

Structure des représentations logiques, polarisation et sous-spécification

Sylvain Kahane

Modyco, Université Paris 10
Lattice, Université Paris 7
sk@ccr.jussieu.fr

Résumé – Abstract

Cet article s'intéresse à la structure des représentations logiques des énoncés en langue naturelle. Par représentation logique, nous entendons une représentation sémantique incluant un traitement de la portée des quantificateurs. Nous montrerons qu'une telle représentation combine fondamentalement deux structures sous-jacentes, une structure « prédicative » et une structure hiérarchique logique, et que la distinction des deux permet, par exemple, un traitement élégant de la sous-spécification. Nous proposerons une grammaire polarisée pour manipuler directement la structure des représentations logiques (sans passer par un langage linéaire avec variables), ainsi qu'une grammaire pour l'interface sémantique-syntaxe.

This paper aims at the structure of logic representations in natural languages. By logic representation we mean a semantic representation including a quantifier scope processing. We show that such a representation basically combines two underlying substructures, a “predicative” structure and a logic hierarchic structure, and that the identification of the two allows for an elegant processing of underspecification. We will propose a polarized grammar that directly handles the structure of logic representations (without using a linear language with variables), as well as a grammar for the semantics-syntax interface.

Mots Clés –Keywords

Logique du premier ordre, calcul des prédicats, représentation sémantique, relation prédicat-argument, quantificateur, grammaire d'unification polarisée, grammaire de dépendance, dag, interface syntaxe-sémantique.

First order logic, predicate calculus, semantic representation, predicate-argument relation, quantifier, polarized unification grammar, dependency grammar, dag, syntax-semantics interface.

1 Introduction

Les objectifs de cet article sont multiples. D'un point de vue mathématique, il s'agit de mieux comprendre la structure des formules logiques du premier ordre (= formules du calcul des

prédicats), notamment lorsque celles-ci sont utilisées comme représentations sémantiques d'énoncés en langue naturelle. Il s'agit en particulier de mieux comprendre la nature des quantificateurs et de leur portée. D'un point de vue linguistique, il s'agit de combiner deux modes de représentations utilisés en sémantique des langues naturelles : des représentations logiques, utilisées en sémantique non lexicale et issues des travaux de Frege et d'autres logiciens, et des représentations issues de la sémantique lexicale, qui traitent tous les signifiés lexicaux comme des prédicats, y compris les quantificateurs et la négation (cf. Dymetman & Coperman 1996 pour une problématique analogue).

La section 2 introduira différentes « écritures » pour une représentation logique et mettra en évidence les deux structures sous-jacentes, tandis que la section 3 proposera une grammaire permettant de manipuler directement de telles structures, de réaliser une interface sémantique-syntaxe et de contrôler la sous-spécification.

2 Représentation logique

2.1 Variations sur la représentation logique

Cette section s'articulera autour des différents modes de représentation du sens d'un énoncé comportant plusieurs quantificateurs. Nous travaillerons avec l'énoncé suivant :

- (1) *Tout homme aime une femme.*

On lui associe généralement la représentation logique suivante :

- (2) $\forall x [\text{homme}'(x) \rightarrow \exists y [\text{femme}'(y) \wedge \text{aimer}'(x,y)]]$

En s'autorisant la confusion entre un prédicat et son extension, on peut utiliser la représentation suivante, plus proche de la langue naturelle :

- (3) $\forall x \in \text{homme}' [\exists y \in \text{femme}' \text{aimer}'(x,y)]$

Dans une telle représentation, un quantificateur est lié à trois objets : une *variable*, une *restriction* (sur cette variable) et la *portée*. Pour le quantificateur \forall en (3), il s'agit respectivement de la variable x , de la restriction $\text{homme}'(x)$ et de la sous-formule $\exists y \in \text{femme}' \text{aimer}'(x,y)$. De nombreuses approches en sémantique (Woods 1975, Copestake *et al.* 1999) rendent cette structure explicite en associant à (1) la représentation suivante :

- (4) $\text{tout}'(x, \text{homme}'(x), \text{un}'(y, \text{femme}'(y), \text{aimer}'(x,y)))$

On pourra ainsi représenter les autres quantificateurs (*certaines*, *quelques*, *la plupart de*, etc.) de la même façon que *tout* et *un*, c'est-à-dire par un prédicat à trois arguments. La formule en (4) possède une structure arborescente que nous donnons en (5) :

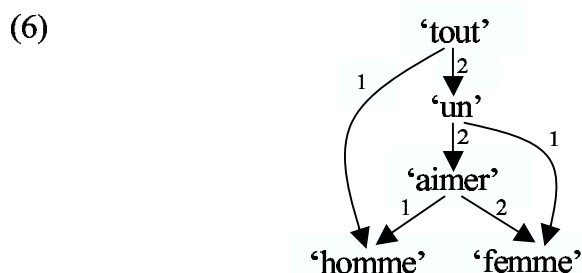
- (5)
-
- ```

graph TD
 tout["tout'"] --- x1["x"]
 tout --- homme["homme'"]
 tout --- un["un'"]
 homme --- x2["x"]
 un --- y1["y"]
 un --- femme["femme'"]
 un --- aimer["aimer'"]
 femme --- y2["y"]
 aimer --- x3["x"]
 aimer --- y3["y"]

```

La représentation des quantificateurs comme des prédicats à trois arguments est conséquente à l'utilisation d'une variable liée, qui n'a pas de contribution sémantique, mais uniquement un rôle structural en liant différentes positions de la formule. La variable n'est utile qu'en raison

de l'écriture linéaire<sup>1</sup> de la formule en (4). On peut en effet supprimer la variable en (5) en associant à (4) non pas un arbre, mais un dag<sup>2</sup>, c'est-à-dire une structure réentrante où les différentes occurrences de la variable correspondent à un seul nœud. Mais ce nœud fait alors double emploi avec le nœud de 'homme' et on peut considérer que la variable  $x$  n'est en fait qu'une réification de cet élément sémantique. Ceci est assez cohérent avec le fait que le sémantème 'homme', le signifié du nom *homme*, représente en fait une entité (un objet du monde), qui même s'il est indéterminé, n'est pas en soi un prédicat unaire (le prédicat unaire 'homme' que nous avons introduit correspond à l'expression *être un homme* plutôt qu'au nom *homme* proprement dit). Nous obtenons ainsi une nouvelle représentation pour (1)<sup>3</sup> :



Bien que directement liée aux représentations logiques les plus classiques, cette représentation n'a jamais été proposée à notre connaissance. Polguère (1992) propose une représentation très similaire, en ajoutant à une structure prédicative (voir caractérisation plus bas) un argument de portée pour les quantificateurs ; mais contrairement à ici, la portée est explicitement représentée comme une portion de la structure prédicative, alors que nous l'encodons, comme nous le verrons plus loin, par une relation hiérarchique.

Dans cette nouvelle représentation, les quantificateurs ('tout' et 'un') sont explicitement des opérateurs à deux arguments (cf. Barwise & Cooper 1981). En fait, même dans la représentation (4), en raison du caractère lié de la variable, le quantificateur est aussi un opérateur à deux arguments, dont on peut donner le lambda-terme  $\lambda P\lambda Q.[\text{tout}'(x,P(x),Q(x))]$ , ou même  $\lambda P\lambda Q.[\text{tout}'(\lambda x[P(x),Q(x)])]$ , en explicitant le caractère lié de la variable  $x$ .

Nous adopterons dorénavant la représentation proposée en (6), que nous appellerons la *représentation sémantique* de l'énoncé (1). Nous allons maintenant étudier plus en détail la structure de cette représentation. Elle combine en fait deux sous-structures : une structure prédicative et une structure hiérarchique liée à la portée.

Notons auparavant que notre représentation sémantique n'est pas moins riche que la représentation logique traditionnelle (en (2), (3) ou (4)), puisqu'on peut revenir à la représentation logique en réifiant (à la façon de Davidson 1967 pour les événements) les sémantèmes sur lesquels pointent les quantificateurs, c'est-à-dire en leur associant une variable, en leur ajoutant cette variable comme argument et en donnant cette variable comme argument aux prédicats pointant sur eux (voir section 3.4 pour un exemple).

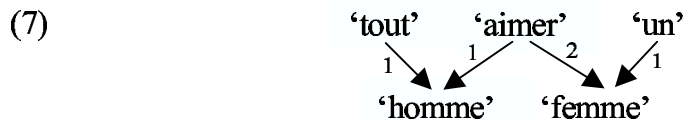
<sup>1</sup> On utilise aussi la variable pour désigner les référents de discours. Néanmoins, cet usage devrait être nettement séparé de l'autre, car dans ce deuxième cas la variable renvoie à un objet du monde et ne devrait pas figurer comme argument d'un sémantème, si l'on distingue sens et dénotation.

<sup>2</sup> On appelle *dag* un graphe orienté acyclique (de l'anglais *directed acyclic graph*).

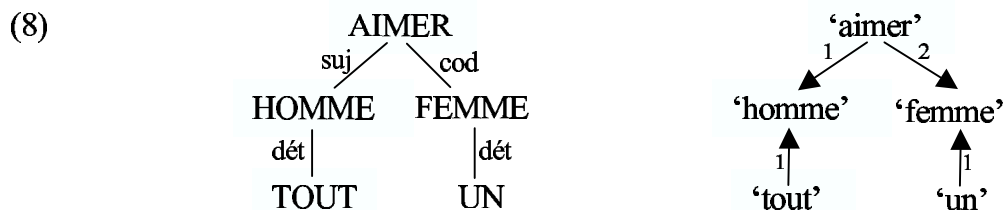
<sup>3</sup> La structure en (5) est un arbre ordonné, c'est-à-dire que les fils de chaque nœud sont ordonnés, indiquant ainsi pour chaque prédicat quel est son premier (deuxième ...) argument. En (6), nous adoptons une autre convention : les différents arcs du graphe sont étiquetés pour indiquer les différentes places argumentales. Par ailleurs, comme nous le verrons, ce graphe est partiellement hiérarchisé : les arcs droits correspondent à la structure arborescente sous-jacente (la structure hiérarchique logique, section 2.2).

## 2.2 Représentation logique et structure prédicative

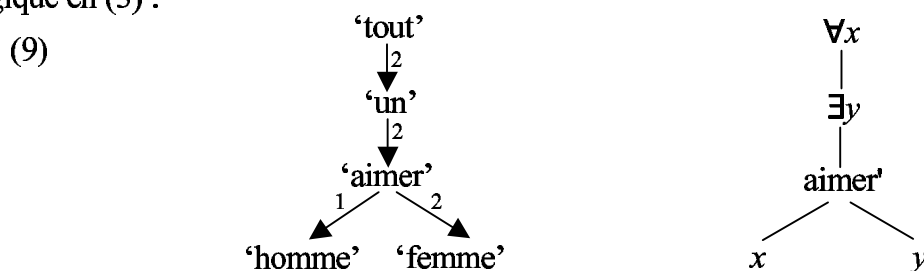
La *structure* que nous appelons *prédicative* (qu'on aurait pu aussi appeler structure argumentale) est un graphe de relations prédicat-argument, où les prédicats représentent des sémantèmes, c'est-à-dire les signifiés des éléments lexicaux de l'énoncé. Une telle structure est utilisée comme représentation sémantique par la théorie Sens-Texte (Mel'cuk 1988, Polguère 1992, Kahane 2001). Une relation prédicative correspond à une relation actancielle ou modificative et doit être « validée » par une relation de dépendance syntaxique<sup>4</sup>. En tant que prédicat sémantique, les quantificateurs n'ont qu'un argument : le nom qu'ils déterminent. La portée est d'une nature différente : cela n'aurait pas de sens de considérer qu'il existe en (1) une relation prédicative entre *tout* et *un*, car il n'y a aucune relation syntaxique entre eux. La structure prédicative de (1) est donc :



Par définition, la structure prédicative est très proche de la structure syntaxique, lorsque celle-ci est représentée par un arbre de dépendance syntaxique. Nous donnons en (8) l'arbre syntaxique de (1) ainsi que le graphe hiérarchisé obtenu en superposant au graphe en (7) la hiérarchie de l'arbre syntaxique.



On voit que la structure syntaxique est en quelque sorte une hiérarchisation de la structure prédicative. Or la représentation sémantique possède également une hiérarchie sous-jacente, mais celle-ci est différente et rend compte des relations de portée<sup>5</sup>. Nous donnons en (9) la *structure hiérarchique logique* de (1), ainsi que l'arbre de décomposition de la formule logique en (3) :



Comme on le voit en (9), c'est la structure hiérarchique que privilégie la représentation logique traditionnelle, au détriment de la structure prédicative. On peut noter des distorsions

<sup>4</sup> Il existe en fait un certain nombre de distorsions possibles entre la structure prédicative et la structure syntaxique (Mel'cuk 1988, Kahane 2001). C'est le cas par exemple des phénomènes de montée, comme dans *Pierre semble dormir*, où la relation prédicative entre 'dormir' et 'Pierre' est validée par la relation syntaxique sujet entre *semble* et *Pierre*.

<sup>5</sup> Nous laissons de côté une question fort intéressante, mais qui dépasse le cadre de cet article. Il s'agit des liens entre la structure informationnelle et la structure logique. La théorie Sens-Texte n'a par exemple jamais introduit de structure hiérarchique explicite, considérant que celle-ci n'est qu'une conséquence de la structure communicative et notamment de la partition thème-rhème (Polguère 1992, Mel'cuk 2003).

notables entre la hiérarchie logique et la hiérarchie syntaxique, sources de réelles difficultés pour l'interface sémantique-syntaxe, lorsque la structure sémantique repose sur la hiérarchie logique comme en (2)-(4). La structure logique est d'ailleurs plus directement liée à l'ordre linéaire qu'à la structure syntaxique, à notre avis. Ainsi, en (10)a et b, dans l'interprétation préférée<sup>6</sup>, 'tout' est dans la portée de 'un', à l'inverse de (1), alors que l'ordre des quantificateurs dans la phrase est également l'inverse de celui de (1) :

- (10) a. *Une femme est aimée de tout homme.*  
b. *Il y a une femme que tout homme aime.*

La représentation sémantique que nous avons adoptée peut être étendue, par exemple pour traiter des modificateurs. Dans *un homme heureux*, 'heureux' sera traité comme un prédicat unaire ayant comme argument 'homme' et se trouvant dans la portée de 'homme' (la relation prédicat-argument et la portée auront des sens opposés). Ce prédicat peut lui-même être modifié, par exemple, par 'très'. La portion de la représentation se trouvant dans la portée d'un sémantème représentant une entité restreint l'extension de ce sémantème et contribue à la restriction du quantificateur portant sur cette entité<sup>7</sup>.

Nous allons maintenant proposer une grammaire permettant de générer nos représentations sémantiques, avant de l'étendre pour en faire une interface sémantique-syntaxe.

### 3 Une grammaire pour les représentations logiques

Nous allons proposer une grammaire « lexicalisée » pour les représentations logiques, c'est-à-dire une grammaire qui construit la représentation en combinant des morceaux de structure associés aux différents sémantèmes. Nous nous tournons vers un formalisme capable de manipuler des graphes, les grammaires d'unification polarisées (Kahane 2004). Nous rappellerons brièvement ce formalisme avant de présenter notre grammaire sémantique.

#### 3.1 Grammaires d'unification polarisées

Les grammaires d'unification polarisées sont des grammaires permettant de générer des ensembles de structures finies. Une structure repose sur des *objets*. Par exemple, pour un graphe (orienté), les objets sont les nœuds et les arcs. Chaque arc est lié à deux nœuds par les

---

<sup>6</sup> Comme on le sait, il est possible qu'un quantificateur en deuxième position ait une portée large. Par exemple (10)a peut éventuellement recevoir la même interprétation que (1), mais une telle interprétation est plus difficile d'accès et doit être conditionnée par le contexte ou les connaissances du monde.

<sup>7</sup> Nous pouvons traiter d'autres types de restrictions, comme une relative ou une participiale. Ainsi, pour la phrase *Tout homme aimant une femme est heureux*, les relations de portées seront :

'tout' —2→ 'heureux' —1→ 'homme' —→ 'un' —2→ 'aimer' —2→ 'femme'.

La relation de portée entre 'homme' et la participiale est « validée » par la relation prédicative 'aimer' —1→ 'homme', qui est dans le sens inverse (comme pour un adjectif).

Les syntagmes nominaux complexes, comme dans *Tout locuteur de deux langues est heureux*, posent un problème intéressant. Ici les relations de portées sont :

'tout' —2→ 'deux' —2→ 'heureux' —1→ 'locuteur' —1→ 'langue'.

Il est également intéressant de comparer des exemples comme (i) *Les soldats fatigués se sont arrêtés* et (ii) *Les soldats, fatigués, se sont arrêtés*. La structure prédicative est la même ('s'arrêter' —1→ 'soldat' ←1— 'fatigué'), mais, en (i) seulement, 'fatigué' restreint la portée du quantificateur et doit être dans la dépendance logique de 'soldat'.

fonctions source et cible. Ce sont ces fonctions qui fournissent la structure proprement dite. Une *structure polarisée* est une structure dont les objets sont polarisés, c'est-à-dire étiquetés par une valeur appartenant à un ensemble fini  $P$  de polarités. L'ensemble  $P$  est muni d'une opération commutative et associative notée «  $\cdot$  », appelée *produit*. Un sous-ensemble  $N$  de  $P$  contient les polarités dites *neutres*. Une structure polarisée est dite *neutre* si tous ses objets sont neutres.

Nous allons utiliser un système de polarités  $P = \{\bullet, \circ, \ominus\}$ , avec  $N = \{\bullet, \ominus\}$ , et un produit défini par le tableau suivant (où  $\perp$  représente l'impossibilité de se combiner). Nous appellerons nos polarités  $\bullet$  = noir = saturation,  $\circ$  = blanc = contexte (obligatoire) et  $\ominus$  = gris = neutre absolu.

|           |           |           |           |
|-----------|-----------|-----------|-----------|
| $\cdot$   | $\bullet$ | $\circ$   | $\bullet$ |
| $\bullet$ | $\bullet$ | $\circ$   | $\bullet$ |
| $\circ$   | $\circ$   | $\circ$   | $\bullet$ |
| $\bullet$ | $\bullet$ | $\bullet$ | $\perp$   |

Les structures peuvent être combinées par *unification*. L'unification de deux structures  $A$  et  $B$  donne une nouvelle structure  $A \oplus B$  obtenue en « collant » ensemble ces structures par l'identification d'une partie des objets de la première structure avec ceux de la deuxième. Lorsque deux structures polarisées  $A$  et  $B$  sont unifiées, la polarité d'un objet de  $A \oplus B$  obtenu par identification de deux objets de  $A$  et  $B$  est le produit de leurs polarités.

Une *grammaire d'unification polarisée* (GUP) est définie par une famille finie  $T$  de types d'objets (avec des fonctions attachées aux différents types d'objets), un système  $(P, \cdot)$  de polarités, un sous-ensemble  $N$  de  $P$  de polarités neutres, et un ensemble fini de structures élémentaires polarisées, dont les objets sont décrits par  $T$  et dont une est éventuellement marquée comme la structure initiale (et appelée *top* dans la suite). Les structures *générées* par la grammaire sont les structures neutres obtenues par combinaison de l'éventuelle structure initiale et d'un nombre fini de structures élémentaires.

Rappelons que le formalisme est monotone (avec l'ordre  $\bullet < \circ < \ominus$  sur les polarités) et que les structures peuvent être combinées absolument dans n'importe quel ordre.

### 3.2 Une grammaire sémantique

Nous allons présenter notre grammaire en plusieurs étapes pour en faciliter la compréhension et pour en montrer le caractère élémentaire.

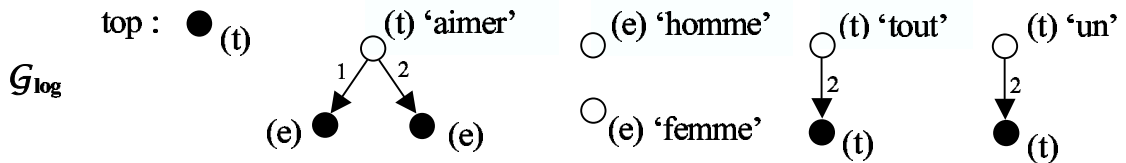
La structure prédictive est un graphe. Elle peut donc être générée par une grammaire aussi simple que  $G_{\text{préd}}$ , où chaque prédicat ancre un morceau du graphe avec ses arguments : le nœud du prédicat est saturé et les positions des arguments doivent être remplies et sont polarisées en blanc.



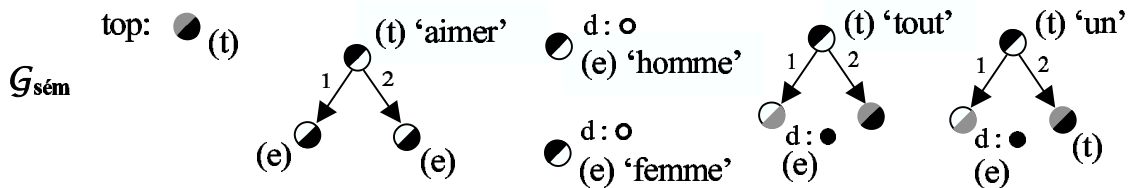
Nous avons ajouté une polarité supplémentaire  $d : \circ$  et  $d : \bullet$  sur certains nœuds. La polarité  $d : \circ$  indique que ces sémantèmes sont indéterminés. Un sémantème d'une entité déterminée, comme 'Paul', recevra une polarité  $d : \bullet$ . La polarisation multiple fonctionne ainsi : si un objet  $A$  de polarités  $(x, y)$  et un objet  $B$  de polarités  $(z, t)$  sont identifiés, l'objet résultant reçoit le couple de polarités  $(x, z, y, t)$ .

La structure logique est un arbre. La structure d'arbre est très facile à encoder avec une GUP (Kahane 2004) : il suffit d'assurer que chaque élément autre que la racine ait un unique

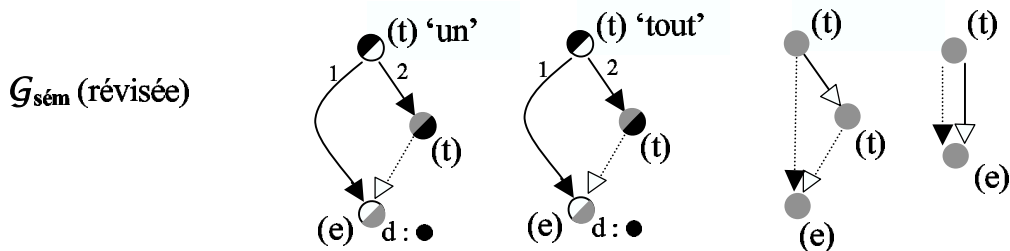
gouverneur (la connexité étant elle assurée par définition). Nous obtenons la grammaire  $\mathcal{G}_{\text{log}}$ , où chaque sémantème ancre la racine d'un sous-arbre. Nous typons, selon l'usage, les nœuds en deux classes : e pour les entités et t pour les propositions (*truth value*).



En superposant<sup>8</sup>  $\mathcal{G}_{\text{préd}}$  et  $\mathcal{G}_{\text{log}}$ , nous obtenons la grammaire sémantique  $\mathcal{G}_{\text{sém}} = \mathcal{G}_{\text{préd}} \times \mathcal{G}_{\text{log}}$ . Du fait de la superposition, chaque nœud hérite d'un couple de polarité. Un nœud qui n'apparaît pas dans une des deux grammaires est considéré comme neutre absolu là où il n'apparaît pas. Nous notons  $(\circ, \bullet) = \blacklozenge$  et  $(\bullet, \circ) = \blacklozenge$ . On remarquera que  $\blacklozenge$  et  $\blacklozenge$  se comportent comme des polarités opposés de type besoin-ressource.



Nous devons encore enrichir un peu  $\mathcal{G}_{\text{sém}}$  pour assurer que la restriction d'un quantificateur est dans sa portée. Une GUP permet facilement de contrôler le chemin entre deux nœuds : nous ajoutons pour cela un lien de dominance dans la structure du quantificateur entre la portée et la restriction. Une grammaire dédiée permet de « réécrire » ce lien de dominance en une chaîne de dépendance. On peut facilement contrôler la nature de cette chaîne, puisqu'une GUP peut simuler n'importe quelle grammaire de réécriture (Kahane 2004). Nous donnons ci-dessous la grammaire  $\mathcal{G}_{\text{sém}}$  révisée, avec les structures complètes pour les quantificateurs et, à droite, les deux règles nécessaires à la propagation de la dominance, où les liens de dominance sont représentés par des flèches pointillées.



### 3.3 Interface sémantique-syntaxe

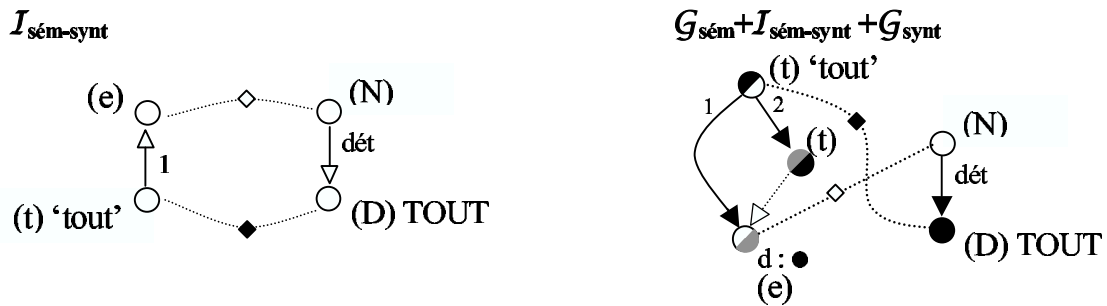
Nous appelons interface sémantique-syntaxe une grammaire capable de faire se correspondre une représentation sémantique et une représentation syntaxique. La difficulté de la tâche vient du fait qu'il y a entre les deux représentations des distorsions importantes au niveau de la hiérarchie entre les éléments<sup>9</sup>. Le problème a été étudié dans le cadre de nombreux

<sup>8</sup> Comme l'a fait remarquer un relecteur, la superposition des grammaires est un cas particulier de synchronisation (voir section 3.3). Celle-ci est possible quand les deux grammaires manipulent les mêmes objets.

<sup>9</sup> Le fait que notre représentation syntaxique soit un arbre de dépendance plutôt qu'un arbre syntagmatique ne change pas fondamentalement le problème. Nous préférons pour notre part traiter la question de l'ordre linéaire séparément et considérer une structure syntaxique sans ordre linéaire.

formalismes à commencer par la grammaire de Montague (1973). Celle-ci privilégie l'organisation logique, alors que des travaux plus récents tendent à favoriser l'organisation syntaxique, comme Copestake *et al.* 1999 avec HPSG ou Kallmeyer & Joshi 1999 avec TAG.

Pour notre part, nous résolvons le problème de manière triviale, en synchronisant notre grammaire sémantique avec une grammaire de dépendance syntaxique, sans privilégier aucun des deux niveaux de représentation. Seule la structure prédicative est prise en compte dans cette interface sémantique-syntaxe ( $I_{\text{sém-synt}}$ ). En combinaison avec  $G_{\text{sém}}$  et  $G_{\text{synt}}$ , elle permet de construire parallèlement les représentations sémantiques et syntaxiques<sup>10</sup>. En GUP, la synchronisation de deux grammaires s'effectue simplement en alignant les deux grammaires et en synchronisant certains nœuds (cf. Shieber & Schabes 1990, Perrier 2004, Kahane & Lareau 2005). La synchronisation est assurée par des liens dit de synchronisation qui obligeront, lorsqu'on identifie deux nœuds dans une des structures, à identifier les nœuds qui sont synchronisés avec eux dans l'autre structure. L'obligation d'unifier les liens de synchronisation (et donc de synchroniser les éléments qu'ils relient) est encore une fois assurée par la polarisation (représentée dans des losanges).



Notre grammaire est équivalente au niveau de l'expressivité aux autres interfaces sémantique-syntaxe évoquées et s'apparente tout particulièrement à la grammaire proposée par Perrier 2004, qui est à notre connaissance le premier à avoir utilisé une grammaire polarisée pour traiter les représentations logiques et la sous-spécification. Notre grammaire peut être utilisée pour l'analyse comme pour la synthèse et sera compatible avec diverses procédures. Une procédure pour une GUP consiste à donner un ordre dans le traitement des structures (par exemple suivre l'ordre imposé par la hiérarchie logique ou syntaxique) et choisir un ordre dans la neutralisation des polarités. L'un de ces ordres mérite quelques commentaires.

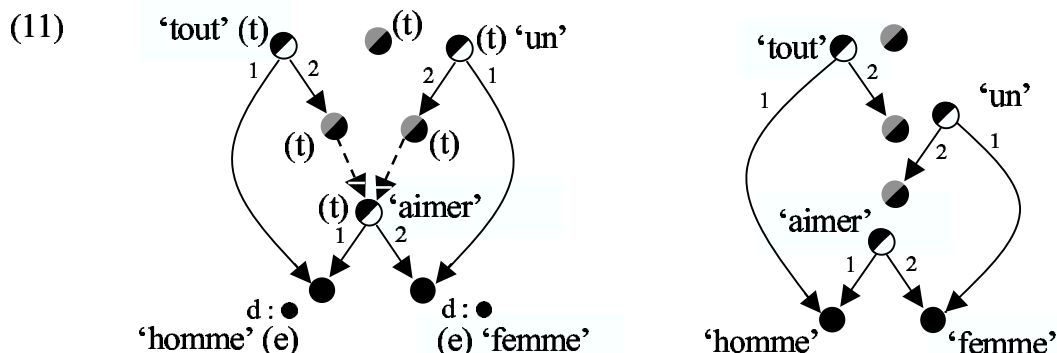
### 3.4 Sous-spécification et neutralisation partielle

Lorsqu'on utilise notre grammaire pour l'analyse, on remarque que la structure syntaxique ne permet pas de décider quelle doit être la portée respective des quantificateurs. Pour ne pas introduire une ambiguïté qu'on n'a pas les moyens de résoudre, on peut préférer construire une représentation sémantique où la hiérarchie logique reste sous-spécifiée. Cette question a été beaucoup étudiée dans la littérature récente (Reyle 1993, Bos 1995, Copestake *et al.* 1999, Kallmeyer & Joshi 1999, Perrier 2004). Dans notre cadre, la solution est élémentaire : il suffit de ne pas chercher à neutraliser les polarités qui contrôlent la hiérarchie logique (c'est-à-dire

<sup>10</sup> Notre grammaire syntaxique  $G_{\text{synt}}$  a été décrite dans de nombreux articles (notamment Kahane 2001 et Kahane & Lareau 2005). Le fait que les noms doivent avoir un unique déterminant est ici assuré par la polarité d de la grammaire sémantique. Il faut néanmoins noter que des éléments sémantiquement déterminés comme 'Paris' ou 'la France' peuvent être ou non syntaxiquement déterminés (*la France* vs *en France*) et qu'un traitement syntaxique de la détermination doit venir compléter le traitement purement sémantique.



de ne pas considérer  $G_{\text{préd}}$  lors de l'analyse). On obtient, pour (1) ou (10), la *représentation sémantique sous-spécifiée* (11) (en version simplifiée à droite, en mettant en vis-à-vis les nœuds qui fusionneront pour obtenir (6)), où comme on le voit les seules polarités non neutres sont les polarités blanches associées à la hiérarchie logique :



Autrement dit, l'interface syntaxe-sémantique est uniquement réalisée entre la structure syntaxique et la structure prédictive. C'est ce que préconise depuis les années 60 la théorie Sens-Texte, originellement conçue pour la traduction automatique, où la levée des ambiguïtés logiques est généralement inutile.

Notons que la représentation sémantique en (11) peut être traduite en une écriture linéaire en réifiant les différents nœuds :  $p$  pour la racine,  $p_1$  et  $p_2$  pour les quantificateurs 'tout' et 'un',  $e$  pour le prédicat 'aimer',  $x$  et  $y$  pour les entités 'homme' et 'femme' et  $h_1$  et  $h_2$  pour les portées des quantificateurs ( $h$  pour *hole*, d'après la Hole Semantics de Bos 1995) :

$$(12) \quad \text{top}:p, p_1:\text{'tout'}(x,h_1), p_2:\text{'un'}(y,h_2), e:\text{'aimer'}(x,y), x:\text{'homme'}, y:\text{'femme'}$$

ou encore avec une double réification des entités (en introduisant les variables  $a_1$  et  $a_2$ ) :

$$(13) \quad \text{top}:p, p_1:\text{'tout'}(x,a_1,h_1), p_2:\text{'un'}(y,a_2,h_2), e:\text{'aimer'}(x,y), a_1:\text{'homme'}(x), a_2:\text{'femme'}(y)$$

Les relations de dominance, ainsi que la nécessité de neutraliser les trous  $h_1$  et  $h_2$ , seront données par des conditions telles que :

$$(14) \quad p \geq p_1 > h_1 \geq e, p \geq p_2 > h_2 \geq e, \{p, h_1, h_2\} = \{e, p_1, p_2\}$$

Même si les deux représentations, (11) et (13)+(14) (qui est la représentation de Copestake *et al.* 1999) sont (quasiment) équivalentes, on peut voir un avantage à utiliser des structures polarisées et à ne pas utiliser d'écriture linéaire forçant l'introduction de multiples variables.

## 4 Conclusion

Notre contribution concerne la représentation sémantique des énoncés et tout particulièrement la structure logique, c'est-à-dire la représentation des phénomènes de portée. Nous n'avons pas réellement abordé de questions de logique, c'est-à-dire la façon dont nos représentations pouvaient servir directement à un calcul logique. De ce point de vue, nous nous sommes couverts par la possibilité de revenir aux formules usuelles de la logique du premier ordre, mais une description directe de l'inférence dans notre formalisme serait certainement intéressante. Notre objectif était d'avoir une représentation qui soit à la fois suffisante pour l'interprétation logique, mais qui permette aussi un traitement adéquat d'autres phénomènes sémantiques qui ne se manifestent pas en termes de portée.

Le véritable point de notre contribution est d'avoir rappelé que la structure sémantique des énoncés n'est pas linéaire et qu'il y a tout intérêt dans ces conditions à utiliser un langage qui

permette de manipuler directement des arbres, des graphes et des produits des deux. Nous avons ainsi montré qu'une représentation sémantique est la superposition d'une structure prédicative et d'une structure hiérarchique encodant les relations de portée. Lorsqu'on ne privilégie plus la relation de portée, comme le fait la logique classique, il devient facile de définir une représentation sous-spécifiée en « neutralisant » une des deux structures seulement.

## Remerciements

Je remercie chaleureusement pour leurs commentaires Pascal Amsili, Laurence Danlos, Laura Kallmeyer, François Lareau, Guy Perrier, Alain Polguère, Igor Mel'cuk, Benoît Sagot et les trois relecteurs de TALN.

## Références

- BARWISE J. & COOPER R. (1981), Generalized quantifiers and natural language, *Linguistics and Philosophy*, 4, 159-219.
- BOS J. (1995), Predicate Logic Unplugged, *Tenth Amsterdam Colloquium*.
- COPESTAKE A., FLICKINGER D. & SAG I. A. (1999), Minimal Recursion Semantics : An Introduction, draft, 26 p.
- DAVIDSON D. (1967), The Logical Form of Action Sentences, in N. Rescher (ed.), *The Logic of Decision and Action*, University of Pittsburgh Press, 81-95. Reprinted in D. Davidson, *Essays on Actions and Events*, Oxford: Clarendon Press, 1990, 105-122.
- DYMETMAN M., COPPERMAN M. (1996), Extended dependency structures and their formal interpretation, *Proceedings CoLing*, Copenhagen.
- KAHANE S. (2001), Grammaires de dépendance formelles et théorie Sens-Texte, Tutoriel, *Actes TALN*, vol. 2, 17-76.
- KAHANE S. (2004), Grammaires d'unification polarisées, *Actes TALN*, Fès, 233-242.
- KAHANE S. & LAREAU F. (2005), Grammaire d'Unification Sens-Texte : polarisation et modularité, *Actes TALN*, Dourdan, 10 p.
- KALLMEYER L. & JOSHI A. (1999), Factoring predicate argument and scope semantics: Underspecified semantics with LTAG, *Proceedings of the 12th Amsterdam Colloquium*.
- MEL'CUK I. (1988), *Dependency Syntax: Theory and Practice*, SUNY Press.
- MEL'CUK I. (2003), *Communicative Organisation of Natural Language*, Benjamins.
- MONTAGUE R. (1973), The proper treatment of quantification in ordinary English, in J. Hintikka (éd.), *Approaches to Natural Language*, 221-242, Reidel.
- PERRIER G. (2004), La sémantique dans les grammaires d'interaction, *Actes TALN*, Fès, Maroc, 351-360.
- POLGUERE A. (1992), Remarques sur les réseaux sémantiques Sens-Texte, in A. Clas (éd.), *Le mot, les mots, les bons mots*, Presses de l'Université de Montréal.
- REYLE U. (1993), Dealing with ambiguities by underspecification, *Journal of semantics*, 10, 123-179.
- SHIEBER S. M. & SCHABES Y. (1990), Synchronous tree-adjointing grammars, *Proceedings CoLing*, vol. 3, 253-258, Helsinki, Finland.
- WOODS W.A. (1975), What's in a link: Foundations for semantic networks, in D. Bobrow & A. Collins, *Representation and Understanding—Studies in Cognitive Science*, 55-82, Academic Press, Orlando.