

## **Représentation sémantique sous-spécifiée pour les conjonctions de subordination**

Laurence Danlos  
TALANA/ LATTICE  
Université Paris 7  
laurence.danlos@linguist.jussieu.fr

### **Mots-clefs – Keywords**

Discours, SDRT (Segmented Discourse representation Theory), DAG (Directed Acyclic Graph), Représentation sémantique sous-spécifiée, LTAG (Lexicalized Tree Adjoining Grammar).

Discourses, SDRT, DAG, Underspecified semantic representation, LTAG.

### **Résumé - Abstract**

Cet article concerne les phrases complexes avec deux conjonctions de subordination. Nous montrerons que de telles phrases peuvent s'interpréter de quatre façons différentes. Il s'agit donc de formes fortement ambiguës pour lesquelles il est opportun d'avoir recours à des représentations sémantiques sous-spécifiées, et c'est ce que nous proposerons.

This paper concerns sentences with two subordination conjunctions. It shows that such sentences can be given four interpretations. They are heavily ambiguous forms for which an underspecified semantic representation is proposed.

## 1 Introduction

Cet article concerne les phrases complexes avec deux conjonctions de subordination. Nous montrerons d’abord (Section 2) que de telles phrases peuvent s’interpréter de quatre façons différentes. Pour deux interprétations, une des conjonctions a portée large sur l’autre, pour les deux autres interprétations, aucune conjonction n’a de portée large<sup>1</sup>. L’existence de ces quatre interprétations conduit au résultat suivant : les structures de dépendance sémantique pour ces phrases complexes sont des graphes qui ne sont pas toujours arborescents (ce qui va contre l’idée couramment répandue que ces structures sont toujours des arbres). Ces graphes sont des graphes dirigés acycliques (“directed acyclic graphs” ou DAG) qui sont de plus étiquetés et ordonnés (“labeled ordered DAG” ou LODAG).

Ensuite, nous montrerons (Section 3) que les phrases qui nous concernent reçoivent deux analyses syntaxiques, l’une obtenue par récursivité, l’autre par itération d’adjoints. Ces analyses seront décrites dans le cadre de LTAG (Lexicalized Tree Adjoining Grammar) (Joshi & Schabes, 1997). L’étude des variantes par l’ordre linéaire - proposition(s) subordonnée(s) postposée(s) ou préposée(s) - permet d’établir que chaque analyse syntaxique peut être interprétée de deux façons (Section 4).

Il s’agit donc de formes fortement ambiguës pour lesquelles il est opportun d’avoir recours à des représentations sémantiques sous-spécifiées pour de nombreuses applications, entre autres la Traduction Automatique. En plagiant (Copestake *et al.*, 1999), nous pouvons en effet affirmer qu’il est fort difficile de désambiguer ces formes alors qu’il est souvent inutile de le faire : leur désambiguation est sans effet sur leur traduction. Nous proposerons donc (Section 5) une représentation sémantique sous-spécifiée (abrégée en RSSS) pour les conjonctions de subordination. Cette représentation s’inspire de formalismes basés sur les contraintes de dominance (Reyle, 1993), (Bos, 1995), (Duchier & Gardent, 2001), qui seront introduits brièvement. Toutefois, elle s’en différencie sur deux plans : d’une part, “la dominance à gauche” remplace la dominance, d’autre part les contraintes de dominance sont résolues dans des graphes qui n’ont pas forcément une structure arborescente. Avec notre proposition, une analyse syntaxique donnée ne reçoit qu’une seule analyse sémantique (sous-spécifiée).

## 2 Représentations de dépendance sémantique

Nous posons que les phrases avec une conjonction de subordination sont du type  $S1 \text{ Conj}_a S2$  - en employant le symbole  $S$  (Sentence) pour une proposition (phrase) - lorsque la subordonnée est postposée à la principale<sup>2</sup>. Nous adoptons une représentation de dépendance sémantique standard, à savoir,  $\text{Conj}_a$  dénote une relation de discours  $R_a$ .  $R_a$  est un prédicat avec deux arguments  $\pi_1$  et  $\pi_2$ , qui correspondent aux représentations sémantiques de  $S1$  and  $S2$  respectivement. Ces arguments sont ordonnés ( $\pi_1$  précède  $\pi_2$ ) et étiquetés avec les étiquettes nucleus (notée N) et satellite (notée S) qui viennent de la RST (Rhetorical Structure Theory) (Mann & Thompson, 1987).  $\pi_1$  est donc le nucleus de  $R_a$ ,  $\pi_2$  son satellite. Cette représentation sémantique correspond au LODAG de la partie gauche de la Figure 1. Sous forme abrégée, ce LODAG s’écrit  $R_a(\pi_1, \pi_2)$ . Il correspond dans le cadre de la SDRT (Segmented Discourse Representation Theory) (Asher, 1993) à la SDRS figurant dans la partie droite de la Figure 1 ( $e_i$  dénote l’éventualité principale de  $S_i$ ).

<sup>1</sup>En comparaison, une proposition avec deux quantifiers ne peut s’interpréter que de deux façons, car un des quantifiers a forcément portée large sur l’autre.

<sup>2</sup>Dans cette section concernant la sémantique, nous supposons que les propositions subordonnées sont toujours postposées. Dans la Section 3 concernant la syntaxe, la position postposée ou préposée des subordonnées sera discutée en détail.

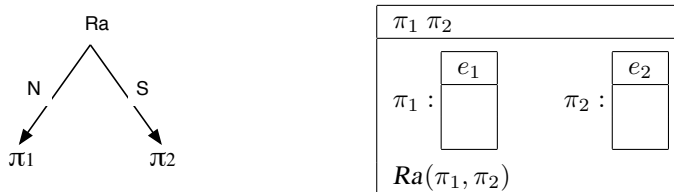


Figure 1: LODAG et SDRS pour  $SI\ Conj_a\ S2$

Nous allons montrer que les phrases avec deux conjonctions de subordination, i.e. les phrases du type  $SI\ Conj_a\ S2\ Conj_b\ S3$ , peuvent être interprétées de quatre façons différentes et proposer des LODAG pour chacune de ces interprétations. Pour cela, nous nous appuyerons sur le principe de compositionnalité suivant : si, à partir d’un discours  $D_n$  avec  $n$  phrases, on peut inférer un sous-discours  $D_p$  avec  $p$  phrases et  $1 < p < n$ , alors le LODAG pour  $D_n$  doit inclure le LODAG pour  $D_p$ .

Considérons la phrase complexe (1a). Elle est ambiguë (à l’écrit), comme le montre le fait que la question en (1b) peut recevoir la réponse (1c) ou (1d). Ces interprétations se distinguent à l’oral par une accentuation ou non accentuation sur *avant*. A l’écrit, en l’absence de toute ponctuation (la ponctuation sera examinée dans la Section 3), l’ambiguïté ne peut être levée que si les habitudes de Fred sont connues, connaissance qui peut être difficile ou impossible à acquérir par un système de traitement automatique. D’où le besoin d’une RSSS qui ne choisit pas entre l’une ou l’autre interprétation. Cette RSSS sera présentée dans la Section 5. Pour le moment, intéressons-nous aux représentations sémantiques spécifiées où l’ambiguïté est levée.

- (1) a. Marie a été surprise parce que Fred a fait la vaisselle avant d’aller à l’école.
- b. Pourquoi Marie a-t-elle été surprise ?
- c. Parce que Fred a fait la vaisselle AVANT d’aller à l’école (d’habitude il la fait en rentrant de l’école)<sup>3</sup>.
- d. Parce que Fred a fait la vaisselle (d’habitude il ne la fait pas).

Dans l’interprétation correspondant à la réponse (1c), nous disons que  $Conj_a = parce\ que$  a portée large sur  $Conj_b = avant\ (que)$  et nous posons que le graphe de dépendance sémantique est topologiquement le LODAG (A) dans la Figure 2. Dans l’interprétation correspondant à la réponse (1d), aucune conjonction n’a de portée large, nous disons que  $S2$  est factorisée et nous posons que le graphe est le LODAG (B) dans la Figure 2. Vérifions que (A) et (B) sont conformes à notre principe de compositionnalité. Dans les deux interprétations, si (1a) est vrai, alors le sous-discours  $S2\ Conj_b\ S3 = Fred\ a\ fait\ la\ vaisselle\ avant\ d’aller\ à\ l’école$  est vrai. Les LOGAD pour (1a) doivent donc inclure le LODAG  $R_b(\pi_2, \pi_3)$  représentant  $S2\ Conj_b\ S3$ , ce qui est le cas dans (A) et (B). En revanche, si (1a) est vrai, le sous-discours  $SI\ Conj_a\ S2 = Marie\ a\ été\ surprise\ parce\ que\ Fred\ a\ fait\ la\ vaisselle$  n’est vrai que dans l’interprétation où  $S2$  est factorisée. (B) contient bien le LODAG  $R_a(\pi_1, \pi_2)$  représentant  $SI\ Conj_a\ S2$ , ce qui n’est pas le cas pour (A).

Le LODAG (B) n’a pas une structure arborescente :  $\pi_2$  a deux pères, c’est le satellite de  $R_a$  et le nucleus de  $R_b$ , d’où l’expression “ $S2$  est factorisée”. Par contre, (A) a une structure arborescente, et cet arbre doit être interprété de façon standard en mathématiques ou informatique, i.e. le satellite de  $R_a$  est son fils droit, i.e. le sous-arbre de racine  $R_b$ . On peut défendre la position qu’un arbre comme (A) ne devrait pas être interprété de façon standard. Cette position est défendue dans le cadre de la RST où les arbres sont interprétés par le “principe de nucléarité”

<sup>3</sup>Le mot *avant* est écrit en majuscules pour indiquer qu’il est accentué à l’oral.

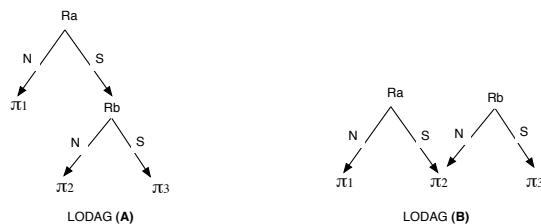


Figure 2: Représentations sémantiques spécifiées pour (1a)

(Marcu, 1996). Avec ce principe, les arguments d’une relation de discours ne peuvent être que les *feuilles* de l’arbre ; par exemple, le satellite de  $R_a$  dans (A) est la feuille  $\pi_2$  (le nucleus de  $R_b$ ). Ceci revient à interpréter (A) comme (B). Mais alors les cas à portée large ne sont plus représentés, ils ne sont pas pris en compte, ce qui est inacceptable. Par conséquent, les graphes de structure arborescente doivent être interprétés de façon standard, ce qui permet que les arguments d’une relation de discours soient des nœuds intermédiaires ou des feuilles.

Il est souvent considéré que les représentations de dépendance sémantique pour les discours ne peuvent être que des arbres (c’est un principe théorique en RST ). De ce fait, certains auteurs ont recours à des arbres qui sont artificiels dans la mesure où certaines relations prédicat-argument sont données par le principe de nucléarité et d’autres par l’interprétation standard, ce que nous critiquons dans (Danlos, 2003). Il faut admettre que les représentations de dépendance sémantique pour les discours sont des LODAG : cette affirmation est justifiée dans cet article portant sur les phrases complexes avec deux propositions subordonnées. Elle est aussi justifiée dans (Danlos, 2003) pour d’autres types de discours contenant trois propositions, et *a fortiori* pour des discours plus complexes contenant  $n$  propositions avec  $n > 3$ .

Concluons cette partie concernant les deux interprétations de (1a) en donnant dans la Figure 3 les SDRS correspondant aux LODAG (A) et (B).

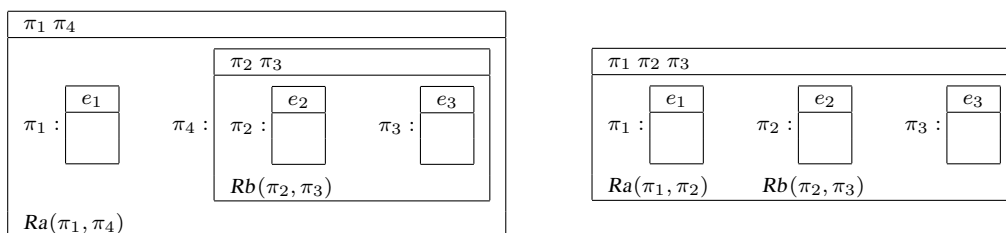


Figure 3: SDRS correspondant à (A) et (B)

Tournons-nous maintenant vers la phrase complexe (2a). Elle est aussi ambiguë (à l’écrit), ce qui est mis en évidence en complétant la dernière proposition par une infinitive, voir (2b) et (2c). Cette ambiguïté suscite les mêmes commentaires que ceux que nous avons faits sur (1a) concernant l’oral, la ponctuation, le besoin d’une RSSS.

- (2) a. Fred fait la vaisselle avant d’aller à l’école si Marie lui demande.
- b. si Marie lui demande de faire la vaisselle AVANT d’aller à l’école (sinon il la fait en rentrant de l’école).
- c. si Marie lui demande de faire la vaisselle (sinon il ne la fait pas).

Dans l’interprétation explicitée en (2b), nous disons que  $Conj_b = si$  a portée large, dans celle explicitée en (2c), nous disons que  $SI$  est factorisée. Ces interprétations sont représentées dans les LODAG (C) et (D) de la Figure 4, qui sont conformes à notre principe de compositionnalité. En effet, dans les deux interprétations, si (2a) est vrai, alors le sous-discours  $SI Conj_a S2 = Fred$

*fait la vaisselle avant d'aller à l'école* est vrai dans certaines occasions, ce qui va de pair avec le fait que (C) et (D) incluent le LODAG  $R_a(\pi_1, \pi_2)$ . En revanche, si (2a) est vrai, le sous-discours  $S1\ Conj_b\ S3 = Fred\ fait\ la\ vaisselle\ si\ Marie\ lui\ demande$  n'est vrai que dans l'interprétation où  $S1$  est factorisée : (D) contient bien le LODAG  $R_b(\pi_1, \pi_3)$ , ce qui n'est pas le cas pour (C).

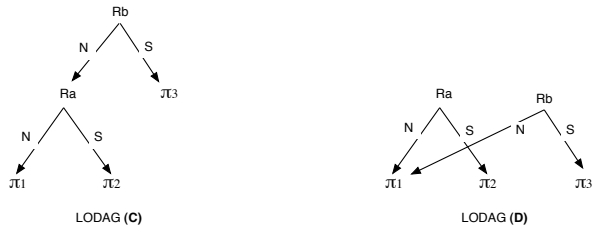
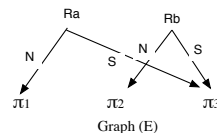


Figure 4: Représentations sémantiques spécifiées pour (2a)

**Résumé sur la sémantique de  $S1\ Conj_a\ S2\ Conj_b\ S3$ .** Une phrase avec deux subordonnées reçoit une des quatre interprétations représentés dans les LODAG (A), (B), (C) et (D), qui n'ont pas forcément une structure arborescente. Elle en reçoit plusieurs dans les cas ambigus (voir la section suivante pour les ambiguïtés syntaxiques). D'où le besoin d'une RSSS pour les conjonctions de subordination (Section 5).

Ce résumé ne prend en compte ni les valeurs lexicales des conjonctions en jeu ni les relations de discours qu'elles dénotent. Nous avons vu dans les exemples (1a) et (2a) que les conjonctions *parce que* et *si* avaient ou non une portée large, tandis que *avant (que)* n'avait de portée large ni en position  $Conj_a$  ni en position  $Conj_b$ . Ces exemples ne nous permettent évidemment pas d'affirmer que *avant (que)* n'a jamais de portée large : seule une étude de corpus permettrait de valider ou d'invalider cette hypothèse. Néanmoins, il se pourrait que certaines conjonctions ou certaines relations de discours n'aient jamais de portée large.

Il semble que les interprétations des phrases  $S1\ Conj_a\ S2\ Conj_b\ S3$  se limitent aux quatre interprétations que nous venons de décrire. En particulier, nous n'avons pas trouvé d'exemple dans lequel  $S3$  serait factorisée. Cette interprétation correspond au LODAG (E) ci-contre. Dans (E), il n'y a pas de relation de dominance à gauche (Section 5) entre  $R_a$  et  $\pi_2$ , alors qu'il y en a une dans les LODAG (A)-(D). Ceci montre que les LODAG pour les discours présentent des contraintes structurelles (Danlos, 2003).



### 3 Analyse syntaxique et ordre linéaire (en LTAG )

Syntaxiquement, les propositions subordonnées sont des modificateurs de phrases<sup>4</sup>. De ce fait, dans XTAG (XTAG Research Group, 1998) et FTAG (Abeillé *et al.*, 2000), les grammaires LTAG de l'anglais et du français, une conjonction de subordination ancre des arbres auxiliaires de racine  $S$ , avec deux arguments syntaxiques de catégorie  $S$ , le nœud pied (la principale) et le nœud à substitution (la subordonnée). Une proposition subordonnée peut être *postposée* (l'ordre linéaire est alors  $S1\ Conj_a\ S2$  et l'arbre auxiliaire est noté  $\beta_1(Conj_a)$ ) ou *préposée* (l'ordre linéaire est alors  $Conj_a\ S2, S1$  et l'arbre auxiliaire est noté  $\beta_2(Conj_a)$ ). En ce qui concerne la ponctuation, une virgule est obligatoire pour séparer une subordonnée préposée de la principale, alors qu'elle est facultative dans le cas d'une subordonnée postposée. La Figure 5 montre les arbres auxiliaires et les arbres de dérivation pour les propositions subordonnées<sup>5</sup>.

<sup>4</sup>Elles peuvent aussi être des modificateurs de GV, mais nous ne traiterons pas ce cas.

<sup>5</sup>Dans un arbre de dérivation, un arc en pointillés indique l'adjonction, un arc plein la substitution ; chaque arc est étiqueté par l'adresse de Gorn de l'argument où a lieu l'opération ;  $\alpha_i$  est l'arbre LTAG représentant  $S_i$ .

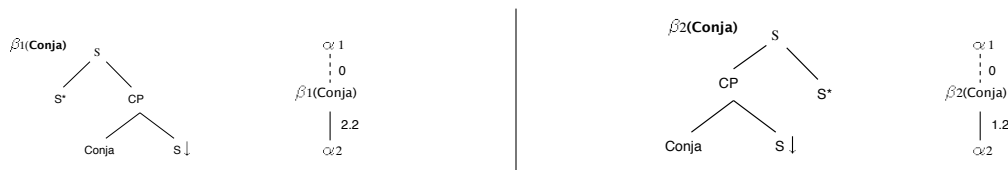


Figure 5: Arbres auxiliaires et arbres de dérivation pour les subordinées postposées et préposées

Une phrase  $S1\ Conj_a\ S2\ Conj_b\ S3$  reçoit deux analyses syntaxiques. Dans la première, obtenue par récursivité et notée AS1,  $Conj_a\ S2$  s'adjoint à  $S1$  et  $Conj_b\ S3$  s'adjoint à  $S2$ . Dans la seconde, obtenue par itération des modifieurs et notée AS2,  $Conj_a\ S2$  et  $Conj_b\ S3$  s'adjoignent à  $S1$ . La Figure 6 montre les arbres dérivés et les arbres de dérivation pour ces deux analyses. Pour AS2, l'arbre de dérivation repose sur une extension de LTAG permettant des adjonctions multiples à un même nœud, extension proposée dans (Schabes & Shieber, 1994)<sup>6</sup>. Ces adjonctions multiples sont ordonnées (de gauche à droite). Les phrases  $S1\ Conj_a\ S2\ Conj_b\ S3$  sans aucune virgule s'analysent de préférence selon AS1. À l'inverse, les phrases  $S1\ Conj_a\ S2, Conj_b\ S3$  avec une virgule avant la seconde conjonction et les phrases  $S1, Conj_a\ S2, Conj_b\ S3$  où la seconde subordinée est encadrée de virgules s'analysent de préférence selon AS2.

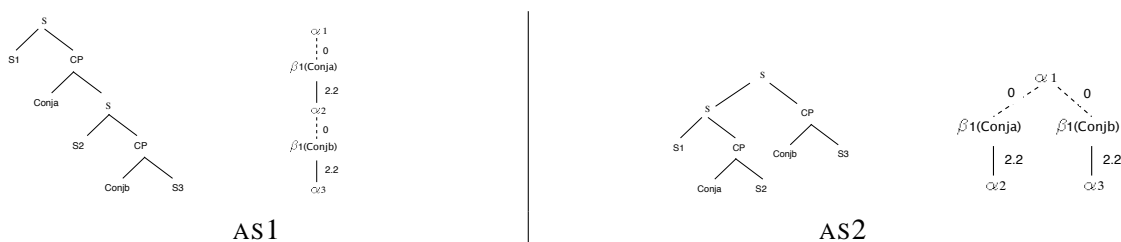


Figure 6: Arbres dérivés et arbres de dérivation pour AS1 et AS2

Examinons les variantes de l'ordre linéaire neutre  $S1\ Conj_a\ S2\ Conj_b\ S3$ , qui correspond au cas où  $Conj_a$  et  $Conj_b$  ancrent un arbre postposé. Pour chaque analyse, on doit considérer les trois cas suivants : (a)  $Conj_a$  ancre l'arbre préposé  $\beta2(Conj_a)$  et  $Conj_b$  l'arbre postposé  $\beta1(Conj_b)$ , (b) symétriquement,  $Conj_a$  ancre  $\beta1(Conj_a)$  et  $Conj_b$   $\beta2(Conj_b)$ , (c)  $Conj_a$  et  $Conj_b$  ancrent un arbre préposé. La Figure 7 montre les ordres linéaires (non neutres) pour AS1 et AS2.

Quelles sont les ambiguïtés syntaxiques de ces variantes ? Les variantes (a2) et (b2) sont toutes deux de la forme  $Conj\ S, S\ Conj\ S$ . De ce fait, chaque phrase de cette forme reçoit systématiquement les deux analyses (a2) et (b2). Nous verrons toutefois dans la Section 5 que ces deux analyses syntaxiques correspondent à la même représentation sémantique sous-spécifiée. Les variantes (a1) et (c2) sont de la forme  $Conj\ S, (,) Conj\ S, S$ , la virgule avant la seconde conjonction étant obligatoire pour (c2) et pratiquement interdite pour (a1). Ces formes ne sont donc quasiment pas ambiguës. Les variantes (b1) et (c1) ne sont pas syntaxiquement ambiguës.

## 4 Interface syntaxe-sémantique

Il nous reste à étudier les interprétations sémantiques des analyses AS1 et AS2. Pour cela, nous allons utiliser comme critère les variantes de l'ordre linéaire neutre, et donc examiner les variantes des discours (1a) et (2a) illustrant les quatre interprétations (A)-(D). Quelle que soit l'interprétation de (1a), i.e. (A) ou (B), les variantes de ce discours correspondant aux variantes

<sup>6</sup>Si on n'intègre pas cette extension de LTAG, on ne peut absolument pas rendre compte de l'interprétation (D).

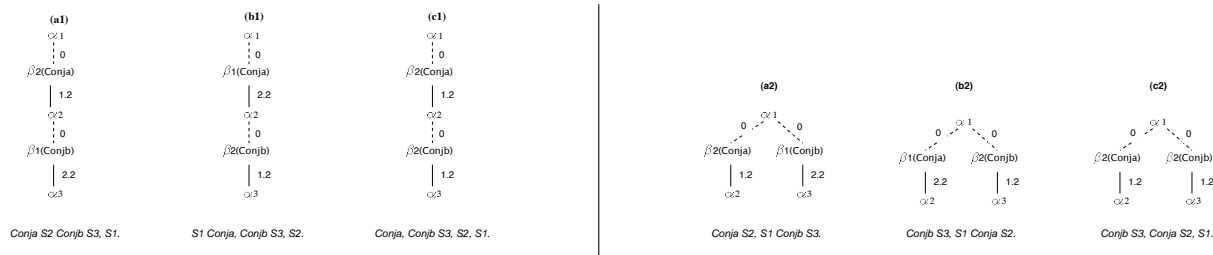


Figure 7: Ordres linéaires non neutres pour AS1 et AS2

autorisées avec l’analyse AS1 (voir Figure 7) donnent des résultats acceptables, plus ou moins heureux, voir (3). Par contre, les variantes correspondant à AS2 sont interdites : les discours de (4) ne font pas sens ou ne décrivent pas la situation dénotée par (1a) (d’où le signe #).

- (3) a. Parce que Fred a fait la vaisselle avant d’aller à l’école, Marie a été surprise.
- b. Marie a été surprise parce que, avant d’aller à l’école, Fred a fait la vaisselle.
- c. Parce que, avant d’aller à l’école, Fred a fait la vaisselle, Marie a été surprise.
- (4) a. # Parce que Fred a fait la vaisselle, Marie a été surprise avant qu’il aille à l’école.
- b. # Avant que Fred aille à l’école, Marie a été surprise parce qu’il a fait la vaisselle.
- c. # Avant que Fred aille à l’école, parce qu’il a fait la vaisselle, Marie a été surprise.

Les interprétations (A) et (B) correspondent donc à l’analyse AS1, ce qui revient à dire que AS1 peut être interprétée selon (A) ou (B). Le lecteur vérifiera sur les variantes de (2a) qu’à l’inverse AS2 peut être interprétée selon (C) ou (D).

**Résumé :** Une phrase  $S1\ Conj_a\ S2\ Conj_b\ S3$  ou une de ses variantes par l’ordre linéaire peut recevoir syntaxiquement une analyse par récursivité AS1 ou une analyse par itération de modifieurs AS2 (ou les deux analyses dans les cas ambigus). AS1 peut être interprétée selon (A) ou (B), AS2 selon (C) ou (D).

## 5 Représentation sémantique sous-spécifiée

**Introduction.** Comme nous l’avons indiqué dans la Section 1, il peut être opportun en TAL d’avoir recours à des représentations sémantiques sous-spécifiées pour certaines formes ambiguës. Le cas le plus connu est celui des propositions avec deux quantifieurs qui présentent des ambiguïtés de portée. Ainsi la proposition (5a) avec les deux quantifieurs *tout* et *une* s’interprète soit selon la formule logique (5b) avec portée large de *tout* (chaque marin a son propre amour féminin) soit selon (5c) avec portée large de *une* (il existe une femme qui est aimée par tous les marins).

- (5) a. Tout marin aime une femme
- b.  $\forall x \exists y \text{ marin}(x) \wedge \text{femme}(y) \wedge \text{aime}(x, y)$
- c.  $\exists y \forall x \text{ marin}(x) \wedge \text{femme}(y) \wedge \text{aime}(x, y)$

Les travaux de (Reyle, 1993), (Bos, 1995) et (Duchier & Gardent, 2001) proposent de représenter de façon sous-spécifiée une proposition ambiguë comme (5a) selon une “description d’arbre”, illustrée (de façon très simplifiée) dans la partie gauche de la Figure 8, où un trait plein représente la dominance immédiate (relation de père à fils), un trait en points la dominance (relation de père, grand-père, arrière grand père, ... à fils). Ces relations de dominance sont “résolues” (respectées) dans les arbres (a) et (b) de la partie droite de la Figure 8 où la portée des quantifieurs est spécifiée : dans (a)  $\text{pourtout}(x)$  a portée large, dans (b)  $\text{c'est existe}(y)$ .

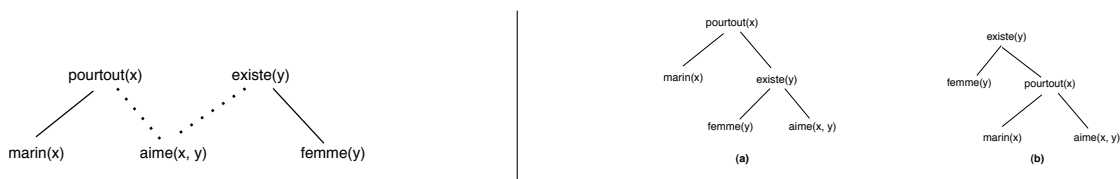


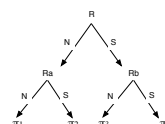
Figure 8: RSSS pour (5a), et représentations avec la portée des quantifieurs spécifiée

La RSSS que nous allons proposer pour les conjonctions de subordination s’inspire de cette idée. Cependant elle en diffère par le fait que nous utilisons “la dominance à gauche” et non la dominance.

**La dominance à gauche.** Nous avons vu dans la Section 2 que le principe de nucléarité est trop restrictif (il élimine d’emblée des interprétations acceptables). Nous verrons à la fin de cette section que la relation de dominance n’est pas assez restrictive. Nous introduisons donc une nouvelle relation, appelée “dominance à gauche”, qui est intermédiaire.

**Définition de la dominance à gauche:** un nœud X d’un arbre domine à gauche un nœud Y ssi Y est un fils de X (dominance immédiate) ou il existe un fils Z de X tel que Y appartient à la frontière gauche de Z.

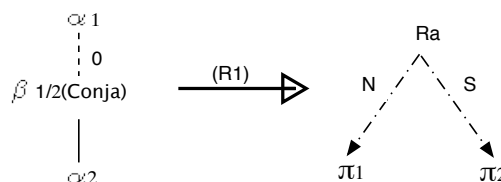
Donnons une illustration avec l’arbre binaire ci-contre. La racine R domine à gauche Ra,  $\pi_1$ , Rb et  $\pi_3$ . Avec le principe de nucléarité, R domine seulement les feuilles  $\pi_1$  and  $\pi_3$ . Avec la relation de dominance, R domine Ra,  $\pi_1$ ,  $\pi_2$ , Rb,  $\pi_3$  et  $\pi_4$ .



**Interface syntaxe-sémantique.** Dans la lignée des travaux de sémantique s’interfaçant avec LTAG (Candito & Kahane, 1998) (Kallmeyer & Joshi, 2002), nous posons comme principes que (i) chaque arbre élémentaire se voit attribuer une représentation sémantique (éventuellement sous-spécifiée), (ii) la combinaison des représentations sémantiques est régie par l’arbre de dérivation. Avec ces principes, nous proposons la règle suivante pour attribuer aux arbres élémentaires d’une conjonction de subordination une représentation sémantique sous-spécifiée.

**Règle (R1):** la représentation sémantique sous-spécifiée pour  $\beta_1(Conj_a)$  et  $\beta_2(Conj_a)$  où  $Conj_a$  dénote une relation de discours  $R_a$  est la description d’un LODAG dans lequel  $R_a$  domine à gauche  $\pi_1$  et  $\pi_2$ , les représentations sémantiques des arguments syntaxiques de  $Conj_a$ .

Cette règle est graphiquement représentée dans la Figure ci-contre, dans laquelle un trait en pointillés et points représente la dominance à gauche. La partie gauche de cette règle est l’arbre de dérivation correspondant à l’un ou l’autre des arbres élémentaires de  $Conj_a$ . Ce qui signifie que la représentation sémantique sous-spécifiée d’une conjonction de subordination (partie droite de la règle) ne prend pas en compte l’ordre linéaire de surface. Cette position suit les travaux de (Melc’uk, 2001) sur la structure communicative. En plus des structures de dépendance sémantique, Melc’uk ajoute des informations communicatives destinées à représenter, entre autres, les différences de sens induites par les variantes d’ordre linéaire.



Montrons comment la règle (R1) permet de calculer les interprétations des analyses syntaxiques AS1 et AS2 (et de leurs variantes par l’ordre linéaire) en fonction de leurs arbres de dérivation.



**Interprétations de AS1 .** A partir de l’arbre de dérivation correspondant à AS1, la règle (R1) calcule la RSSS donnée dans la partie gauche de la Figure 9 et notée RSSS1. Les contraintes de dominance à gauche et d’ordre dans RSSS1 sont résolues dans les LODAG (A) et (B) répétés dans la partie droite de la Figure 9. (B), qui n’est pas arborescent, est identique à RSSS1 excepté que la dominance immédiate remplace la dominance à gauche. Dans (A), qui est arborescent,  $R_a$  domine à gauche  $\pi_2$ , et  $\pi_1$  précède  $\pi_2$  qui précède  $\pi_3$ . Soulignons que les contraintes de RSSS1 ne sont pas résolues dans (C) car la relation entre  $R_b$  et  $\pi_2$  dans (C) est une relation de dominance et non de dominance à gauche. En conclusion, avec (R1), AS1 peut être interprétée seulement selon (A) ou (B), ce qui est conforme aux résultats de la Section 4.

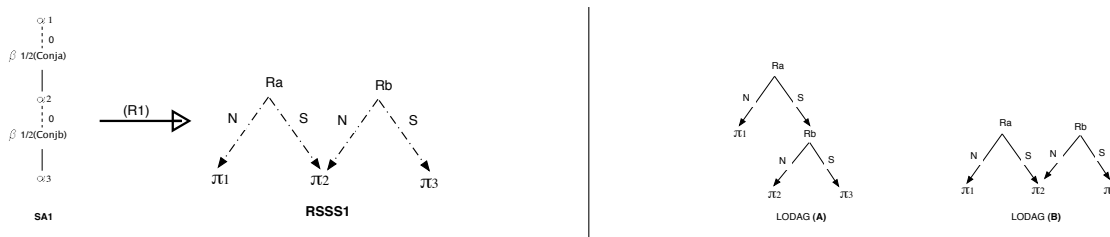


Figure 9: RSSS pour AS1 et représentations avec la portée des conjonctions spécifiée

**Interprétations de AS2 .** A partir de l’arbre de dérivation de AS2, la règle (R1) calcule la RSSS2 de la Figure 10. L’ordre des adjonctions multiples au même nœud dans AS2 est reflété par l’ordre dans RSSS2 :  $\pi_1$  précède  $\pi_2$  qui précède  $\pi_3$ . Les contraintes dans RSSS2 sont résolues dans les graphes (C) et (D). Dans (C),  $R_b$  domine à gauche  $\pi_1$ . (D) est identique à RSSS2 excepté que la dominance immédiate remplace la dominance à gauche. En conclusion, avec (R1), AS2 peut être interprétée seulement selon (C) ou (D), ce qui est conforme aux données.



Figure 10: RSSS pour AS2 et représentations avec la portée des conjonctions spécifiée

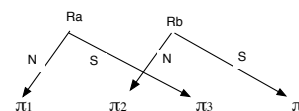
**Travaux proches.** (Duchier & Gardent, 2001) ont mis au point un système incrémental d’interprétation du discours reposant sur une RSSS pour les connecteurs (conjonctions ou adverbiaux). Notre proposition s’inspire de ce travail, avec cependant deux différences : (i) leurs représentations sémantiques spécifiées des discours sont toujours arborescentes, ce qui n’est pas le cas pour nous, (ii) ils utilisent la dominance, tandis que nous utilisons la dominance à gauche. Si la dominance était utilisée dans la règle (R1), il serait possible de résoudre les contraintes de RSSS1 (et donc d’interpréter AS1) selon (C), et de résoudre les contraintes de RSSS2 (interpréter AS2) selon (B), ce qui n’est pas conforme aux données. En revanche, le recours à la dominance à gauche évite ces mauvaises interprétations.

## 6 Perspectives et conclusion

Nous nous sommes concentrée sur les phrases avec deux propositions subordonnées. Nous envisageons deux prolongements à cette étude. D’abord les discours avec d’autres connecteurs, en particulier avec des adverbiaux comme *de ce fait* ou *ensuite*. Notre étude préliminaire sur les discours  $S1 Conn_a S2 Conn_b S2$  dans lesquels  $Conn_{a/b}$  sont des adverbiaux ou des conjonc-

tions permet d'envisager une règle similaire à (R1) pour les adverbiaux sauf que la relation de dominance à gauche entre  $R_a$  et son nucleus  $\pi_1$  serait remplacée par une relation de dominance. Chez (Duchier & Gardent, 2001), les adverbiaux sont traités comme les conjonctions.

Ensuite les discours comprenant plus de trois propositions. Leurs représentations sémantiques sont clairement des LODAG, qui présentent néanmoins des contraintes structurelles. Par exemple, la contrainte postulant que  $R_a$  doit dominer à gauche  $\pi_2$  exclut tout LODAG avec des dépendances croisées comme celles illustrées ci-contre.



## References

- ABEILLÉ A., CANDITO M.-H. & KINYON A. (2000). The current status of FTAG. In *Proceedings of TAG+5*, p. 11–18, Paris.
- ASHER N. (1993). *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer.
- BOS J. (1995). Predicate logic unplugged. In P. DEKKER & M. STOKHOF, Eds., *Proceedings of the 10th Amsterdam Colloquium*, p. 133–142.
- CANDITO M.-H. & KAHANE S. (1998). Can the TAG Derivation Tree represent a Semantic Graph? In *Proceedings of TAG+4*, p. 25–28, Philadelphia.
- COPESTAKE A., FLICKINGER D., SAG I. A. & POLLARD C. (1999). Minimal Recursion Semantics. An Introduction. Manuscript, Stanford University.
- DANLOS L. (2003). Discourse dependency structures as DAGs. In *Proceedings of MTT 2003*, Paris.
- DUCHIER D. & GARDENT C. (2001). Tree descriptions, constraints and incrementality. In R. M. H. BUNT & E. THIJSSE, Eds., *Computing Meaning*, p. 205–227. Dordrecht: Kluwer Academic Publishers.
- JOSHI A. K. & SCHABES Y. (1997). Tree-Adjoining Grammars. In G. ROZENBERG & A. SALOMAA, Eds., *Handbook of Formal Languages*, p. 69–123. Berlin: Springer.
- KALLMEYER L. & JOSHI A. (2002). Factoring Predicate Argument and Scope Semantics: Underspecified Semantics with LTAG. *Research on Language and Computation*. To appear.
- MANN W. C. & THOMPSON S. A. (1987). Rhetorical structure theory: Description and construction of texts structures. In G. KEMPEN, Ed., *Natural Language Generation*, p. 85–95. Dordrecht: Martinus Nijhoff Publisher.
- MARCU D. (1996). Building up rhetorical structure trees. In *The Proceedings of the 13th National Conference on Artificial Intelligence*, p. 1069–1074, Portland.
- MELC'UK I. (2001). *Communicative Organization in Natural Language: The Semantic-Communicative Structure of Sentences*. Amsterdam: John Benjamins Publishing Company.
- REYLE U. (1993). Dealing with ambiguities by underspecification: Construction, representation and deduction. *Journal of Semantics*, **10**, 123–179.
- SCHABES Y. & SHIEBER S. M. (1994). An Alternative Conception of Tree-Adjoining Derivation. *Computational Linguistics*, **20**(1), 91–124.
- XTAG RESEARCH GROUP (1998). *A Lexicalized Tree Adjoining Grammar for English*. Rapport interne 98–18, Institute for Research in Cognitive Science, Philadelphia.