

## Discourse Representation Theory et graphes sémantiques : formalisation sémantique en contexte industriel

Maxime Amblard   Johannes Heinecke   Estelle Maillebuau  
Orange Labs, 2 av. Pierre Marzin, 22307 Lannion cedex  
{maxime.amblard, johannes.heinecke,  
estelle.maillebuau}@orange-ftgroup.com

**Résumé.** Ces travaux présentent une extension des représentations formelles pour la sémantique, de l’outil de traitement automatique des langues de Orange Labs<sup>1</sup>. Nous abordons ici uniquement des questions relatives à la construction des représentations sémantiques, dans le cadre de l’analyse linguistique. Afin d’obtenir des représentations plus fines de la structure argumentale des énoncés, nous incluons des concepts issus de la DRT dans le système de représentation basé sur les graphes sémantiques afin de rendre compte de la notion de portée.

**Abstract.** This works present an extension of the formal representation of semantic for the natural language processing tool developped at Orange Labs. We cover here only issues relating to build the semantic representations, based on linguistic parsing. To obtain more detailed representations of the argumental structure statements, we include insights from the DRT in the system of representation based on the semantic graphs to give an account of scope.

**Mots-clés :** Modèle sémantique, analyse syntaxique en dépendance, DRT.

**Keywords:** Semantic modelling, parsing, DRT.

Un grand nombre d’applications en lien avec les technologies récentes (internet, *etc.*) s’orientent vers des thématiques dites sémantiques. Nous souhaitons associer automatiquement à des données informatisées une interprétation plus large que ce que peut apporter un lexique sémantique standard, tout en maintenant une forte cohérence dans les systèmes de représentation. Bien que la terminologie *sémantique* reste actuellement ambiguë, nous souhaitons nous inscrire dans ces prospectives. Nous proposons une extension à la notion de portée du modèle de représentation sémantique utilisé dans l’outil de traitement automatique des langues de Orange Labs.

Des développements, conduits depuis plus d’une décennie ont permis de proposer un système d’analyse et de génération, appelé TiLT. Chaque traitement linguistique propose des degrés de finesse à ajuster qui permettent son utilisation pour des tâches de TAL diverses. Nous nous sommes focalisés sur l’interface entre syntaxe et sémantique. Ici, la sémantique rend compte de relations via des propriétés logiques, dans la continuité de (Retoré, 2008) et (Blackburn *et al.*, 1999), ainsi que de (Asher & Lascarides, 2003) pour les aspects dynamiques.

Nous utilisons les analyses syntaxiques fournies par l’analyseur TiLT, ainsi que les premières représentations sémantiques. Le modèle choisi est basé sur la notion de graphe conceptuel, (Sowa, 1964). Cette théorie est largement développée comme outil de modélisation de la logique. Afin d’étendre le modèle de représentation nous nous sommes orientés vers la Discourse

---

<sup>1</sup>France Télécom R&D

Representation Theory (DRT), ([Kamp & Reyle, 1993](#)). Cette théorie propose une modélisation simple des notions de portée que nous souhaitons inclure.

Nous illustrerons les différentes représentations précédentes et nouvelles propriétés proposées au travers de l'exemple principal de cet article :

(1) Paul avale un médicament

Nous reviendrons dans un premier temps sur les cadres généraux utilisés (section 1) : contexte de ces développements, graphes sémantiques et DRT. Nous présenterons l'interface syntaxe-sémantique mise en œuvre dans TiLT (section 2). Puis nous expliciterons les choix des concepts issus de la DRT pour la modélisation sémantique qui augmentent la pertinence du modèle (section 3). Enfin nous conclurons en mettant en perspective d'autres utilisations possibles de la DRT dans notre modèle (section 4).

# 1 Cadres généraux : applicatif et formel

## 1.1 Contexte industriel

Les recherches présentées ici sont le prolongement de différents développements réalisés ces dernières années. La question de la modélisation sémantique est à la fois fortement dépendante des représentations syntaxiques, et également corrélée à la structuration des données lexicales. Toutefois, le calcul de représentations sémantiques reste incontournable pour des systèmes de TAL, comme ceux développés par Orange Labs.

La première perspective d'utilisation du niveau sémantique envisagée a été le développement d'un outil de traduction automatique. Le but était de partager des représentations abstraites de la distribution argumentale des prédicats associés aux mots d'un énoncé d'une langue source, afin de les transposer dans une langue cible. Le résultat respecte alors la distribution de départ. Le niveau sémantique a été utilisé et évalué comme substitut à un pivot langagier, ([Iheddadene, 2007](#); [Park et al., 2007](#)), pour une application intervenant dans la relation entre un client francophone et un pharmacien anglophone, outil où la fiabilité est nécessaire. Cependant, cette perspective est actuellement en veille car le modèle sémantique pose des questions structurelles.

Ce système est également utilisé dans une application d'aide à la communication entre personnes sourdes et entendant, ([Kervajan et al., 2007](#)). Pour transcrire en séquence signée un énoncé d'une langue parlée, la simple analyse syntaxique ne suffit pas. Nous avons besoin de mettre en œuvre des concepts plus élaborés. Les représentations sémantiques permettent de récupérer tout ou partie de cette information.

Enfin, un troisième aspect de la modélisation sémantique est utilisé dans la production d'outil liant représentations abstraites (du type ontologique) et données primaires. Une application permettant de maintenir et rechercher des documents dans une grande masse de données à partir de concepts de haut niveau a été développée, ([Heinecke, 2006](#)).

## 1.2 Graphes Sémantiques

La représentation choisie est issue de la théorie des graphes conceptuels de Sowa, ([1964](#)). Cette théorie a pour but de proposer une représentation sous forme de graphes des formules logiques.

L'origine de ces travaux était de proposer une modélisation pour le TAL.

L'avantage que propose de telles représentations est sans conteste une plus grande facilité de lecture et de représentation des éléments intervenant dans ces structures. De plus, cette communauté a également proposé un ensemble d'opérations sur ces graphes permettant efficacement la mise en relation entre eux. Ces propriétés permettent ainsi facilement de relier un graphe associé à un énoncé, à un graphe qui est une modélisation abstraite des relations entre concepts<sup>2</sup>.

Cependant, si les relations prédicats/variables des formules logiques trouvent naturellement leur correspondance dans la représentation sous forme de graphe, il est moins aisé de rendre compte de certains concepts de plus haut niveau, en particulier ce qui a trait à la modélisation des portées des opérateurs. Ces perspectives n'avaient pas été retenues dans la modélisation initiale.

Les graphes sémantiques portent sur des éléments appelés **individus** qui matérialisent un objet physique ou abstrait, une situation, un événement de la représentation. On leur associe alors un numéro unique. Les relations entre individus sont représentées au travers de **prédicats**. Ces derniers ont pour fonction de matérialiser le *signifiant* d'un mot ou d'un morphème. Chacun possède des *participants* (individus) en nombre constant, cependant ils n'apparaissent dans la représentation que dans le cas où ils sont spécifiés. Chacun reçoit alors un rôle thématique ou une fonction prédéfinie par le prédicat.

Cette dernière propriété les différencie de la définition logique traditionnelle de prédicat. Pour marquer cette association entre attribution thématique et argument du prédicat, nous utilisons le symbole « = ». Il faut alors voir cette relation intra-argument comme une forme de transtypage des types sémantiques de base associés aux individus vers les rôles thématiques.

Ainsi, on peut définir formellement un **graphe sémantique** qui représente partiellement le contenu informationnel d'un énoncé en langue naturelle sous forme de graphe orienté acyclique.

Soit  $G$  un graphe sémantique,  $G$  est un quadruplet tel que  $G = \langle V_{pred}, V_{ind}, E, \phi \rangle$  où :

- $V_{pred}$  est l'ensemble des nœuds représentant des prédicats.
- $V_{ind}$  est l'ensemble des nœuds représentant des individus.
- $E$  est l'ensemble des arêtes
- $\phi$  est une fonction de  $E \rightarrow (V_{ind} \times V_{pred})$  qui associe les arêtes à des nœuds.

Les nœuds représentent des prédicats ou des individus et les arêtes les liens entre ces éléments étiquetés par le rôle ou la fonction attribuée. On note  $\mathcal{G}$  l'ensemble des sous-graphes de  $G$ .

### 1.3 Discourse Representation Theory - DRT

La DRT a été conçue pour rendre compte de deux types de problèmes connexes :

1. proposer un mécanisme d'unification des variables. L'idée principale est de proposer un système de représentation permettant la modélisation des relations discursives comme la résolution des anaphores pronominales (intra- et inter-phrastique),
2. rendre compte d'un problème classique relatif aux *Donkey sentences*<sup>3</sup>.

Ces perspectives ont permis d'élaborer un formalisme qui utilise des formules logiques pour proposer une représentation sémantique, ainsi qu'une interface entre sémantique et discours.

<sup>2</sup>Ici, cette représentation est un type d'ontologie.

<sup>3</sup>Les Donkey sentences sont des phrases où les interactions entre la portée d'un quantificateur universel et d'un existentiel sont complexes, par exemple : *tout fermier qui possède un âne le bat*.

Les évolutions de la DRT ont été menées en intégrant les événements et leurs relations dans la construction du discours. On trouvera une présentation détaillée de leur utilisation dans (Busquets *et al.*, 2001). Cette théorie a fait l’objet de nombreuses propositions, en particulier sur la construction des structures de représentation, (Amsili & Roussarie, 2004), et l’utilisation de calculs élaborés, par exemple la sous-spécification dans (Schilder, 1998).

Afin de modéliser les relations entre les éléments présents dans l’énoncé, la DRT utilise la logique du premier ordre. Ainsi, à chaque mot est associé un prédicat au sens traditionnel. Les liens entre les prédicats sont réalisés par des variables.

Pour l’exemple (1), nous obtiendrons les prédicats  $avaler(e, x, y)$ ,  $médicament(z)$  et  $Paul$ . L’unification entre les variables est déléguée à un mécanisme externe. Les variables mises en jeu dans cette construction sont déclarées grâce à des quantificateurs qui sont représentés par des Discourse Representation Structures (DRS). Les DRS sont des structures composées de deux parties : la première exhibant une liste de variables relatives à la DRS et la seconde la liste des formules sur lesquelles les variables ont portée.

Ainsi, on donne de manière effective la portée d’un quantificateur sur les variables d’une description. La figure 1 présente les DRS pour les quantifications existentielle et universelle, correspondant aux syntagmes *un médicament* et *les médicaments*.

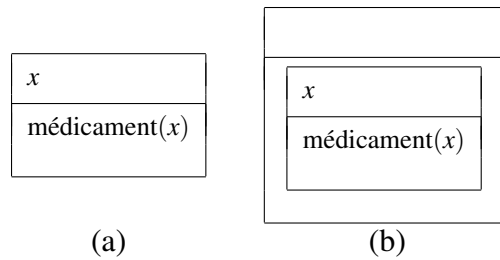


FIG. 1 – Représentation sous forme de DRS de la quantification existentielle (a) du syntagme «un médicament» et de la quantification universelle (b) du syntagme «les médicaments»

Afin de représenter de manière simplifiée ces DRS, nous reprenons la notation utilisée dans (Amblard, 2007) à partir des crochets. La quantification existentielle est notée :  $[x|médicament(x)]$  et la quantification universelle :  $[[x|médicament(x)]]$ .

Un objectif important de la DRT est de modéliser les relations de discours. Pour cela, elle adopte la position Davidsonienne, (Davidson, 1967), et réifie (désigner l’action ou le fait par une variable particulière) les formules. Elle associe des relations discursives entre variables d’événement. Nous n’intégrons pas ces propriétés dans nos représentations, mais supposons que leur intégration, au moins pour les relations intra-phrastiques entre événements, est une prolongation pertinente de ces travaux.

## 2 Fonctionnement des représentations sémantiques

Les mises en œuvre de l’analyse sémantique offrent des perspectives nombreuses et diverses. Nous nous attacherons dans cet exposé uniquement à ce qui concerne l’analyse compositionnelle de la sémantique, laissant les questions liées à la sémantique du discours, ainsi que celles relatives à la sémantique lexicale. L’outil principal que nous développons fournit une représentation sémantique. La modélisation choisie en utilise un fragment pour donner des graphes

sémantiques. Ils ont pour but de fournir une représentation des relations prédicatives extraites à partir de l'analyse syntaxique des énoncés. Ce sont des graphes orientés acycliques. Pour parvenir à ce calcul, plusieurs types de données et représentations doivent être mobilisées, en particulier les données lexico-sémantiques qui sont regroupées et organisées dans des lexiques et un *thesaurus* où des prédicats sont associés aux entrées lexicales. Cet ensemble permet de spécifier des sens différents en fonction de l'utilisation faite des mots.

Pour l'énoncé (1), on utilise un prédicat associé au mot « avaler » qui prend deux arguments. L'un en relation d'**agent**, en l'occurrence *Paul*, et l'autre en relation de **thème**, ici *un médicament*. Les assignations de rôles thématiques sont alors vues comme un trans-typage des arguments. Donc, ce prédicat est: *OPERATION.aval*(agent= $x_{19}$ , situation= $x_{17}$ , thème= $x_{18}$ ). Nous avons fait le choix d'introduire des arguments servant à représenter les événements décrits (situations). Comme nous ne construisons pas des formules, nous ne parlerons pas de réification. Cependant, nous adoptons la position Davidsonienne, précédemment évoquée. Le problème est d'unifier les variables  $x_{17}$ ,  $x_{18}$ ,  $x_{19}$  qui dénotent des individus de la représentation. Ainsi, l'agent du prédicat *OPERATION.aval* doit avoir le même référent que l'entité *npr:Paul*( $x_{19}$ ).

Le graphe de la figure 2 est construit à partir de l'analyse syntaxique (en dépendance) réalisée par le système TiLT. Pour cela, l'analyse sémantique utilise les données du type catégorie, traits morpho-syntaxiques, etc. Un ensemble de règles se charge de créer ces relations. Pour

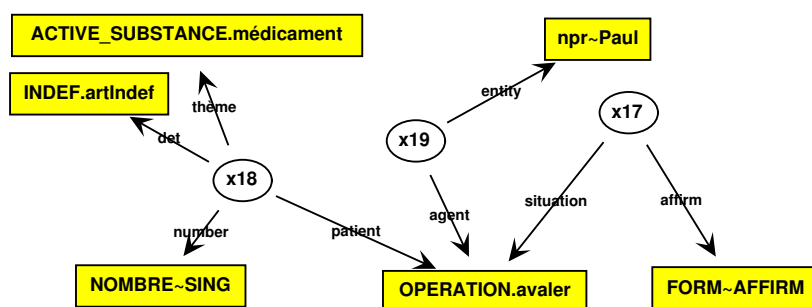


FIG. 2 – Graphe Sémantique de « Paul avale un médicament ».

chaque fonction syntaxique détectée dans l'arbre de dépendance, le système cherche une règle qui lui correspond. Par exemple, pour la relation de dépendance *SUJ* entre *npr:Paul* et *OPERATION.aval* correspond la règle:

$$\text{SUJ} : P/\text{subj} = D/\text{narg}_0$$

Cette règle peut s'interpréter comme «dans la relation *SUJ*, les variables doivent être unifiées entre l'item qui a la fonction sujet dans la principale et l'item qui a la fonction  $\text{narg}_0$  du dépendant». Il s'agit alors d'identifier à quelles variables correspondent *subj* et  $\text{narg}_0$ . Cette information est contenue dans un trait particulier de position fonctionnelle, appelé *FONCPOS*. La valeur de ce trait spécifie à quelle position argumentale correspond chaque fonction syntaxique:

$$\text{Paul} : \text{FONCPOS}/\text{narg}_0 = P_0$$

$$\text{aval} : \text{FONCPOS}/\text{varg}_0 = P_0/\text{subj} = P_1/\text{objd} = P_2$$

L'argument  $P_0$  du prédicat *npr:Paul* correspond au même individu que celui de la position  $P_1$  du prédicat *OPERATION.aval*. L'information sémantique correspondant à ces positions se trouve dans le *thesaurus*. Ainsi, on est en mesure de reconstruire la règle entièrement:

$$\text{SUJ} : P/(\text{subj}-P_1-\text{agent}) = D/(\text{narg}_0-P_0-\text{entity})$$

### 3 Extension des graphes sémantiques

Les graphes sémantiques permettent de retrouver une des parties importantes de l'information à modéliser pour la sémantique compositionnelle : les liens entre variables qui interviennent dans les structures argumentales des prédicats. Cependant, aucun mécanisme ne permet structurellement de rendre compte de la notion de portée. On retrouve cette question tant dans la modélisation des quantificateurs, qui est classique en sémantique, que pour celle de la négation. Bien qu'il soit complexe de proposer une réponse globale pour les phénomènes liés à la négation, même pour des exemples élémentaires, cette modélisation échoue à proposer des réponses cohérentes. Nous proposons dans cette section de revenir sur ces deux aspects en présentant une modélisation basée sur les graphes sémantiques qui reprend des propriétés de la DRT. Nous appelons DRS-graphes ces nouvelles structures.

#### 3.1 Quantification dans les DRS-graphes

Du point de vue de la modélisation sémantique, la particularité de la DRT est de modéliser de manière simple la portée d'un opérateur par des DRS. Nous reprenons ce concept que nous ajoutons à notre modèle. Ainsi, au lieu de représenter uniquement une variable par un nœud du graphe de  $V_{ind}$ , nous lui associons une portée si celle-ci fait l'objet d'une quantification. La portée est alors représentée du point de vue formel par le biais d'une fonction et par une boîte pour la représentation graphique.

Soit  $G \in \mathcal{G}$  un graphe sémantique, nous construisons le DRS-graphe  $H$  en associant au quadruplet  $G$  une fonction  $\mathcal{P}$  de  $V_{ind} \rightarrow \mathcal{G}$  qui représente la portée des quantifications. Elle associe à tout élément de  $V_{ind}$  (les individus de la représentation) un graphe  $G' \in \mathcal{G}$  (sous-graphe du graphe initial  $G$ ). Les DRS-graphes sont donc des quintuplets contenant un graphe sémantique et une fonction de  $V_{ind} \rightarrow \mathcal{G}$ . On remarquera que cette définition ne contraint pas les portées possibles et impossibles. De fait, un DRS-graphe est un ensemble ordonné de graphes.

Dans le cas général, les portées sont récursivement contenues les unes dans les autres. Après identification de structures de portées complexes par exemple pour des énoncés contenant plusieurs informations du discours, nous intégrerons une heuristique de normalisation des formes obtenues pour intégrer une représentation simplifiée (unique) dans nos outils.

En reprenant l'exemple (1), nous ne souhaitons plus seulement obtenir le graphe de la figure 2, mais une représentation graphique où les portées sont représentées par des boîtes, comme présenté dans le graphe de la figure 3. Dans cet exemple, nous modélisons les noms propres non pas par une constante du système, mais par un prédicat désignant l'ensemble des propriétés les décrivant. Implicitement, on réalise un type-raising<sup>4</sup> systématique sur ces catégories.

Apparaît alors un problème classique concernant la portée des quantificateurs et qui justifie la production d'un ensemble de résultats : la présence de plusieurs d'entre eux dans un même énoncé peut impliquer que les portées différentes génèrent des interprétations différentes. Par exemple, dans l'énoncé « les musées reçoivent une exposition », si le quantificateur des musées domine celui des expositions, il y a autant de musées que d'expositions ; dans le cas contraire, il y a une unique exposition qui est reçue par tous les musées.

<sup>4</sup>ou *montée de type*, ce qui correspond à la modification du statut foncteur/argument dans le système de types, permettant de considérer un objet par sa définition en intension (l'ensemble de ses propriétés).



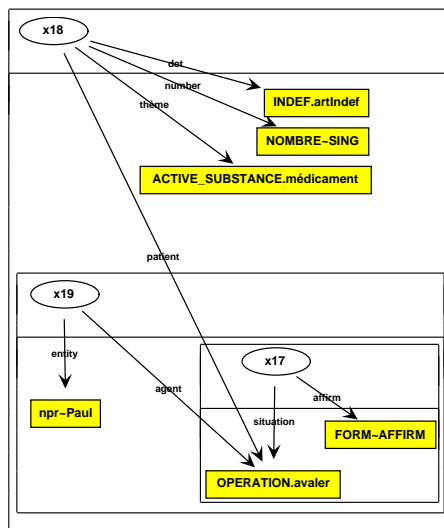


FIG. 3 – Représentation sous forme de DRS-graphe de l'exemple.

Actuellement, nous ne nous attachons pas à proposer une interprétation des résultats fournis par les DRS-graphes, mais une modélisation des phénomènes présents.

Nous introduisons la notion de DRS *ouvertes* qui permet de rendre compte de la portée partielle, c'est-à-dire de relations sous-spécifiées de portée. Ainsi, au cours de la construction de la représentation sémantique, nous associons à un individu quantifié non pas le graphe entier sur lequel il a portée dans la représentation finale, mais le sous-graphe sur lequel il a nécessairement portée. Pour modéliser les DRS ouvertes, lorsque nous ferons appel à la notation simplifiée des DRS, nous n'inclurons pas le crochet fermant.

En reprenant l'exemple, nous construisons :

$$(i) [x_{18} \mid OPERATION.avalier(x_{17}, x_{18}, x_{19}) \wedge \{x_{17}\} \wedge \{x_{18}\}]$$

où nous notons  $\{y\}$  l'ensemble des prédicats utilisant uniquement la variable  $y$ . D'autre part, nous construisons :

$$(ii) [x_{19} \mid OPERATION.avalier(x_{17}, x_{18}, x_{19}) \wedge \{x_{17}\} \wedge \{x_{19}\}]$$

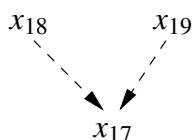


FIG. 4 – Portées sous-spécifiées de l'exemple 1.

La représentation sous forme d'arbre de ces relations sous-spécifiées peut-être vue de manière classique comme dans la figure 4. Ainsi, on représente que  $(i) \cap (ii) = \{x_{17}\}$ .

La réduction des relations sous-spécifiées permet alors de fournir plusieurs DRS-graphes pour chacune des portées possibles. Les algorithmes traditionnels de la DRT font souvent appel à la structure hiérarchique des DRS, notamment pour la résolution d'anaphore. Même si cette structure ne se retrouve que partiellement au fur et à mesure de la construction des DRS-graphes, la fonction de portée et les intersections de DRS-graphes permettent de retrouver la hiérarchie des DRS en jeu.

Cependant, nous proposons un critère de normalisation afin de ne pas démultiplier les solutions proposées par nos applications. Nous supposons que la forme normale d'un DRS-graphe (ou le représentant unique d'une classe de DRS-graphe) est celle où les quantificateurs respectent l'ordre linéaire gauche droite, à l'exception des descriptions des noms propres, et où la quantification événementielle ne porte sur aucune autre variable.

Pour l'exemple 1, c'est le DRS-graphe de la figure 3 qui sera proposé. On appelle **extension sémantique** d'un DRS-graphe  $H$  l'ensemble des DRS-graphes partiels appartenant à la même classe sémantique que  $H$ . La figure 5, sur laquelle nous reviendrons, montre un DRS-graphe (fig 5 (b)) et un membre de l'extension sémantique de ce DRS-graphe (fig 5 (b')). Nous revenons dans la section suivante sur la quantification des événements au travers de la question de la modélisation de la négation.

### 3.2 Négation dans les DRS-graphes

Du point de vue de la DRT, il est nécessaire de réifier les graphes sémantiques (pour ne pas dire les formules) et ainsi adopter la position Davidsonienne, (Davidson, 1967), en désignant l'action ou le fait par une variable particulière. Dans les DRS-graphes, nous respectons cette position comme nous l'avons précédemment évoqué. Dans ce cas, les événements ont un statut particulier dans la description de la structure à construire. Mais si ce statut est particulier, il n'en est pas différent et ces variables appartiendront à l'ensemble  $V_{ind}$ .

Nous supposons que la quantification standard pour ces variables est la quantification existentielle et que la modification du type de quantification doit être justifiée par des indices linguistiques présents dans l'énoncé. Dans nos graphes sémantiques traditionnels, la négation est un prédicat qui porte sur la variable de réification. Dans la transcription de ces graphes sous forme de DRS-graphes, les implications de la modélisation de la portée sont directes sur l'interprétation de la négation. En effet, si on analyse la version négative de l'exemple 1 :

(2) Paul n'avale pas de médicament.

On s'aperçoit que l'extension sémantique n'est pas réduite à un unique représentant, c'est-à-dire que plusieurs DRS-graphes peuvent être construits. Cependant, en supposant que la quantification sur la variable *Paul* ne porte pas d'enjeu, seulement deux sont pertinentes:

1. l'un où la quantification événementielle a portée sur toute la construction (fig 5.(b)),
2. l'autre où il n'a portée que sur les prédicats n'utilisant que cette variable (fig 5.(b')).

Dans le premier cas, le fait que l'événement soit nié implique que l'ensemble des descriptions sur lesquelles il a portée le sont aussi. Ainsi, le fait qu'il n'existe pas de médicament est une condition qui rend la modélisation recevable. Or il n'est pas acceptable pour une modélisation pragmatique de supposer qu'il n'existe pas l'une des entités définies, à moins que cela soit l'expression d'une figure stylistique. La négation d'un événement  $x$  ne doit alors porter que sur l'événement lui-même (et sur les prédicats ne portant que sur cette variable ( $\{x\}$ )).

Nous supposerons que l'impact de la négation sur la construction des DRS-graphes est la fermeture de la portée de l'événement sur lequel il porte. Ainsi, l'ambiguïté de portée de quantification est déléguée à la résolution de la structure sous-spécifiée finale, mais cette procédure la réduit. Si ce phénomène est important dans l'interprétation de la négation simple, il l'est également pour tout ce qui a trait à la description de l'événement. La négation des prédicats utilisés



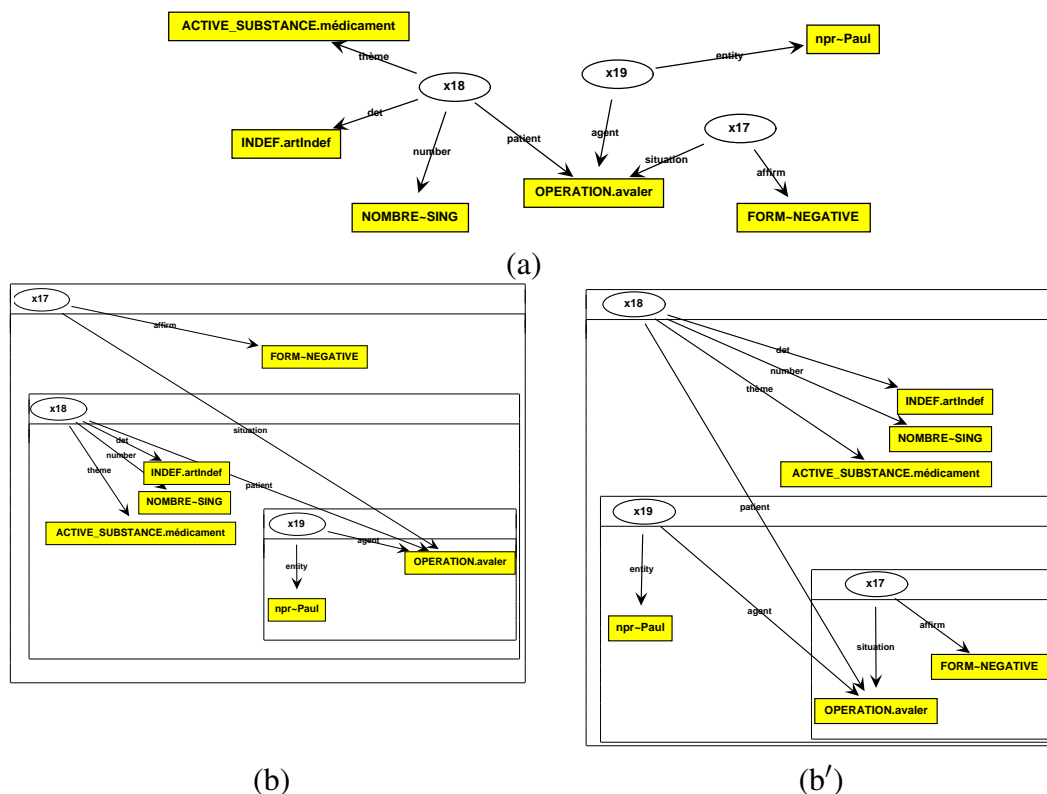


FIG. 5 – Graphe sémantique pour la négation (a) et de DRT-Graphes équivalents (b) et (b').

pour spécifier les phénomènes de *temps et aspects* n'ont pas nécessairement d'implication sur l'interprétation des énoncés.

Ainsi, pour l'analyse de l'exemple 2, on construit le DRS-graphe  $[e|avalier(e, x, y)]$ . L'ajout de la négation implique la construction du DRS-graphe  $[e|avalier(e, x, y) \wedge \neg(e)]$ . La suite de l'analyse permet alors de ne construire que les structures sous-spécifiées où les quantifications ont portée sur la variable d'événement. Le choix fait de positionner en bas de la structure hiérarchique les événements niés et le soucis d'homogénéité du calcul, explique la position exposée précédemment sur les formes normales des énoncés affirmatifs.

## 4 Conclusion et perspectives

Les raisons qui nous ont conduit à réaliser les choix initiaux sont principalement dues à des problèmes de robustesse. En effet, ajouter le traitement des quantificateurs entraîne une nécessaire sur-génération (ou ambiguïté). Bien que nous soyons en mesure de faire un choix parmi les résultats obtenus, en prenant par exemple la représentation qui possède la plus longue chaîne connexe (ces graphes sont acycliques par définition), nous souhaitons à présent intégrer des représentations plus élaborées permettant une représentation plus fine. Ceci est réalisé via des notions issues de la sous-spécification. Une raison pour laquelle cette nécessité ne s'est pas fait sentir de manière importante pour le moment est que les domaines sur lesquels les outils ont été développés étaient délibérément limités.

Les autres points importants dont il nous faudrait rendre compte sont les problèmes liés à la co-

ordination et à l'implication (cause et effet). Si cette approche permet d'augmenter la pertinence de la modélisation, elle se pose comme une étape vers la définition d'un système complet.

Une fois ces propriétés formelles introduites dans le système de calcul, nous avons envisagé de tester sa robustesse. Pour cela, nous avons mis en place un protocole dont la première étape a été la construction d'un corpus sémantique. Ce dernier comporte une soixantaine de phrases réparties en dix domaines relatifs à différentes questions de sémantique. La seconde phase, celle d'implémentation de ces propositions et d'intégration dans l'outil TiLT, est en cours. Elle permettra de finaliser les tests prévus.

## Références

- AMBLARD M. (2007). *Calculs de représentations sémantique et syntaxe générative: les grammaires minimalistes catégorielles*. PhD thesis, LaBRI, Université Bordeaux 1.
- AMSILI P. & ROUSSARIE L. (2004). Vers une lambda-drt étendue. In P. BLACHE, Ed., *Actes de TALN 2004 (Traitement automatique des langues naturelles)*, Fès, Maroc: ATALA LPL.
- ASHER N. & LASCARIDES A. (2003). *Logics of Conversation*. Cambridge University Press.
- BLACKBURN P., BOS J., KOHLHASE M. & DE NEVILLE H. (1999). Inference and computational semantics. In H. BUNT & E. THIJSSE, Eds., *Third International Workshop on Computational Semantics (IWCS-3)*, p. 5–21.
- BUSQUETS J., VIEU L. & ASHER N. (2001). La SDRT : une approche de la cohérence du discours dans la tradition de la sémantique dynamique. *Verbum*, **XXIII**(1), 73–101.
- DANLOS L. (2007). Un formalisme pour le discours inspiré des TAG synchrones. In F. BENARMARA, N. HATOUT, P. MULLER & S. OZDOWSKA, Eds., *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, p. 389–398, Toulouse: ATALA IRIT.
- DAVIDSON D. (1967). The logical form of action sentences. In N. RESHER, Ed., *The Logic of Decision and Action*, p. 81–95: Pittsburgh University Press.
- HEINECKE J. (2006). Génération automatique des représentations ontologiques. In P. MERTENS, C. FAIRON, A. DISTER & P. WATRIN, Eds., *Actes de la 13e conférence TALN*, vol. 2, p. 502–511. Louvain: Presses universitaires de Louvain.
- IHEDDADENE M. (2007). *Traduction Automatique : étude et réalisation d'un module de génération à partir d'une représentation sémantique interlingue*. PhD thesis, Université de Provence, France.
- KAMP H. & REYLE U. (1993). *From Discourse to Logic*. Dordrecht: Kluwer Publishers.
- KERVAJAN L., GUIMIER DE NEEF E. & BRETON G. (2007). Vers un système de traduction automatique français/langue des signes française. *Traitement Automatique des Langues*, **48**(2).
- PARK J., MAILLEBAU E., GUIMIER DE NEEF E., VINESSE J. & HEINECKE J. (2007). Evaluating an Interlingual Semantic Representation. In K. GERDES, T. REUTHER & L. WANNER, Eds., *Meaning - Text Theory 2007*, München - Wien.
- RETORÉ C. (2008). Les mathématiques de la linguistique computationnelle. *Gazette des Mathématiciens*, **115**, 35–62.
- SCHILDER F. (1998). An underspecified segmented discourse representation theory (USDRT). In *COLING-ACL*, p. 1188–1192.
- SOWA J. (1964). Conceptual graphs for a data base interface. *IBM J. RES. DEVELOP.*, p. 336–357.