait d'avoir une structure élémentaire séparée pour *à* permet, par ex., de pouvoir relier *à* à *qui* avant que *parler* ne soit analysé dans l'analyse incrémentale d'une phrase comme *Pierre se demande à qui tu penses parler*.

De même, on pourra indiquer par un lien blanc dans la structure d'un nom que celui-ci requiert un déterminant.

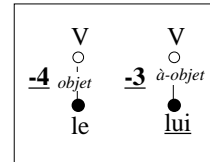


Des structures élémentaires avec liens blancs

Certains mots, comme les pronoms clitiques ou relatifs, peuvent occuper des positions argumentales, mais ne pas se trouver en position canonique. Dans

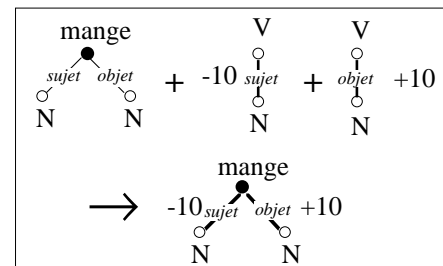
les grammaires₁ TAG, le placement d'un argument quel qu'il soit est toujours assuré par la structure élémentaire de son gouverneur, ce qui oblige à considérer des structures particulières pour les éléments qui ne sont pas en position canonique. Avec le croisement des phénomènes, on obtient pour certains verbes des centaines de structures élémentaires (Abeillé 91, X-TAG Group 95, Candito 99). Nous pensons préférable que le placement de tels éléments soit encodé dans leur propre structure élémentaire. La structure élémentaire d'un clitique argumental indiquera donc le lien avec le gouverneur (par un lien blanc évidemment) avec un poids sur ce lien indiquant son placement. Ce poids devra venir se substituer au poids indiqué dans la structure élémentaire du gouverneur. Ceci sera réalisé à l'aide d'un trait prioritaire, souligné dans la figure, qui viendra remplacer le trait ordinaire de même nature (au lieu de s'unifier avec lui). Notons que la "montée" des clitiques (= clitic climbing) nécessite en fait des structures élémentaires plus complexes pour les clitiques. Une solution de même nature que celle que nous proposons pour les extractions est possible.

Des structures élémentaires pour les clitiques avec liens blancs et traits prioritaires



Avec un tel jeu de règles, une forme verbale pourra ne recevoir qu'une seule structure élémentaire⁵, où les compléments seront placés en position canonique. Les autres placements seront assurés par des règles de placement prioritaires associées au dépendant.

Ce formalisme permet également de construire une grammaire₁ non lexicalisée, assez comparable à une grammaire₁ Sens-Texte (Mel'cuk 88). En effet, l'introduction des arcs blancs permet de dissocier en partie la grammaire₂ du lexique : toute structure élémentaire peut être obtenue par la combinaison de structures comportant un unique élément noir, nœud ou arc.



Une structure élémentaire vue comme combinaison de structures encore plus élémentaires

⁵ D'autres extensions du formalisme seront nécessaires pour traiter les arguments optionnels et les disjonctions de requêtes (une même position argumentale peut être occupée par des éléments de natures assez différentes, comme une infinitive ou une complétive).

Les structures avec un nœud noir construisent un nœud et indique sa valence active, tandis que les structures avec un arc noir, auquel est attaché un poids, contrôlent la linéarisation. Ces dernières, bien qu'écrites dans un formalisme différent, sont d'ailleurs exactement les règles d'une grammaire₂ Sens-Texte (voir Mel'cuk & Pertsov 87 pour une grammaire₂ de l'anglais). (Sur le passage d'une grammaire₁ non lexicalisée à une grammaire₁ lexicalisée, voir également Candito & Kahane 98a et Candito 99.)

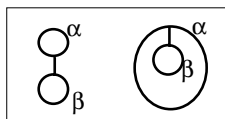
2. Arbres à bulles et représentation des phrases à extractions

Pour représenter syntaxiquement les phrases comportant des extractions, nous utiliserons, une structure plus riche que l'arbre de dépendance, l'arbre de dépendance à bulles.

Définition des arbres à bulles et projectivité

L'utilisation d'arbres à bulles remonte au moins à Tesnière 59 (67 : 46), dont le stemma comprend des bulles appelées nucléus où sont regroupés, pour des raisons différentes des nôtres, des éléments comme l'auxiliaire et le participe, la copule et l'adjectif ou l'article et le nom. La première étude formelle des arbres à bulles se trouve à notre connaissance dans Gladkij 68. Notre étude repose directement sur Kahane 97.

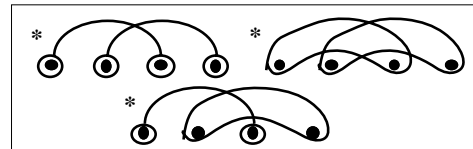
Intuitivement, un **arbre à bulles** est un arbre dont certains nœuds sont occupés par des bulles contenant elles-mêmes des nœuds pouvant à leur tour être occupés par des bulles ou avoir des dépendants. Formellement, un arbre à bulles est un quadruplet $(X, B, \rightarrow, \varphi)$ où X est l'ensemble des **nœuds de base** (correspondant aux mots dans le cas d'une représentation syntaxique), B est l'ensemble des **bulles**, φ est une application qui associe à chaque bulle son **contenu** (qui est une partie non vide de X) et \rightarrow est une **relation de dépendance** sur B (définissant un arbre sur B). On impose de plus que deux bulles soient toujours disjointes ou incluses l'une dans l'autre (ainsi qu'un certain nombre d'autres propriétés; cf. Kahane 97), ce qui nous donne deux situations pour la dépendance entre deux bulles: la dépendance pure et l'emboîtement.



Les deux situations possible pour la dépendance $\alpha \rightarrow \beta$
la dépendance pure et l'emboîtement

Pour simplifier les représentations, la bulle la plus emboîtée de contenu $\{x\}$ est identifiée avec le nœud de base x .

On appelle **projection** d'une bulle α la réunion des contenus des bulles dominées par α , α compris, pour la relation de domination \rightarrow^* obtenue par transitivité de \rightarrow . Un arbre à bulle est dit **linéairement ordonné** s'il est assorti d'un ordre linéaire sur ses nœuds de base. Un arbre à bulles linéairement ordonné est dit **projectif** si la projection de toute bulle forme un segment continu. Comme dans le cas des arbres, la projectivité peut être exprimée plus géométriquement: en particulier, deux arcs, deux bulles ou un arc et une bulle ne se coupent jamais (Gladkij 68, Kahane 97).



Quelques cas de non projectivité
pour un arbre à bulles

Représentation de l'extraction par des arbres à bulles

Certains groupes de mots, que nous appellerons, suivant Tesnière 59, des **nucléus**, peuvent se comporter comme un seul mot du point de vue de certains phénomènes syntaxiques comme l'extraction, mais aussi la cliticisation, les coordinations la négation, la comparaison, etc. Nous distinguerons deux familles de nucléus:

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

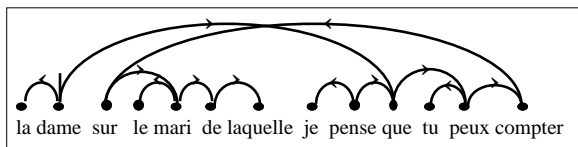
Le traitement de l'extraction des compléments de nom des sujet et objet direct (*la dame dont le mari est mort; la dame dont je connais le mari; *la dame dont je parle au mari*) nécessite de considérer qu'un verbe et son sujet ou un verbe et son complément d'objet direct peuvent former un nucléus verbal.

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

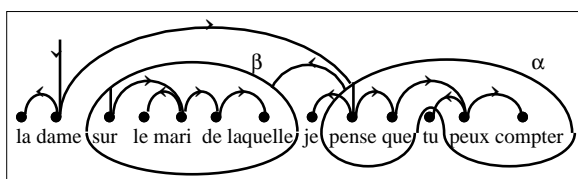
Marie les pompiers.; ?*Jean ne **veut que** tu **parles** à personne).

Dans la suite, α désignera un nucléus verbal et β un nucléus nominal.

On peut représenter les nucléus par des bulles et attribuer, sous certaines conditions, des liens à ces bulles. Nous présentons ci-dessous les représentations du syntagme nominal *la dame sur le mari de laquelle je pense que tu peux compter* par un arbre de dépendance ordinaire et par un arbre à bulles.



Unarbre de dépendance (non projectif)



Un arbre à bulles (projectif)

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

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

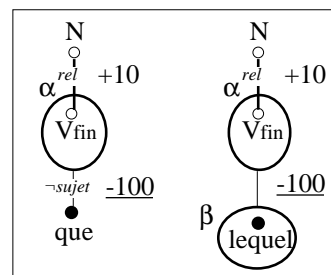
3. Une grammaire₁ à bulles pour l'extraction

Nous allons proposer une grammaire₁ pour la relativisation avant de l'étendre aux autres phénomènes d'extraction.

Structures élémentaires pour les relatives

La structure élémentaire d'un pronom relatif indique que celui-ci appartient à un nucléus β dépendant d'un nucléus α dépendant lui-même de l'antécédent. Plus précisément, c'est le pronom relatif qui introduit dans la structure de représentation les bulles α et β ainsi que le lien entre l'antécédent et la relative (qui sont donc en noir). Un nœud est également prévu dans la

bulle α pour recevoir le verbe principal de la relative, qui doit être fini (V_{fin}) (le fait qu'il est s'agit du verbe principal est indiqué par la dépendance avec α).



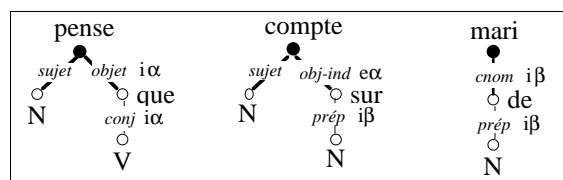
Structures élémentaires pour les pronoms relatifs *que* et *lequel*

On peut encore noter que c'est le pronom relatif qui assure le placement antéposé du nucléus β (par l'intermédiaire du trait prioritaire -100). Par ailleurs, les pronoms *que* et *lequel* se distinguent par le fait que *que* est directement lié au nucléus α et n'appartient pas à un nucléus β (ce qui traduit un caractère plus complémenteur de *que* par rapport à *lequel*).

Nous pensons, comme Tesnière 59, que le pronom relatif joue un double rôle : un rôle de pronom (en remplissant une position dans la relative) et un rôle de translatif ou complémenteur (en autorisant une proposition à venir modifier un nom). Ainsi, dans cette analyse, à la différence par ex. de l'analyse usuelle en TAG (Kroch 87, Abeillé 91, XTAG 95), c'est le pronom qui donne à la relative le pouvoir de s'attacher à l'antécédent. Les grammaires₁ catégorielles propose une analyse comparable à la notre, où le pronom relatif s'attache à la fois à l'antécédent et à la relative (cf. la catégorie $(N/N)/(S \uparrow NP)$ du pronom relatif *which* dans Moortgat (89: 113) signifiant que *which* s'adjoint à un N et requiert un S auquel manque un NP), mais où le pronom ne se combine pas directement avec son gouverneur syntaxique et ne contrôle pas lui-même la formation des nucléus (ce qui rend d'ailleurs l'expression des contraintes sur l'extraction plutôt délicate dans ce formalisme).

Opération de combinaison avec les nucléus

Les opérations de combinaison avec nucléus reposent sur un marquage explicite des liens qui peuvent ou non figurer dans une bulle α ou β .

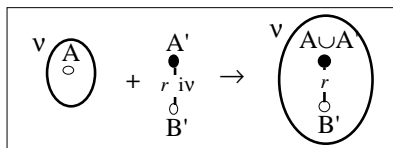


Structures élémentaires avec étiquetage des liens intra et axtranucléaires

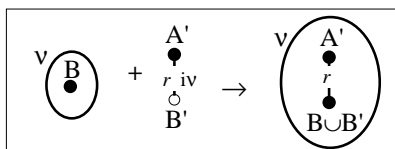
Un lien qui peut être à l'intérieur d'un nucléus α (resp. β) sera étiqueté $i\alpha$ (resp. $i\beta$). Un lien qui peut être attribué à un nucléus α sera étiqueté $e\alpha$.

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

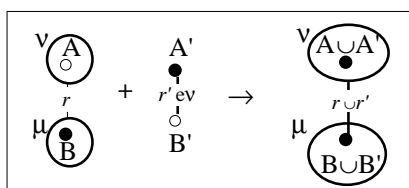
Lorsqu'un nœud A fusionne avec un nœud d'une bulle v, les dépendants de A sont placés par défaut à l'extérieur de v. Par conséquent, pour qu'un lien soit placé dans une bulle, une opération particulière est requise. Les opérations de combinaison suivantes indiquent à quelles conditions un nœud peut être ou non inclus dans une bulle d'un type donné. La règle R1, qui suit, se lit comme ça : si un nœud A' fusionne avec un nœud A inclus dans un nucléus v et si B' dépend de A' par un lien $i\alpha$, B' peut être inclus dans v. Cette règle peut être a priori utilisée pour n'importe quel type de nucléus, mais la description des extractions nécessite uniquement le cas $v = \alpha$. La règle R2 est quasiment symétrique à R1 et sera utilisée avec $v = \beta$. Enfin, la règle R3 concerne le lien entre deux nucléus ; à noter que avant la fusion B n'est pas a priori identifié comme le sommet de μ , mais qu'après la fusion, il est nécessairement le sommet de μ (ce qu'indique le lien entre μ et B).



R1: Fusion avec lien intranucléaire



R2: Fusion avec lien intranucléaire



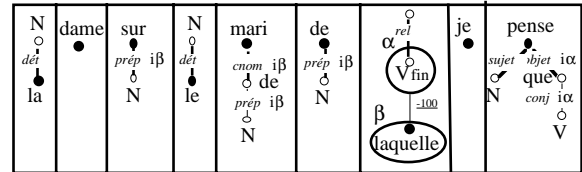
R3: Fusion avec lien extranucléaire

Dans notre représentation par arbre à bulles, nous perdons le fait que le lien $e\alpha$ promu sur le nucléus v par l'opération R3 est à l'origine issu de A'. On pourrait garder cette information par un enrichissement du formalisme de représentation en "reliant" tout lien d'un nucléus à un nœud intérieur.

Avant de commenter la portée des opérations précédentes, nous allons les illustrer par l'analyse incrémentale du syntagme :

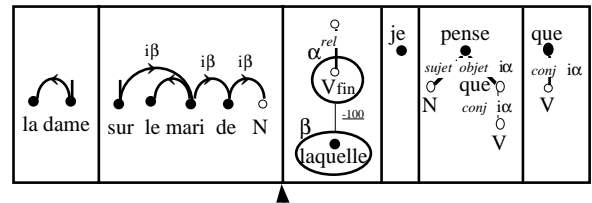
(1) *la dame sur le mari de laquelle je pense que tu peux compter*

Pour chaque mot du syntagme, une structure élémentaire est chargée ; dès qu'une réduction est possible, celle-ci est effectuée.



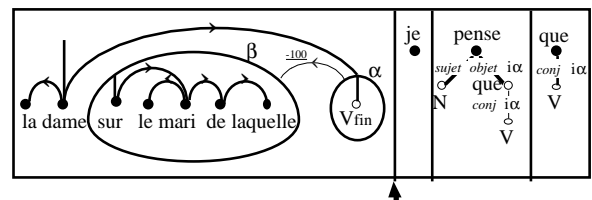
Etape 0: la pile avant réduction

Lorsque le pronom relatif *laquelle* est considéré, le segment initial de la pile est réduit à deux cases contenant respectivement *la dame* et *sur le mari de N*.



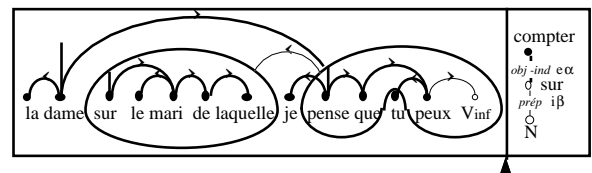
Etape 6: la pile en cours de réduction

Le nœud ancré par *laquelle* peut fusionner avec la requête de *sur le mari de N*. Comme tous les liens de la chaîne *sur le mari de N* sont de type $i\beta$, l'ensemble de la chaîne sera incluse dans la bulle β (R2), permettant au nœud supérieur de la structure de *laquelle* de fusionner avec *dame* (tout en respectant les contraintes imposées par la projectivité).



Etape 7: la pile en cours de réduction

Le verbe *pense* pourra, dès qu'il est considéré, venir fusionner avec le nœud V de la bulle α , et la requête *que V de pense* sera incluse dans la bulle α (R1).



Etape 12 : la pile avant la dernière réduction

Lorsque *compter* viendra prendre sa place dans la bulle α , la requête *sur N* pourra venir s'unifier avec la tête

de la bulle β (R3), donnant ainsi un arbre à bulles entièrement noir. On voit que l'analyse ne réussit que si l'un des verbes du nucléus α prend le sommet du nucléus β dans sa valence active.

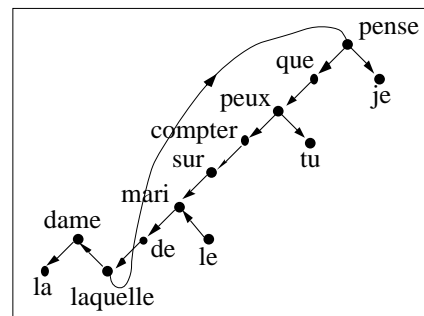
Commentaires sur les opérations

Dans notre description des extractions, le groupe extrait, représenté par un nucléus β , ne dépend pas de son gouverneur syntaxique, mais de la bulle α (nous appelons gouverneur syntaxique de β le nœud du nucléus α qui a introduit le lien entre α et β et qui serait le gouverneur dans une représentation par un arbre de dépendance ordinaire). Or la projection du nucléus α est l'ensemble de la proposition relative et le nucléus α domine directement le nucléus β et le verbe principal de la relative: notre représentation est donc tout à fait comparable à la représentation de la Syntaxe X' (Chomsky 77, Jackendoff 77), où la proposition relative correspond à un nœud S' dans la projection duquel on trouve un nœud COMP pour le groupe extrait et un nœud S pour le reste de la relative. Par ailleurs, nos opérations R1, R2, R3 peuvent être comparées aux opérations considérées par les grammaires₁ basées sur la Syntaxe X'. Prenons le cas de HPSG (Pollard & Sag 94): l'opération R3 correspond à l'introduction d'un trait non local (ou trait Slash), tandis que l'opération R1 correspond à la propagation de ce trait. Plusieurs descriptions proposées dans le cadre des grammaires₁ de dépendance (Hudson 90, Lombardo & Lesmo 98, Kahane, Nasr & Rambow 98) utilisent un appareillage équivalent au trait Slash de HPSG, permettant de remonter le groupe extrait sur le verbe principal de la proposition où a lieu l'extraction. Notre analyse diffère de ces analyses sur au moins un point : nous ne propageons pas le contenu du nucléus β (ou du groupe extrait) ; nous l'attachons simplement au nucléus α et c'est l'appartenance au nucléus α qui se propage. D'autre part, nous utilisons des opérations similaires pour décrire l'extraction proprement dite (le nucléus α) et le phénomène dit du pied-piping (le nucléus β), contrairement, par ex., à TAG (qui utilise la propagation de traits pour le pied-piping) ou aux grammaires₁ catégorielles (où le pied-piping n'est pas traité à ma connaissance).

Graphe de dérivation

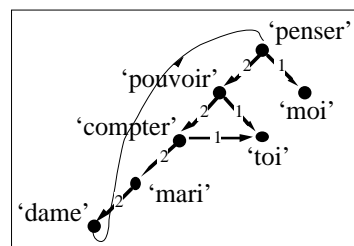
De la même façon que cela est fait en TAG, on peut associer à l'analyse d'une phrase par notre grammaire₁ un **graphe de dérivation** (en TAG, il s'agit d'un arbre). Chaque nœud du graphe de dérivation représente une structure élémentaire et chaque lien la combinaison de deux structures élémentaires. Chaque lien est dessiné avec son gouverneur syntaxique en haut et son dépendant syntaxique en bas. De plus, chaque lien est orienté par une flèche qui indique

laquelle des deux structures a construit le lien syntaxique qui les unit. En conséquence, une flèche vers le bas correspond à une substitution et une flèche vers le haut à une adjonction.



Le graphe de dérivation de (1)

On note que le pronom relatif *laquelle* possède deux gouverneurs: il s'adjoint à l'antécédent *dame* et se substitue dans *de*. Quant à la flèche de *laquelle* à *pense*, elle correspond à la substitution de *pense* dans la structure élémentaire de *laquelle* (il s'agit d'une flèche vers le bas, bien que *pense* soit placé au-dessus de *laquelle*, puisqu'il domine par ailleurs *laquelle*). Cette analyse n'est pas sans rappeler l'analyse de Tesnière 59 (67 : 560), où les trois mêmes dépendances sont considérées pour un pronom relatif. Un graphe assez similaire est obtenu dans un formalisme à la TAG proposé par Candito & Kahane 98b. On peut encore noter que ce graphe de dérivation est un pas certain vers une représentation sémantique : en effet, si l'on collapse avec leur voisin les nœuds qui correspondent à la même unité sémantique (c'est-à-dire *que* avec *pense*, *sur* avec *compter*, *de* avec *mari* et aussi *laquelle* avec *dame*), on obtient une **représentation sémantique** à la **Sens-Texte** (Mel'cuk 88, Kahane & Mel'cuk à paraître)⁶ (cf. Candito & Kahane 98a pour l'interprétation de l'arbre de dérivation TAG en ce sens).



La représentation sémantique Sens-Texte de (1)

Les nœuds de la représentation Sens-Texte correspondent aux unités sémantiques du syntagme. Chaque unité sémantique peut être vue comme un

⁶ Le seul arc, après collapsage, qui ne correspond pas à une dépendance sémantique est l'arc entre *dame-laquelle* et *pense-que*. Bien qu'il ne corresponde effectivement pas à une relation prédicat-argument, ce lien, que l'on retrouve dans la théorie Sens-Texte au niveau syntaxique profond, traduit le fait que la configuration communicativement dominée par *pense* vient modifier *dame*.

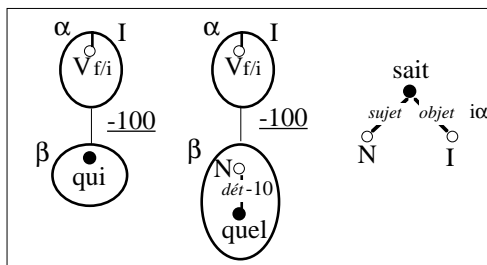
prédicat pointant vers ses arguments : on obtient ainsi un **graphe** étiqueté (les étiquettes des arcs représentant les relations prédicat-argument ou **dépendances sémantiques** indique simplement le numéro d'actant). A ce graphe vient se superposer une hiérarchie sur les nœuds, que nous appellerons, suivant Polguère 90, **hiérarchie communicative**, qui exprime l'importance communicative des sens exprimés dans la phrase et qui se retrouve encodée par la structure d'arbre au niveau syntaxique. Dans la tradition Sens-Texte, seuls les **sommets hiérarchiques** (appelés nœuds dominants chez Polguère 90) de certains sous-graphes (notamment le rhème et le thème) sont indiqués. Nous préférons indiquer toute la hiérarchie communicative. Lorsque la dépendance communicative correspond à une dépendance sémantique, nous l'encodons simplement en plaçant le gouverneur communicatif au dessus de son dépendant communicatif, sinon nous l'indiquons par une flèche fine (comme entre 'dame' et 'penser'). Les règles de passage de l'arbre dérivation à la représentation sémantique doivent indiquer la correspondance entre un prédicat et ses arguments et une configuration syntaxique.

Les autres types d'extraction

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

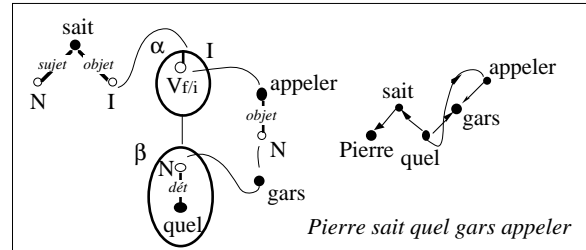
A l'instar de notre analyse des relatives, nous allons 1) assurer l'introduction des nucléus par l'élément qui permet l'extraction (le pronom interrogatif, le cliveur *c'est...que/qui* ou la topicalisation) et 2) chercher à obtenir un graphe de dérivation proche d'une représentation sémantique.

A l'instar d'un pronom relatif, un pronom interrogatif indirect introduit un nucléus β dépendant d'un nucléus α dont le sommet est un verbe fini ou infinitif.



Structures élémentaires pour les pronoms interrogatifs indirects *qui* et *quel* et pour le verbe "interrogatif" *savoir*

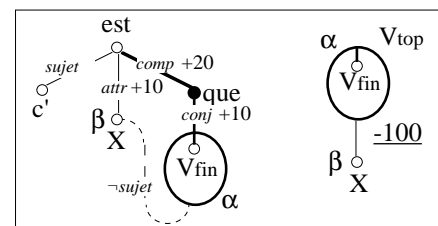
C'est le nucléus α qui va venir fusionner avec la requête du verbe "interrogatif" ; le nucléus reçoit donc une étiquette comme n'importe quel nœud lexical: il s'agit ici d'une étiquette ad hoc I que l'on retrouvera comme requête d'un verbe "interrogatif" (comme *savoir*).



Dérivation d'une interrogative indirecte

D'autre part, en attribuant au lien *objet* d'un verbe "interrogatif" l'étiquette $i\alpha$, on autorise ce lien à appartenir à un nucléus α , ce qui permet l'analyse d'une double extraction du type *les pommes que je sais à qui donner* ou *le livre dont Pierre sait quelle page il a arrachée*.

Pour la topicalisation, on utilisera une structure élémentaire similaire à celle des pronoms interrogatifs, mais sans ancre lexicale. Cette structure introduit simplement une bulle α et permet donc, par le biais de l'opération R3, à un complément d'un des verbes du nucléus α d'être topicalisé. Un nucléus α dont un complément est topicalisé reçoit l'étiquette $V_{top} = [cat:V, top:+, \dots]$. Par le biais du trait *top*, tout nœud verbal d'une structure élémentaire peut spécifier s'il accepte de recevoir ou non un verbe (ou nucléus verbal) dont un complément a été topicalisé. En particulier selon que le nœud V sommet du nucléus introduit par une topicalisation est ou non $top:+$, on bloquera ou non la double topicalisation.



Structures élémentaires pour le clivage et la topicalisation

Enfin, la description du clivage nécessite un léger enrichissement du formalisme : en effet, le cliveur *c'est...que* introduit une bulle α dont l'un des compléments devient un complément du cliveur (ce qu'indique notre lien courbe en pointillé). Ce lien pointillé, qui n'est pas une dépendance syntaxique, devra néanmoins fusionner avec un lien noir pour que la structure soit saturée. Cette fusion sera réalisée par une variante sans surprise de l'opération R3.

Conclusion

Bien que notre grammaire₁ se situe à la croisée de différents courants en linguistique formelle (notamment TAG, HPSG, les grammaires₁ catégorielles et les grammaires₁ de dépendance), nous pensons qu'elle permet une description originale des extractions. Cette description présente différents avantages. Premièrement, la représentation par bulle donne à l'outil utilisé pour traiter les dépendances une véritable existence *géométrique* (à la différence, par ex., des grammaires₁ utilisant la propagation d'un trait non local). Deuxièmement, notre analyse peut être intégrée à une grammaire₁ lexicalisée (bien que la lexicalisation ne soit pas une nécessité, cf. fin de la Section 1) sans provoquer la multiplication des structures élémentaires. Enfin, le nucléus (et les opérations de combinaison qui vont avec) n'est pas une notion ad hoc, puisqu'il intervient dans la description de nombreux autres phénomènes syntaxiques, dont les coordination, la négation et la comparaison. L'extension de la grammaire à ces phénomènes est en cours.

Références

- ABEILLE Anne, 1991, *Une grammaire lexicalisée d'Arbres Adjoints pour le français*, Thèse de Doctorat, Univ. Paris 7.
- ABEILLE Anne & GODARD Danièle, 1998, "French word Order and lexical Weight", in Borsley (ed), *Syntactic Categories, Syntax and Semantics*, NY : Acad. Press.
- AHO Alfred V. & ULLMAN Jeffrey D., 1972, *The Theory of Parsing, Translation and Compiling, Vol.1*, London : Prentice-Hall.
- BAR-HILLEL Y., 1953, "A quasi-arithmetical notation for syntactic description", *Language* 29:1, 47-58.
- BLINKENBERG A., 1928, "L'ordre des mots en français moderne", *Bianco Lunos Bogtrykkeri, Historik-filologiske Meddelelser XVII:1 et XX:1*, Kopenhagen.
- CANDITO Marie-Hélène, 1999, *Organisation modulaire et paramétrable de grammaires éle-ctroniques lexicalisées. Application au français et à l'italien*, Thèse de Doctorat, Univ. Paris 7.
- CANDITO Marie-Hélène & KAHANE Sylvain, 1998a, "Un grammaire TAG vue comme une grammaire Sens-Texte précompilée", *TALN'98*, Paris, 24-34.
- CANDITO Marie-Hélène & KAHANE Sylvain, 1998b, "Defining DTG derivations to get semantic graphs", *TAG+4 Workshop*, Philadelphie, 25-28.
- CHOMSKY Noam, 1977, "On Wh-movement", in P. Culicover et al. (eds), *Formal Syntax*, Dordrecht : Reidel.
- COURTIN & GENTHIAL Damien, 1998, "Parsing with dependency relations and robust parsing", in Kahane & Polguère (eds), *Workshop on dependency-based grammars*, COLING/ACL'98, Montréal.
- GLADKIJ Aleksej V., 1966, *Leckii po matematiceskoj lingvistike dlja studentov*, Novosibirsk : NGU (trad. franç. : *Leçons de linguistique mathématique*, fasc. 1, 1970, Paris: Dunod).
- GLADKIJ Aleksej V., 1968, "On describing the syntactic structure of a sentence" (en russe avec résumé en angl.), *Computational Linguistics*, 7, Budapest, 21-44.
- HUDSON Richard A., 1990, *English Word Grammar*, Oxford: Blackwell.
- IORDANSKAJA Lidija, 1963, "O nekotorykh svojstvax pravil'noj sintaksiceskoj struktury (na materiale russkogo jazyka)" [On some properties of correct syntactic structure (on the basis of Russian)], *Voprosy jazykoznanija*, 4, 102-12.
- JOSHI Aravind K., 1987, "Introduction to Tree Adjoining Grammar", in Manaster Ramer (ed.), *The Mathematics of Language*, Amsterdam : Benjamins, 87-114.
- JACKENDOFF Ray, 1977, *X' Syntax: A Study of Phrase Structure*, Linguistic Inquiry Monograph, MIT Press.
- KAHANE Sylvain, 1997, "Bubble trees and syntactic representations", in Becker & Krieger (eds), *Proc. 5th M. of Mathematics of Language (MOL'5)*, Saarbrücken : DFKI, 70-76.
- KAHANE Sylvain & MEL'CUK Igor, à paraître, "La synthèse des phrases à extraction en français contemporain. Du réseau sémantique à l'arbre syntaxique", *T.A.L.*, 61p.
- KAHANE Sylvain, NASR Alexis & RAMBOW Owen, 1998, "Pseudo-projectivity: a polynomially parsable non-projective dependency grammar", *ACL-COLING'98*, Montréal, 646-52.
- LECERF Yves, 1961, "Une représentation algébrique de la structure des phrases dans diverses langues naturelles", *C. R. Acad. Sc. Paris*, 252, 232-234.
- KROCH Anthony, 1987, "Unbounded dependencies and subadjacency in a Tree Adjoining Grammar", in A. Manaster Ramer (ed.), *The Mathematics of Language*, Amsterdam : Benjamins, 143-71.
- LOMBARDO Vincenzo & LESMO Leonardo, 1998, "Formal aspects and parsing issues of dependency theory", *COLING'98*, Montréal, 787-93.
- MEL'CUK Igor, 1988, *Dependency syntax: theory and practice*, NY : State Univ. of NY Press.
- MEL'CUK Igor & PERTSOV Nicolas, 1987, *Surface syntax of English – A formal model within the Meaning-Text framework*, Amsterdam : Benjamins.
- MOORTGAT Michael, 1989, *Categorial Investigations. Logical and Linguistic Aspects of the Lambek Calculus*, Dordrecht : Foris.
- NASR Alexis, 1995, "A formalism and a parser for lexicalised dependency grammars", *4th Int. Workshop on Parsing Technologies*, State Univ. of NY Press.
- NASR Alexis, 1996, *Un modèle de reformulation automatique fondé sur la Théorie Sens-Texte – Application aux langues contrôlées*, Thèse de Doctorat, Univ. Paris 7.
- POLGUÈRE Alain, 1990, *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*, Thèse de Doctorat de l'Univ. de Montréal.
- POLLARD Carl & SAG Ivan A., 1994, *Head-Driven Phrase Structure Grammar*, CSLI series, Chicago : Univ. of Chicago Press.
- XTAG Research Group, 1995, *A Lexicalized TAG for English*, Technical Report IRCS 95-03, Univ. of Pennsylvania, (version mise à jour sur le web).
- TESNIERE Lucien, 1959 (1967), *Éléments de syntaxe structurale*, Paris : Klincksieck, 670 p.