

Hierarchisation d'analyses basée sur des informations dépendanciennes dans le cadre des LTAGs

Alexandra Kinyon

TALANA, UFR de Linguistique,
Université Paris 7, Case 7003
2, place Jussieu, 75251 Paris Cedex 05
Alexandra.Kinyon@linguist.jussieu.fr

Résumé

Depuis [Kimball 73], les préférences d'attachement telles que "l'association droite" et "l'attachement minimal" ont essentiellement été formulées en termes d'arbres de constituants (e.g. forme, nombre de noeuds ...).

Nous présentons 2 principes de préférence d'attachement formulés en termes d'arbres de dérivation (i.e. d'information dépendancielle) dans le cadre du formalisme des Grammaires d'Arbres Adjoints Lexicalisées (LTAG). Nous montrons pourquoi ce type d'approche permet de remédier aux défauts des approches structurales exprimées en termes d'arbres de constituants et rendent compte d'heuristiques largement acceptées (i.e. argument / modifieur, idiomes).

1. Motivation

La formalisation de principes de préférence d'attachement en analyse intéresse à la fois les psycholinguistes, qui tentent de rendre compte des processus d'analyse chez l'humain et se focalisent sur la notion de "principes syntaxiques universels" et les linguistes computationnels, qui tentent de réduire le non-déterminisme en analyse automatique, et de ce fait adoptent une approche plus pragmatique (i.e. utilisation d'heuristiques).

Les caractéristiques inhérentes au formalisme LTAG (i.e. la lexicalisation, l'adjonction, et un domaine de localité étendu), ainsi que son appartenance à la famille des formalismes "légèrement sensibles au contexte" [Joshi 85] rendent ce dernier utile au domaine du traitement automatique des langues : le formalisme LTAG permet une analyse dans le pire des cas en temps polynomial [Vijay-Shanker 87],[Schabes 90] et permet également de modéliser de façon psycholinguistiquement pertinente le langage humain¹. Tout cela a rendu possible le développement de grammaires à large couverture, pour le français [Abeillé 91], pour l'anglais [Xtag group 95] ainsi que la mise au point d'outils de génération semi-automatique et de maintenance de ces grammaires [Candito 96]. Hélas, l'utilisation de grammaires à large couverture engendrent des taux d'ambiguïté élevés : pour l'anglais [Doran & al. 94] obtiennent 7.46 analyses / phrase sur un corpus de 18 730 phrases du "Wall Street Journal". Pour le français, [Abeillé & Candito 99] obtiennent 2.96 analyses / phrase sur un jeu de 1643 phrases test grammaticales issues de TSNLP [Estival & Lehman 97]. Il faut donc faire baisser ces taux, tant d'un point de vue pratique (i.e. sorties d'analyses trop nombreuses et donc inexploitable) que théorique (i.e. nombre de ces ambiguïtés ne sont pas perçues par l'humain). Une tentative en ce sens est présentée dans [Srinivas & al. 95], qui formulent des

¹ e.g. [Frank 92] montre que l'opération d'adjonction est pertinente dans le cadre de l'acquisition du langage chez les enfants, [Joshi 90] discute la pertinence psycholinguistique du formalisme concernant les phénomènes de dépendances croisées et enchâssées.

heuristiques indépendantes du domaine pour "hiérarchiser" les analyses (i.e. parse ranking). Mais cette approche concerne uniquement l'anglais, est essentiellement pragmatique, n'établit pas le lien avec les résultats psycholinguistiques existants et n'exploite pas la notion de "dérivation".

Pour réduire le taux d'ambiguïté, on peut distinguer trois types d'approches syntaxiques² : les approches "probabilistes", les approches "lexicales" et les approches "structurales".

Dans une première partie, nous présentons le formalisme LTAG. Puis nous discutons chacune des ces trois approches. Enfin, nous montrons pourquoi il est souhaitable de formuler des principes structuraux en termes d'arbre de dérivation (i.e. informations dépendancielle) plutôt qu'en termes d'arbres dérivés (i.e. constituants). Enfin, nous présentons les principes que nous utilisons, expliquons ce dont ils rendent compte et communiquons quelques résultats préliminaires obtenus pour le français.

2. Présentation du formalisme LTAG

Une LTAG consiste en un ensemble fini d'arbres élémentaires de hauteur finie. Chaque **arbre élémentaire** ancre un ou plusieurs items lexicaux. L'ancre principale est appelée **tête**, les ancres secondaires **co-têtes**. Toutes les feuilles d'un arbre élémentaire sont soit **ancre**, soit **noeud pied** (noté *), soit **noeud à substituer** (noté ↓). Tout arbre élémentaire comporte au plus un noeud pied, de même catégorie que sa racine. Les arbres élémentaires sont de deux types : **initial** ou **auxiliaire**³. Un arbre auxiliaire comporte exactement un noeud pied. Les arbres élémentaires qui ne sont pas auxiliaires sont initiaux. Les arbres élémentaires se combinent à l'aide de 2 opérations : la **substitution**, et **l'adjonction**. La substitution est obligatoire et utilisée essentiellement pour les arguments (i.e. sujets, compléments verbaux et nominaux). Une substitution consiste à remplacer dans un arbre (élémentaire ou non) un noeud à substituer X par un arbre initial de racine de même catégorie que X. L'adjonction est facultative (mais peut être rendue obligatoire ou interdite par l'ajout de contraintes spécifiques) . Elle concerne essentiellement les déterminants, les modificateurs, les auxiliaires et modaux, les verbes à montée (e.g. sembler) ainsi que certains compléments phrastiques. Une adjonction consiste à insérer dans un arbre à la place d'un noeud X un arbre auxiliaire de racine de même catégorie que X. Les descendants de X deviennent alors les descendants du noeud pied de l'arbre auxiliaire. Contrairement aux règles de réécriture hors-contexte, l'historique des dérivations doit être explicité car un même arbre dérivé (i.e. arbre de constituants) peut être obtenu par des dérivations distinctes. C'est pour cela qu'avec LTAG on obtient un **arbre de dérivation**, encodant des informations dépendancielle, et à partir duquel on construit un **arbre dérivé**. Les branches d'un arbre de dérivation ne sont pas ordonnées. La figure 1 illustre les arbres élémentaires ancrés lors de l'analyse de "Aujourd'hui Jean brise la glace"⁴, ainsi que les arbres de dérivations correspondant à l'interprétation figée ainsi qu'à l'interprétation littérale de la phrase. La figure 1 montre également que les deux arbres de dérivation engendrent le même arbre dérivé⁵.

En outre, des principes de bonne formation des arbres élémentaires ont été formulé [Abeillé 91][Frank 92] [Abeillé & Candito 99]:

- Principe de cooccurrence prédicat-argument : un arbre élémentaire comporte un noeud feuille pour chaque argument réalisé de la tête d'un arbre élémentaire.

² Nous réalisons l'importance des facteurs sémantiques et/ou contextuels pour la désambiguïté [Crain & Steedman 85]. Toutefois, cela dépasse le cadre de ce travail.

³ Traditionnellement, les arbres initiaux sont nommés α , les arbres auxiliaires β .

⁴ Tous nos exemples reprennent les analyses linguistiques présentées dans [Abeillé 91], excepté pour les complétives que nous substituons lorsqu'il n'y a pas d'extraction. Aussi nous n'utilisons pas de noeud VP ni de traces. Mais cela n'a aucune incidence sur nos principes de préférence.

⁵ Dans un arbre de dérivation, un trait plein indique une adjonction, un trait en pointillés une substitution.

Hierarchisation d'analyses pour les LTAGs

- Consistance sémantique : un arbre élémentaire ne peut être sémantiquement vide.
- Minimalité sémantique : un arbre élémentaire correspond au plus à une unité sémantique

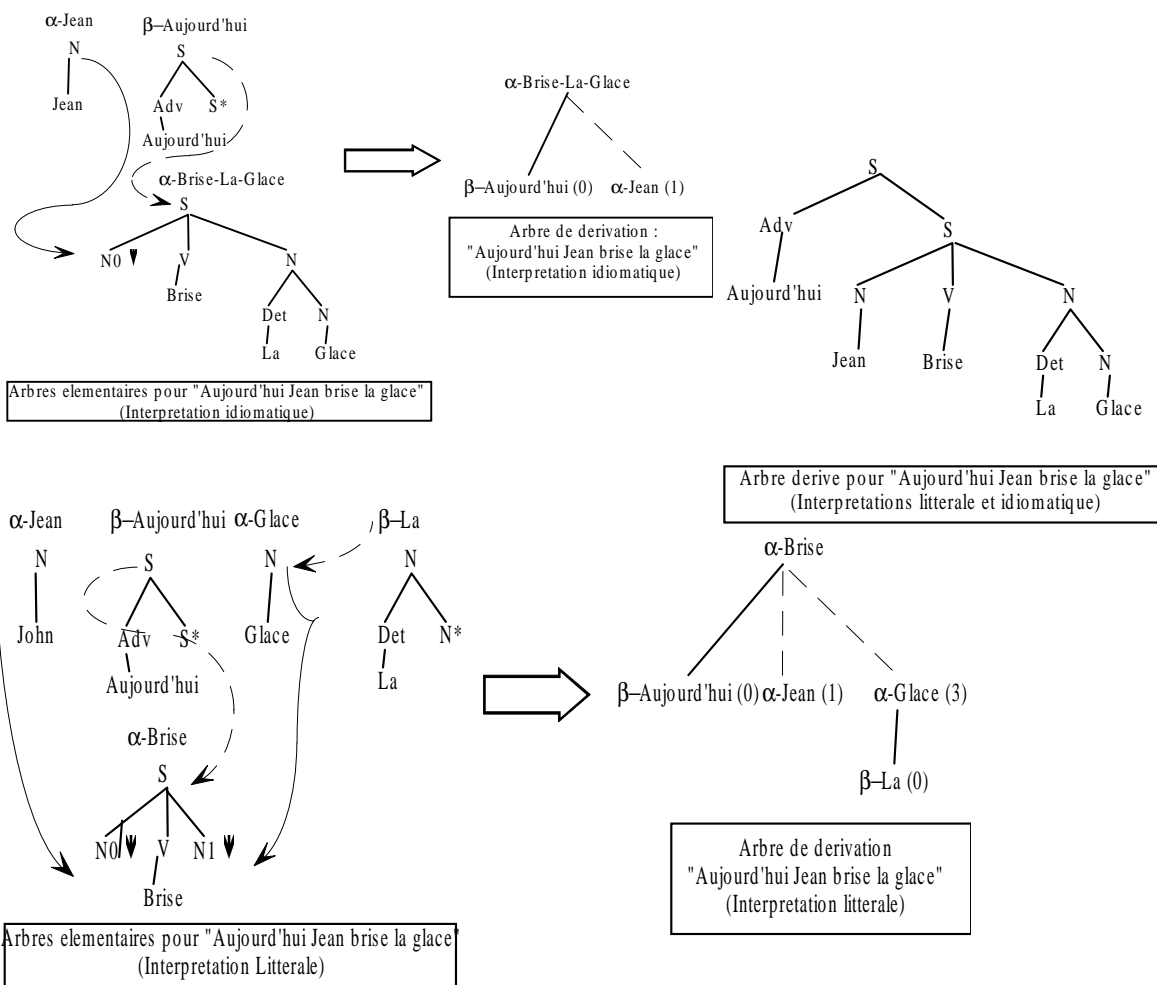


FIGURE 1 : Illustration d'une LTAG et du Principe 1 (idiomes vs interprétation littérale)

3. Comparaison entre approches probabilistes, lexicales et structurales

Nous avons vu précédemment que l'on peut recourir à des approches probabilistes, lexicales ou structurales pour réduire le nombre d'analyses syntaxiques associées à chaque phrase.

3.1 Approches probabilistes

Une approche probabiliste "stricte" consisterait à recourir à un large corpus arboré de référence (annoté manuellement). Lorsqu'une nouvelle phrase est analysée, on conserve la structure syntaxique la plus probable (i.e. la plus fréquente dans le corpus de référence). Cette approche est discutable d'un point de vue pratique : il est coûteux d'obtenir un corpus arboré de référence et cette approche n'est pas indépendante de la langue ni du domaine. Et d'un point de vue théorique, cette approche n'a pas de pouvoir explicatif.

3.2 Approches lexicales

Une approche lexicale consiste à établir à partir de données empiriques des préférences de sous-catégorisation (e.g. pour les verbes). Il est largement admis que ce type de préférence

est crucial [Hindle & Rooth 93]. Toutefois, d'un point de vue pratique les inconvénients sont les mêmes que pour les approches probabilistes strictes (i.e. coût des larges corpus, dépendance à la langue et au domaine). D'un point de vue théorique, 3 problèmes essentiels demeurent :

Le traitement des mots inconnus : Dans "Il grutine à gare", "à la gare" sera préféré comme argument plutôt que comme modifieur.

L'interaction entre deux sous-catégorisations préférées : si "remercier N1 de N2" et "organisateur de N1" sont sous-catégorisations préférées, on préférera cependant attacher "de la manifestation" à "organisateur" dans "L'homme remercie l'organisateur de la manifestation".

L'interaction avec d'éventuelles préférences structurales : Bien que "put N1 in N2" soit un cadre fréquent de sous-catégorisation pour "put" une phrase telle que "I put the book that you were reading in the library"⁶ semble incomplète, bien que syntaxiquement correcte.

Approches lexicales et approches structurales ne sont donc pas nécessairement antagonistes, mais plutôt complémentaires : il existe des préférences lexicales, mais il semble exister également des principes plus généraux d'ordre structural.

3.3 Approches structurales

Concernant les approches structurales [Kimball 73] a introduit le principe d'association droite (RA pour *right association*), qui rend compte de l'interprétation préférentielle de la phrase ambiguë (a) ("yesterday" attaché à "left" plutôt qu'à "said) et [Frazier & Fodor 78] celui d'attachement minimal (MA pour *minimal attachment*), rendant compte de l'interprétation préférentielle de (b) ("for Sue" attaché à "bought" plutôt qu'à "flowers"). Ces principes sont basés sur la forme d'arbres de constituants.

- (a) Tom said that John left yesterday (*Tom a dit que John partait hier*)
- (b) Tom bought the flowers for Sue (*Tom a acheté les fleurs pour Sue*)

D'un point de vue théorique, ce type d'approches a soulevé de nombreuses critiques : elles reposent lourdement sur la *forme* des arbres utilisés. L'interaction entre ces principes est peu claire. Il est difficile d'y intégrer des facteurs sémantiques et/ou pragmatiques [Schubert 84]. La distinction entre modifieurs et arguments [Ferreira & Clifton 86] et l'influence de facteurs de fréquence ne sont pas pris en compte. En outre, ces travaux portent essentiellement sur l'anglais et les principes formulés ne sont pas universels : des résultats expérimentaux infirment la validité du principe RA pour l'espagnol [Cuetos & Mitchell 88] et le hollandais [Brysaert & Mitchell 96].

Par ailleurs, ce type d'approche ne permet pas de rendre compte de préférences syntaxiques largement acceptées : par exemple il est établi que l'interprétation idiomatique d'une phrase est préférée à son interprétation littérale [Abeillé 95] [Gibbs 85][Gibbs & Nayak 89]. De plus, on préfère les arguments aux modifieurs [Abney 89][Britt & al. 92]. S'il on ajoute à cela l'influence des facteurs lexicaux évoqués précédemment, il est saisissant que ces 3 types les plus largement acceptés de préférence syntaxique se révèlent impossible à formaliser en termes d'arbres de constituants.

En pratique toutefois, ce type d'approche est très peu coûteux : les principes sont simple à implémenter, indépendants du domaine et de la langue et il n'est pas nécessaire de disposer de large corpus ...

Dans ce qui suit nous tentons de pallier aux problèmes théoriques liés aux approches structurales que nous venons d'évoquer. Nous traitons les préférences "idiomatique" vs

⁶ Je laisse le livre que tu lisais dans la bibliothèque

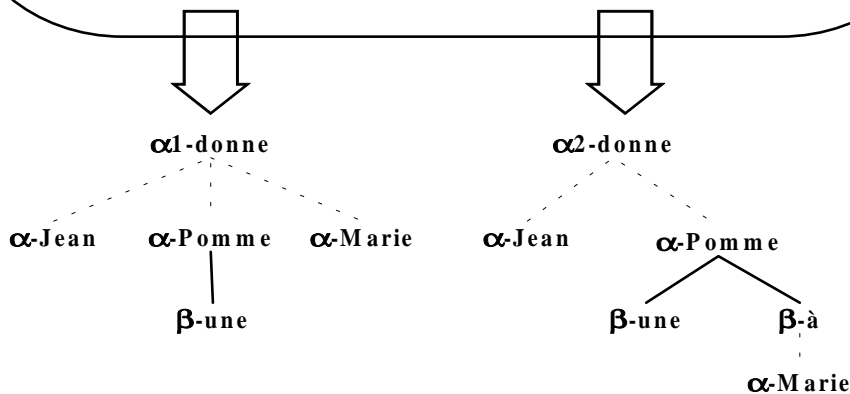
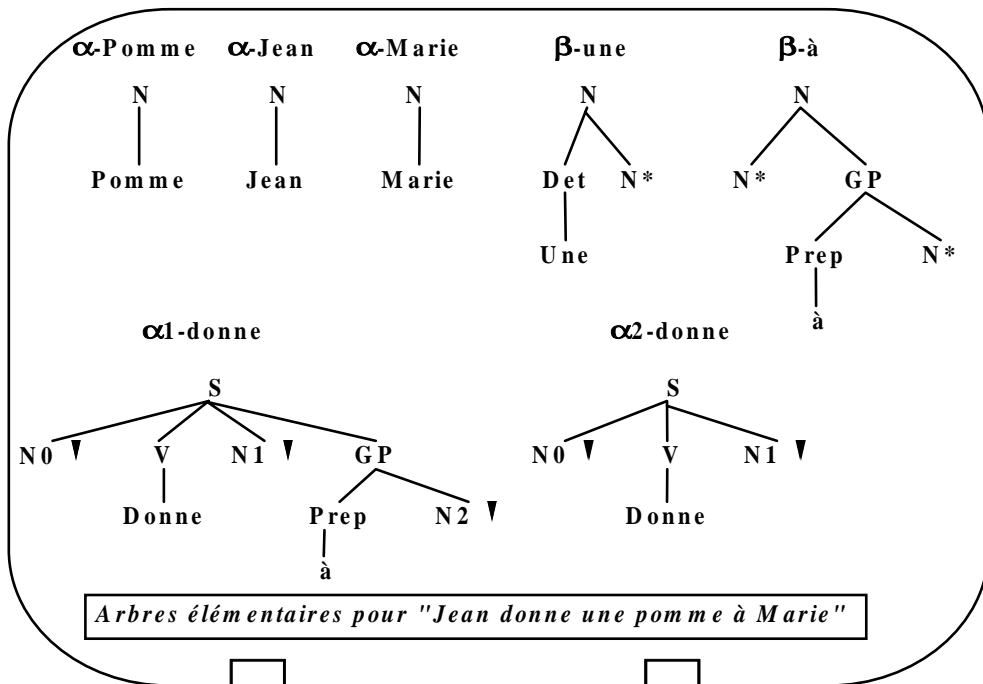
"littéral" ainsi que "argument" vs "modifieur", laissant de coté les préférences d'ordre lexical (bien que le formalisme LTAG soit bien adapté pour exprimer ce type de préférence).

4. Deux principes basés sur les arbres de dérivation

Pour remédier à certains défauts des approches basées sur les arbres de constituants, nous formulons 2 principes de préférence d'analyse en termes d'arbre de dérivation :

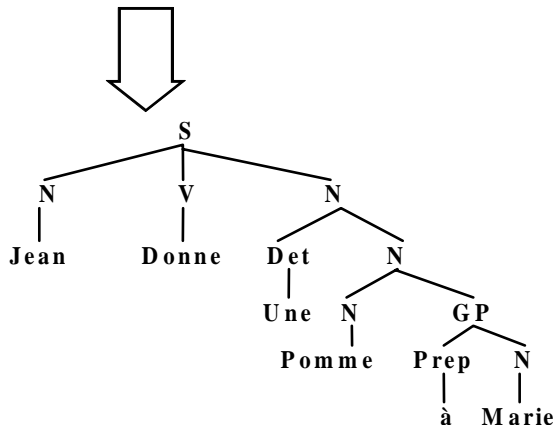
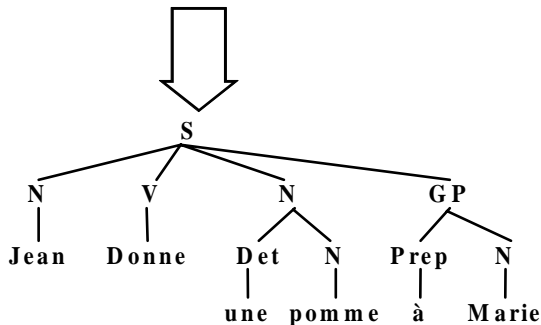
- 1 Préférer l'arbre de dérivation comportant le moins de noeud
- 2- Préférer l'attachement bas d'un arbre α dans un arbre de dérivation

Le principe 1 l'emporte sur le principe 2



Arbre de dérivation 1 (préféré)

Arbre de dérivation 2 (non préféré)



Arbre dérivé 1 correspondant

Arbre dérivé 2 correspondant

FIGURE 2 : Illustration du Principe 1 (arguments vs modifieurs)

4.1. Ce dont ces principes rendent compte

Le principe 1 postule l'existence d'un principe universel d'économie en favorisant l'analyse nécessitant le moins d'opérations. Il permet de capturer la préférence que nous avons pour l'interprétation idiomatique d'une phrase : avec LTAG tous les éléments figés d'une expression idiomatique sont représentés dans un même arbre élémentaire. La figure 1 illustre les dérivations obtenues pour l'analyse de "Aujourd'hui Jean brise la glace". L'arbre de dérivation qui correspond à l'interprétation idiomatique comporte moins de noeuds que celui correspondant à l'interprétation littérale. Ce principe d'économie permet également de capturer la préférence argument vs modifieur : par exemple pour (c) "à Marie" peut être modifieur, mais ce n'est pas l'interprétation préférentielle. La Figure 2 illustre comment le Principe 1 prédit l'arbre de dérivation préféré dans ce cas (i.e. "à Marie" COI de "donne") en favorisant l'arbre de dérivation comportant le moins de noeuds. Une version préliminaire de ce travail comportait une spécialisation de ce principe d'économie, stipulant qu'à nombre égal de noeud, on préférera l'arbre de dérivation comportant le moins de noeud "arbre β ", ce qui revient à dire que l'opération d'adjonction est plus coûteuse que l'opération de substitution [Kinyon 99]. Cela permet de rendre compte de la préférence argument vs modifieur dans le cas d'un verbe transitif (i.e. "Jean peint le matin" \rightarrow "le matin" est préféré comme argument de "peint"). Toutefois, cette "spécialisation" n'a pas donné de bons résultats d'un point de vue empirique, notamment du fait d'interaction avec des critères lexicaux et sémantiques (par exemple en favorisant "est" comme copule plutôt qu'auxiliaire). Nous avons donc laissé de côté cette "spécialisation" du principe d'économie dans l'évaluation empirique⁷.

Le principe 2 est lui d'ordre structural. Il dit que lorsqu'un argument peut être attaché haut (eg. au verbe principal) ou bas (eg au COD de ce même verbe), l'attachement bas est préféré. En (d1) "de cette manifestation" est attaché à "organisateur" plutôt qu'à "soupçonne" tandis qu'en (d2) "de cette manifestation" est forcément attaché à "soupçonne" puisque "Jean" ne sous-catégorise pas de GP (figure 2). Le principe 2 permet également de favoriser la bonne analyse pour (e) : "A qui" sera attaché à "dit" plutôt qu'à "donne", rendant ainsi compte des phénomènes de "filled gap" mis à jour en psycholinguistique [Crain & Fodor 85][Stowe 86].

- (c) Jean donne une pomme à Marie
- (d1) Il soupçonne l'organisateur de cette manifestation
- (d2) Il soupçonne Jean de cette manifestation
- (e) A qui Marie dit elle que Jean donne des fleurs ?

4.2. Pourquoi LTAG ?

Le formalisme LTAG permet donc de modéliser de manière compacte et élégante sur un même objet (i.e. arbre de dérivation) les préférences *argument vs modifieur* et *idiome vs interprétation littérale*, ce qui n'était pas possible avec les approches structurales traditionnelles. Il permet également de représenter des préférences structurales en tenant compte de la distinction *argument/modifieur* et en n'étant pas tributaire de choix d'implémentation de la grammaire (i.e. forme des arbres élémentaires ...), échappant ainsi aux principaux écueils des approches structurales traditionnelles formulées en termes d'arbre de constituant : notre principe 2 comporte des similitudes avec le principe d'association droite, mais encode des informations en termes de dépendants et non de constituants et ne concerne

⁷ Concernant une interaction possible entre le Principe 1 général et sa version "spécialisée", précisons que nous n'avons jamais rencontré associé à une phrase deux arbres de dérivations A et B où A comporterait plus de noeud que B, mais cependant moins de noeud β . Il s'est même avéré impossible de construire un tel exemple, si ce n'est en recourant à des langages formels n'ayant aucun lien avec le langage naturel.

que les arguments. Si l'on tente de formuler ces préférences dans le cadre du formalisme GB, on retombe sur les mêmes problèmes que pour les approches traditionnelles (i.e. analyse en

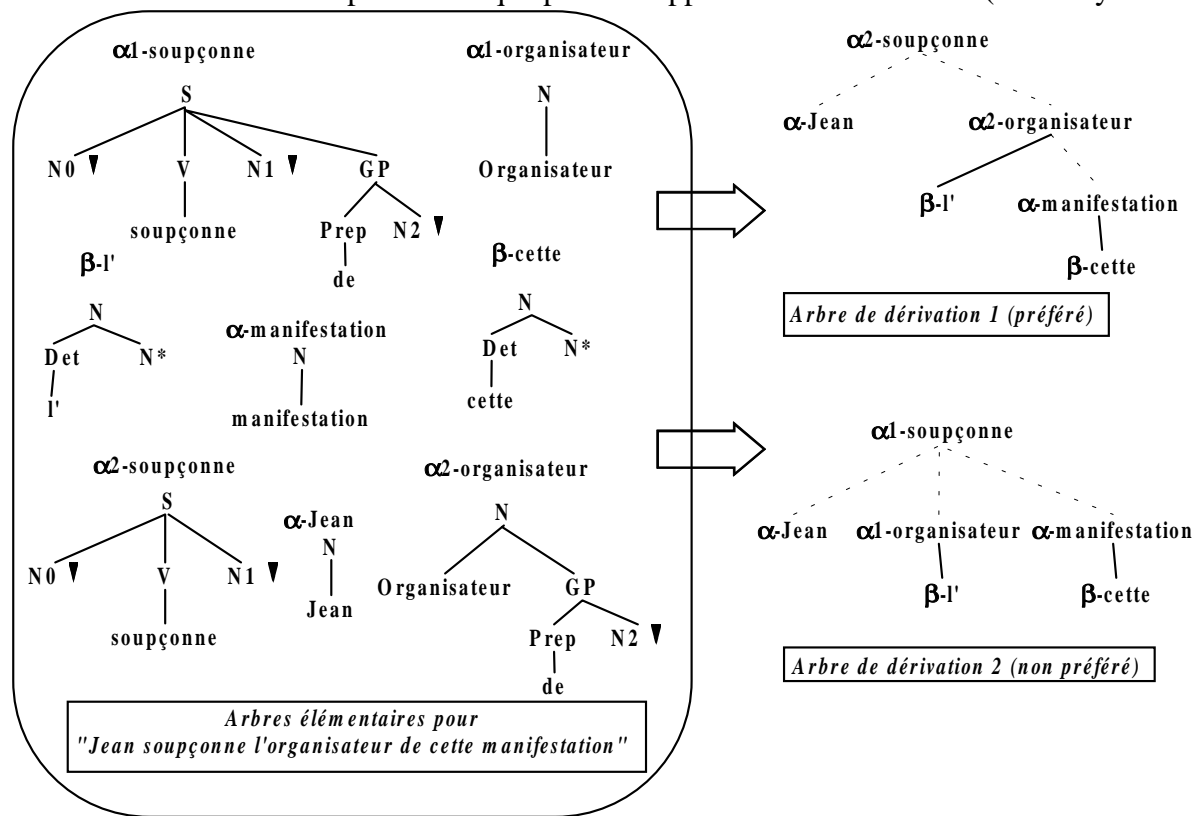


FIGURE 3 : Illustration du Principe 2

constituants). Si l'on tente de faire de même dans le cadre de LFG, il s'avère ardu de formuler ces préférences sur un même objet : les informations de figement se trouvent au niveau des entrées lexicales, la distinction argument/modifieur dans la F-structure, et les informations structurales dans la C-structure. Cela entraînerait donc un "codage" assez peu élégant..

4.3 Pourquoi nous ne traitons pas l'attachement des modifieurs

Les trois principes formulés ci-dessus ne traitent pas explicitement du problème de l'attachement des modifieurs mais tentent plutôt d'attacher les arguments de manière aussi exacte que possible. Deux raisons ont motivé ce choix :

Tout d'abord, nous avons vu que l'existence de principes universels d'attachement des modifieurs était peu probable. Ensuite, nous voulions évaluer dans quelles mesures l'attachement correct des arguments influe sur le taux d'ambiguïté, tous les autres paramètres demeurant inchangés (i.e. ambiguïté artificielle et/ou d'attachement des modifieurs).

5. Résultats préliminaires

Nous avons utilisé 1074 phrases grammaticales (i.e. "1" dans la terminologie TSNLP) issues de TSNLP et correspondent aux phrases de catégorie S (ou augmenté à S) hors coordination, acceptées par l'analyseur Xtag lors de l'évaluation d'une grammaire du français à large couverture et indépendante du domaine. Pour chaque phrase analysée, un humain a choisi un ou plusieurs arbres de dérivation jugés "corrects"⁸. Le principe 1, puis les principes 1 & 2 furent alors appliqués sur les arbres de dérivation afin d'en éliminer certains. En outre,

⁸ Plus d'un arbre de dérivation pouvait être jugé correct lorsqu'une ambiguïté non artificielle d'attachement des modifieurs demeurait.

nous avons ensuite appliqué 3 heuristiques dépendants de la langue et de la grammaire pour améliorer les résultats en diminuant les ambiguïtés artificielles :

- 1- Préférer l'attachement d'un adverbe à un verbe plein plutôt qu'à un auxiliaire
ex : *Paul a souvent mangé des pommes*
- 2- Préférer les catégories grammaticales aux catégories lexicales, notamment :
Préférer les clitiques aux noms : *Elle court* → "Elle" peut être cl ou n
Préférer les auxiliaires aux verbes lexicaux :
ex : *Elle est venue* → aux + pp / vb + n
Préférer les déterminants aux adjectifs numériques :
ex : *Je vois deux hommes* → "deux" num ou det
- 3- Ne pas préférer l'impersonnel

La table 1 illustre les résultats obtenus.

	Avant application des principes	Après application du Principe 1	Après application des Principes 1 & 2	Après application des Principes 1 & 2 + heuristiques
# total de Phrases	1074	1074	1074	1074
# total de dérivations	3057	2474	2334	1616
# de phrases ayant au moins 1 analyse correcte	1070 (99,6%)	1055 (98,2 %)	1054 (98,1 %)	1054 (98,1 %)
# de phrases ambiguës	537	427	424	214
# de phrases non ambiguës	537	647	650	860
# de phrases totalement désambiguës	n.a.	110	113	323
# de dérivations/ph rase (pour toutes les phrases)	2,85	2,3	2,17	1,5

TABLE 1 : Application des Principes 1 & 2 + Heuristiques

5.1 Quelques commentaires sur les résultats

Nous n'avons pas pu tester nos principes sur de plus larges données du fait du manque de ressources (i.e. banques d'arbres annotées) pour le français. Nous fumes quand même surpris par les résultats obtenus après application des principes 1 & 2 : la proportion de phrases comportant au moins une analyse jugée correcte par un humain n'a baissée que marginalement. Ceci semble indiquer que nos principes 1 & 2 sont pertinents. Dans le même temps, le nombre moyen d'analyses par phrase est passé de 2.85 à 2.17. Bien qu'il ne s'agisse pas d'une chute spectaculaire, cela représente quand même une baisse de 24 %. Il est crucial de garder à l'esprit que nous ne traitons pas les problèmes d'attachement strict des modifieurs . Aussi, une phrase comme "Jean est venu hier" reste associée à 5 arbres de dérivation distincts, du fait d'ambiguïtés artificielles et réelles pour l'attachement des adverbes (i.e. "hier" attaché

à S ou à V). Ceci implique que bon nombre de phrases ne seront pas désambiguïsées par les principes 1 & 2. C'est par exemple le cas des phrases ancrant un verbe principal intransitif. Les heuristiques appliqués par la suite visent à éliminer uniquement les ambiguïtés artificielles, et ne traitent pas non plus l'attachement "strict" des modifieurs, ce qui permet d'atteindre 1,5 analyses par phrase (- 44%) sans affecter du tout la proportion de phrases recevant une analyse "correcte". Si nous regardons de plus près les phrases effectivement affectées par au moins un de nos 2 principes, le nombre d'analyse / phrase pour ces dernières passe de 6.76 à 2.94 après application des principes 1 & 2 (i.e. -56.5 %). Ces résultats sont présentés dans la table 2.

	Avant application des Principes	Après application du Principe 1	Après application du Principe 2
# de Phrases	189	189	189
# de dérivations	1279	696	556
# d'analyses / Phrase	6.77	3.68	2.94

TABLE 2 : résultat pour les Phrases sur lesquelles les principes 1 & 2 s'appliquent

5.2 Différences entre théorie et pratique

Contrairement à nos attentes, le principe 1 ne s'est avéré utile qu'une seule fois pour préférer l'interprétation idiomatique d'une phrase. Par contre, il s'est avéré très efficace pour préférer les arguments aux modifieurs : les arbres de dérivation avec des arguments tendent à avoir moins de noeuds du fait des co-têtes. Par exemple, ce principe a permis de préférer systématiquement l'analyse "passifs avec agent" pour les phrases comportant un GP en "par". Les quelques cas où le principe 1 a conservé les mauvais arbres de dérivation sont essentiellement le fait d'un phénomène : les impersonnels sont préférés dans une phrase comme "Il est venu une nuit", bien que ce phénomène syntaxique soit plus rare en Français qu'en Italien par exemple. Ceci montre l'utilité d'ajouter des heuristiques dans un but pratique. Le Principe 2 a favorisé comme prévu l'attachement bas d'argument pour des phrases telles que (g). Mais, il s'est avéré utile uniquement en conjonction avec le Principe 1 et jamais utilisé seul : il a permis de désambiguer plus avant des phrases déjà partiellement désambiguïsées par le Principe 1 en préférant un ou plusieurs arbres de dérivation parmi ceux comportant un même nombre peu élevé de noeuds. Lorsque les préférences exprimées par le Principe 2 se sont avérées incorrectes, ce fut essentiellement du à des facteurs sémantiques ou pragmatiques dans des phrases telles que (h).

(g)- L'ingénieur obtient l'accord de l'entreprise

(h)- Le prisonnier remercie l'organisation pénitentiaire de sa libération

6. Conclusion

Nous avons présenté 3 principes de désambiguïsation exprimés en termes d'arbres de dérivation dans le cadre du formalisme LTAG. Ces principes, qui visent à attacher les arguments de façon aussi "correcte" que possible, sont indépendants du domaine, de la langue et de toute application particulière. Cependant comme ils sont simples à implémenter, il est aisé de les intégrer dans une application de hiérarchisation d'analyses ou encore de les intégrer directement dans un analyseur.

Les résultats préliminaires obtenus sont encourageant quant à la pertinence de ces deux principes. A court terme ces principes seront testés sur des données plus large (e.g. Le Monde) et raffinés dans un but pratique (i.e. ajout d'information de fréquence de sous-catégorisation, heuristiques pour l'attachement de modifieurs). Egalement, ces principes sont à l'heure actuelle testés sur des langues autres que le français (WSJ pour l'anglais), ce qui permettra de comparer nos résultats avec ceux de [Srinivas & al. 95]. Comme c'est la première fois à notre

connaissance que des principes d'attachement sont exprimé en termes de "dépendances" , il serait intéressant également de déterminer dans quelles mesures cette approche peut être adaptée aux grammaires dépendanciellees..

Références

- Abeillé A. (1991) : Une grammaire lexicalisée d'arbres adjoints pour le français : application à l'analyse automatique. Thèse de doctorat. Université Paris 7.
- Abeillé A. (1995) : The flexibility of French idioms. In *Idioms* LEA. Schenk & al. (eds).
- Abeillé A., Candito M.H. (1999) : FTAG : A lexicalized Tree Adjoining Grammar for French. In *Tree Adjoining Grammars*. Abeillé, Ranbow(eds). CSLI, Stanford.
- Abney S. (1989) : A computational model of human parsing. *Journal of psycholinguistic Research*, 18,
- Britt M, Perfetti C., Garrod S, Rayner K. (1992) : Parsing and discourse : Context effects and their limits. *Journal of memory and language*, 31, 293-314.
- Brysbaert M., Mitchell D.C. (1996) : Modifier Attachment in sentence parsing : Evidence from Dutch. *Quarterly journal of experimental psychology*, 49a, pp.664-695.
- Candito M.H. (1996): A principle based hierarchical representation of LTAG. COLING'96. Copenhagen.
- Crain S., Fodor J.D. (1985) : How can grammars help parsers? In *Natural language parsing : Psychological, computational and theoretical perspectives* pp. 94-127. D. Dowty, L. Karttunen, A. Zwicky (eds). Cambridge. Cambridge University Press.
- Crain S. Steedman M. (1985) : On not being led up the garden path : the use of context by the psychological parser. In *Natural Language Parsing*. Dowty, Karttunen, Zwicky (eds). Cambridge. Cambridge University Press.
- Cuetos F., Mitchell D.C. (1988) ; Cross linguistic differences in parsing : restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30, pp.73-105.
- Doran C., Egedi D., Hockey B.A., Srinivas B., Zaidel M. (1994) : Xtag System- a wide coverage grammar for English. COLING'94. Kyoto. Japan.
- Estival D., Lehman S (1997) : "TSNLP: des jeux de phrases test pour le TALN", TAL 8:1
- Ferreira F. Clifton C. (1986) : The independence of syntactic processing. *Journal of Memory and Language*, 25, pp.348-368.
- Frank R. (1992) : Syntactic Locality and Tree Adjoining Grammar : Grammatical Acquisition and Processing Perspectives. PhD dissertation. University of Pennsylvania.
- Frazier L, Fodor J.D. (1978) : "The sausage machine" : a new two stage parsing model. *Cognition* 6.
- Gibbs R. (1985): On the process of understanding idioms. *Journal of psycholinguistic research*, 14.
- Gibbs R., Nayak (1989) : Psycholinguistic studies on the syntactic behaviour of idioms. *Cognitive Psychology*, 21, pp. 100-138.
- Hindle D. Rooth M. (1993) : Structural ambiguity and lexical relations. *Computational Linguistics*, 19
- Joshi A. (1990) : Processing crossed and serial dependencies : an automaton perspective on the psycholinguistic results. *Language and cognitive processes*, 5:1, pp. 1-27.
- Joshi A. (1985) : How much context-sensitivity is necessary for characterizing structural descriptions. In *Natural Language Parsing*. Dowty, Karttunen, Zwicky (eds). Cambridge. Cambridge University Press.
- Kimball J. (1973) : Seven principles of surface structure parsing in natural language.. *Cognition* 2.
- Kinyon A. (1999) : Parsing preferences with LTAGs: exploiting the derivation tree. *Actes ACL-student'99*.
- Schubert L. : (1984). *On parsing preferences. Proceedings 10th International Conference on Computational Linguistics* (Coling 84), Stanford. pp. 247-250.
- Mel'cuk I. (1988) : *Dependency Syntax : Theory and Practice*. Albany. SUNY press.
- Schabes Y. (1990). *Computational and Mathematical Properties of Lexicalized Grammars*. PhD Dissertation. Univ. of Pennsylvania, Philadelphia..
- Srinivas B., Doran C., Kulick S. (1995) : Heuristics and Parse Ranking. IWPT'95. Prag. Czech Republic.
- Stowe L.A. (1986) : Evidence for on-line gap location. *Language and Cognitive Processes* 1:227-245.
- Vijay-Shanker K : (1987). *A study of Tree Adjoining Grammars*. PhD dissertation. Univ. of Pennsylvania..
- Xtag group (1995) : A Lexicalized Tree Adjoining Grammar for English. Technical Report IRCS 95-03. University of Pennsylvania.