



**HAL**  
open science

## Degré de relation sémantique dans une ontologie pour la commande en langue naturelle

Laurent Mazuel, Nicolas Sabouret

### ► To cite this version:

Laurent Mazuel, Nicolas Sabouret. Degré de relation sémantique dans une ontologie pour la commande en langue naturelle. Ingénierie des Connaissances 2007 - IC2007, Jul 2007, Grenoble, France. pp.73-83. hal-00379231

**HAL Id: hal-00379231**

**<https://hal.science/hal-00379231>**

Submitted on 28 Apr 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Degré de relation sémantique dans une ontologie pour la commande en langue naturelle

Laurent Mazuel, Nicolas Sabouret

Laboratoire d'Informatique de Paris 6 (LIP6)

104 av du Président Kennedy

75016 Paris – FRANCE

laurent.mazuel@lip6.fr

<http://www-poleia.lip6.fr/~mazuel>

**Résumé** : Dans la communication homme-machine en langue naturelle, le système doit être capable d'associer les termes de l'utilisateur avec les concepts manipulés par l'agent artificiel. Dans cet article, nous proposons une mesure de degré de relation sémantique entre deux concepts d'une ontologie pour faciliter cette phase dite « d'interprétation sémantique ». Nous proposons une généralisation de la distance de Jiang & Conrath en prenant appui sur d'autres travaux autour de l'utilisation des relations non-hiérarchiques dans les ontologies. Nous expliquons comment nous l'avons intégré dans notre plateforme d'agent conversationnels, en utilisant une architecture basée sur une interprétation sémantique à trois niveaux.

**Mots-clés** : Degré de relation sémantique. Ontologie. Système d'interaction homme-machine. Interprétation sémantique.

## 1 Introduction

La grande majorité des systèmes d'interaction homme-machine actuels en langue naturelle pour la commande, les systèmes de question/réponses, le dialogue... utilise une ontologie pour l'interprétation sémantique (Corby *et al.*, 2004; Dzikovska *et al.*, 2003; Milward & Beveridge, 2003). Cette ontologie comprend, par hypothèse de connectivité sémantique, au moins un concept pour chaque mot pouvant apparaître dans les phrases de l'utilisateur (Sadek *et al.*, 1997). Elle définit aussi les concepts manipulés par le système en lui-même et, éventuellement, des relations entre ces concepts. Par exemple, dans (Corby *et al.*, 2004), chaque document est indexé par un ensemble de concepts. L'interprétation sémantique consiste alors à associer les termes de l'utilisateur avec les concepts manipulés par l'agent artificiel en utilisant cette ontologie. Cette approche permet de concevoir des systèmes robustes et capables de comprendre des commandes ou des requêtes imprécises ou au contraire sur-spécifiées et de proposer des alternatives à l'utilisateur. Par exemple (tiré de Corese (Corby *et al.*, 2004)) :

**User :** *Je veux une liste de magasin vendant des cartes Ethernet dans ma ville.*

**System :** *Je n'ai pas trouvé de magasins avec cartes Ethernet, voulez-vous les magasins avec cartes WiFi ?*

Pour obtenir ce résultat, différentes méthodes et algorithmes d'interprétation sémantique basés sur les ontologies et des *mesures d'approximation sémantique* (ou simplement *mesures sémantiques*) existent et sont utilisées. Dans le cas de Corese, par exemple, l'approximation est un score de similarité sémantique (celui de (Zhong *et al.*, 2002)) permettant de déterminer quels concepts équivalents aux mots de la requête sont considérés acceptables par le système pour générer une réponse pertinente.

Selon la définition de (Resnik, 1995), une mesure sémantique peut être définie en utilisant uniquement la relation de subsomption (on parle alors de *mesure de similarité sémantique*), ou bien en utilisant tous les types de relations comme la méronymie (*part-of*), l'antonymie, ou tout autre relation (on parle alors de *mesure de degré de relation sémantique*), ce qui permet de considérer les critères fonctionnels décrits par les relations (par exemple, les concepts de « voiture » et « essence » ont une faible similarité mais un fort degré de relation). Néanmoins, le calcul automatique d'un degré de relation sémantique est considéré comme beaucoup plus difficile à définir et c'est pourquoi la majeure partie des travaux dans le domaine se sont focalisés sur les mesures de similarité, souvent basé sur des hiérarchies connues (*e.g.* WordNet (Budanitsky & Hirst, 2006) ou Wikipedia (Gabrilovich & Markovitch, 2007)). En particulier, dans les systèmes d'interaction homme-machine en langue naturelle actuels, les ontologies sont en fait réduites à des hiérarchies de concepts et l'interprétation sémantique s'appuie uniquement sur une mesure de similarité.

Dans cet article, nous proposons au contraire d'utiliser une mesure de degré de relation sémantique au sein d'un système de commande en langue naturelle d'agents conversationnels déployés sur le web, décrit dans (Mazuel & Sabouret, 2006). Nos agents sont munis d'une ontologie de concepts reliés par un ensemble de relations non contraint (hyponymie, méronymie ou tout autre relation plus spécifique au domaine de l'agent en particulier, comme *estPlusGrandQue*). Dans notre système, une commande de l'utilisateur peut être vue comme un ensemble structuré de concepts issus de l'ontologie. Réciproquement, chaque agent connaît l'ensemble des commandes formelles (c'est-à-dire décrites dans le notre langage agent) qu'il est capable de traiter. Comme pour les commandes de l'utilisateur, chaque commande formelle peut être représentée par un ensemble structuré de concepts issus de l'ontologie. Notre analyse sémantique consiste donc à mesurer, à partir de l'ontologie, quelle est la commande formelle représentant le mieux la commande de l'utilisateur et, de plus, d'estimer le degré de pertinence de ce choix.

La prochaine section présente les travaux sur lesquels nous nous sommes appuyés pour la définition de notre mesure de degré de relation sémantique. La section 3 présente notre architecture pour l'interprétation sémantique et notre proposition de mesure de degré de relation sémantique sur les ontologies. La section 4 conclue l'article.

## 2 Travaux existants

Nous détaillons dans cette section les travaux servant de base à notre mesure. Nous présenterons d'abord la distance de (Jiang & Conrath, 1997), définie initialement pour la relation de subsomption, et nous montrons comment elle peut-être étendue aux autres relations. Nous présenterons ensuite les travaux sur les appariements d'ontologie de (Aleksovski *et al.*, 2006) et (Hollink, 2006) dont nous nous sommes inspirés pour définir notre stratégie de choix d'un chemin entre deux concepts<sup>1</sup>.

### 2.1 Mesure de distance de Jiang & Conrath

(Jiang & Conrath, 1997) définissent leur mesure de distance comme « basée sur les longueurs de chemins » (d'après la mesure de (Rada *et al.*, 1989)) « avec la théorie de l'information comme facteur de décision ». À chaque noeud est associée une quantité d'information (Resnik, 1995) qui est utilisée pour valuer les chemins. La distance est alors dépendante non seulement du chemin parcouru, mais aussi de la quantité d'information véhiculée par les noeuds traversés par ce chemin.

La quantité d'information  $IC(c)$  pour un concept  $c$  est calculée en utilisant la formule de (Seco *et al.*, 2004) qui se base uniquement sur la structure de la hiérarchie :

$$IC(c) = 1 - \frac{\log(\text{hypo}(c) + 1)}{\log \text{max}_{WN}}$$

où  $\text{hypo}(c) \in \mathbb{N}$  représente le nombre de sous-classes de  $c$ ,  $\text{max}_{WN}$  le nombre de concept total dans la hiérarchie (en l'occurrence WordNet (Fellbaum, 1998)).

Le calcul de la longueur du chemin entre  $c_1$  et  $c_2$  s'appuie sur la recherche du plus petit subsumant (*closest common parent*) de  $c_1$  et  $c_2$ , noté  $\text{cpp}(c_1, c_2)$ . Le plus court chemin  $\text{sp}(c_1, c_2)$  entre  $c_1$  et  $c_2$  est alors l'unique chemin passant par  $\text{cpp}(c_1, c_2)$ <sup>2</sup>.

Enfin, pour chaque arête de l'arbre  $a$  reliant deux concepts  $x$  et  $y$ , Jiang & Conrath proposent de définir un poids  $TC(x, y) \in [0, 1]$ . La mesure de distance  $\text{dist}_{JC}$  de Jiang & Conrath est alors définie par :

$$\text{dist}_{JC}(c_1, c_2) = \sum_{c \in \text{sp}(c_1, c_2) \setminus \text{cpp}(c_1, c_2)} [IC(c) - IC(\text{parent}(c))] \times TC(c, \text{parent}(c))$$

Cette mesure de distance est probablement la plus utilisée actuellement et la plus efficace pour déterminer la proximité sémantique entre deux concepts Budanitsky & Hirst (2006). Comme toute mesure de similarité, elle se limite à la relation de subsomption, mais elle offre la possibilité de prendre en compte des valuations différentes pour chaque arête de l'arbre des concepts (les  $TC(c, \text{parent}(c))$ ). Cette fonctionnalité n'as jamais été prise en compte. Dans l'article initial (Jiang & Conrath, 1997), l'évaluation fut en effet effectué pour un poids d'arête constant ( $TC = 1$ ). Depuis, les systèmes

<sup>1</sup>En effet, si on considère l'ontologie comme un graphe, la prise en compte de toutes les relations augmentent considérablement le nombre de chemin possibles.

<sup>2</sup>Dans le cas d'une hiérarchie simple de concepts, il est unique. Dans le cas de multi-héritage, comme dans WordNet, le consensus actuel est de choisir le  $\text{cpp}$  de profondeur maximale, afin de maximiser la quantité d'information (Budanitsky & Hirst, 2006).

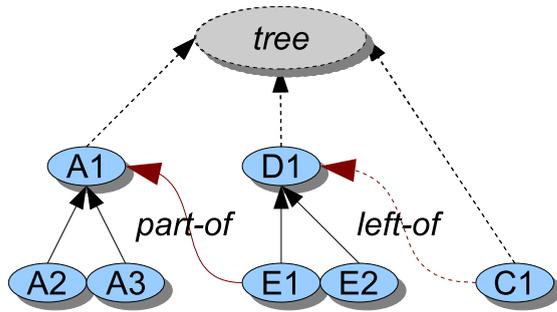


FIG. 1 – Exemple de problème de hiérarchie. Les flèches noires sont des relations de subsumption (*isa*). Les autres flèches correspondent à différents types de relation sémantique.

utilisant la formule de Jiang & Conrath se servent de la version simplifiée sans poids sur les arêtes<sup>3</sup>.

Au contraire, nous proposons de :

1. généraliser la mesure proposée par Jiang & Conrath aux graphes de relations ;
2. attribuer des poids différents en fonction des types de relations traversés.

## 2.2 Recherche de chemins acceptables

La généralisation de la mesure de Jiang & Conrath aux graphes pose un problème d'unicité de chemin. En effet, alors que le plus court chemin est unique dans le cas d'une hiérarchie, plusieurs chemins candidats sont envisageables dans le cas des graphes, mais les travaux récents sur l'alignement d'ontologie (Hollink, 2006) et (Aleksovski *et al.*, 2006) nous montrent que tous ne sont pas acceptables du point de vue de la sémantique.

Par exemple, sur la figure 1, le plus court chemin entre *C1* et *A2* passe par *D1*, mais rien ne dit que ce chemin soit sémantiquement cohérent.

Pour l'alignement de deux ontologies  $\mathcal{O}_1$  et  $\mathcal{O}_2$ , (Aleksovski *et al.*, 2006) utilisent une méthode à base d'ontologie de référence  $\mathcal{O}_r$ . L'algorithme général se découpe en deux étapes :

1. L'ancrage des concepts de  $\mathcal{O}_1$  et  $\mathcal{O}_2$  dans  $\mathcal{O}_r$  : l'objectif est de trouver, pour chaque concept  $c$  de  $\mathcal{O}_1$  (resp.  $\mathcal{O}_2$ ) un concept  $c' \in \mathcal{O}_r$  qui lui correspond.
2. L'alignement des concepts ancrés : l'objectif est de constituer des couples  $(c'_1, c'_2)$  de concepts reliés par un chemin « acceptable » dans  $\mathcal{O}_r$ . C'est ce deuxième aspect qui nous intéresse dans nos travaux.

Prenons par exemple deux concepts  $c_1 \in \mathcal{O}_1$  et  $c_2 \in \mathcal{O}_2$ . Alors il existe au moins un chemin entre  $c'_1$  et  $c'_2$  dans  $\mathcal{O}_r$ . La question est alors de savoir quels critères au niveau de la structure de ce chemin permet de considérer que  $c_1$  est une bonne approximation de  $c_2$ , sachant que ce chemin emprunte un nombre quelconque de relations différentes.

<sup>3</sup> $dist_{JCsimple}(c_1, c_2) = (IC(c_1) + IC(c_2)) - 2 \times IC(cpp(c_1, c_2))$

Aleksovski propose un ensemble de règles (*i.e.* ou contraintes) à respecter pour qu'un chemin soit jugé « acceptable sémantiquement ». Une contrainte simple serait de n'autoriser que les relations hiérarchiques (on retombe alors sur une mesure de similarité sémantique !). L'évaluation de ces contraintes montre que l'utilisation mixte des relations d'hyponymie (*isa*) et de méronymie (*partof*) produit des appariements incorrects. Par exemple, cette utilisation mixte permettait d'apparier le concept *Dental pulp* avec *Bronchus*<sup>4</sup> dans l'ontologie FMA<sup>5</sup>. Ils proposent alors comme « meilleure » contrainte d'utiliser un nombre indéterminé de relation *partof*, sans *isa*, puis un nombre indéterminé de *isa* (sans *partof*).

Pour généraliser ces travaux au cas des graphes quelconques, en s'appuyant sur les résultats de (Aleksovski *et al.*, 2006), nous proposons la contrainte suivante :

1. Utilisation d'un seul type de relation dans un chemin en plus des relations *isa* ;
2. Utilisation de l'hyponymie (*isa*) *uniquement après* l'autre relation choisie.

Pour reprendre l'exemple précédent sur la figure 1, le chemin entre le concept *C1* et le concept *E1* qui passe par la relation *leftof* est donc incorrect pour notre modèle.

### 3 Analyse sémantique

Cette section présente le module d'analyse sémantique de notre architecture de compréhension des commandes en langue naturelle. Nous commençons par donner un rapide aperçu de notre système d'agents conversationnels, puis nous décrivons ensuite notre mesure de degré de relation sémantique. Enfin, nous montrons comment elle est utilisée au sein de nos algorithmes d'interprétation sémantique.

#### 3.1 Aperçu de notre architecture

Notre système de compréhension des commandes en langue naturel repose sur une approche ascendante (*i.e.* bottom-up) comme il est possible d'en voir dans (Paraiso & Barthès, 2004). L'objectif est d'apparier la commande de l'utilisateur avec les actions possibles de l'agent, définies dans une *liste de compétences*. Contrairement à l'approche classique, l'approche *ascendante générative* que nous utilisons construit dynamiquement cette liste de compétence à partir de l'analyse du code de l'agent à l'exécution. Dans notre modèle d'agent (Sabouret & Sansonnet, 2001), nous appelons *événements* les commandes formelles de l'agent.

Le principe général de notre interprétation sémantique est d'apparier les événements de l'agent avec la commande en langue naturelle de l'utilisateur. Pour ce faire, nous associons à chaque événement un score d'appariement qui évalue la proximité de l'événement avec la commande de l'utilisateur, en s'appuyant sur la mesure de degré de relation sémantique. Ce score permet à un module de gestion de dialogue utilisateur de déterminer la stratégie de dialogue la plus appropriée. Nous utilisons pour cela un

<sup>4</sup>via le chemin *Dental pulp isa Loose connective tissue partof Right main bronchus isa Bronchus*.

<sup>5</sup><http://sig.biostr.washington.edu/projects/fm/AboutFM.html>

système de seuil inspiré de celui proposé dans (Maes, 1994) qui permet de faire la différence entre les commandes parfaitement comprises, les commandes incertaines et les commandes non-comprises.

L'architecture est définie plus en détails dans (Mazuel & Sabouret, 2006).

### 3.2 Une mesure de degré de relation sur les ontologies

Nous définissons dans cette section notre mesure de degré de relation sémantique dans une ontologie. Nous utilisons pour cela la formule de Jiang & Conrath avec la fonction de pondération des arêtes (section 2.1) ainsi que les travaux pour la recherche de chemin sémantiquement correct utilisés dans l'alignement d'ontologie (section 2.2). En généralisant la mesure de *distance* de Jiang & Conrath, nous obtenons une mesure de *non-relation* sémantique (*i.e.* plus le score de notre mesure est faible, plus les concepts sont reliés), que nous notons  $dist_{ONT}$ . Cette mesure étant bornée, nous pouvons alors la convertir en une mesure de degré de relation  $sim_{ONT}$  en utilisant les méthodes proposées par (Resnik, 1995; Jiang & Conrath, 1997).

Notre mesure de « non-relation » sémantique  $dist_{ONT}$  est séparée en deux composantes distinctes. La composante hiérarchique correspond la mesure de distance de Jiang & Conrath simple. La composante relationnelle (c'est-à-dire l'utilisation d'une unique autre relation avant la composante hiérarchique, comme expliqué en section 2.2) repose sur une pondération des arêtes et sur la longueur à parcourir sur cette relation, en généralisant la formule de Jiang & Conrath. La difficulté est alors que la notion de quantité d'information d'un concept  $IC(c)$  n'est plus exploitable directement.

Par définition de la généralisation, notre mesure doit être en conformité avec la mesure initiale de Jiang & Conrath dans le cas des ontologies ne comprenant qu'une relation hiérarchique. La composante hiérarchique de notre mesure est donc  $dist_h(c_1, c_2) = dist_{JCsimple}(c_1, c_2)$  et la composante relationnelle doit être nulle lorsqu'aucune relation n'est utilisée.

Le calcul de la composante relationnelle pour une relation donnée  $X$  est calculé directement à partir du plus court chemin suivant  $X$  dans le sous-graphe correspondant à l'ontologie limitée à la relation  $X$ . Pour  $c_1$  et  $c_2$  dans l'ontologie, nous noterons  $sp_X(c_1, c_2)$  le plus court chemin suivant  $X$  entre  $c_1$  et  $c_2$  et  $|sp_X(c_1, c_2)|$  la longueur de ce chemin ( $\infty$  lorsque  $c_1$  et  $c_2$  sont dans deux sous-graphes disjoints). Afin de retrouver la croissance logarithmique de la quantité d'information  $IC$  et afin de rester dans un intervalle borné (pour une ontologie donnée), la composante relationnelle est définie par :

$$dist_{rel}(c_1, c_2) = \frac{(n-1)}{n} \times TC_X$$

où  $n = |sp_X(c_1, c_2)|$  est le nombre d'arêtes du plus court chemin entre  $c_1$  et  $c_2$  suivant  $X$  et  $TC_X$  est la pondération associée au type d'arête  $X$ .<sup>6</sup>

Pour déterminer la mesure de distance généralisée, nous devons alors considérer toutes les relations  $X$  possibles pour la composante relationnelle. Nous choisissons

<sup>6</sup>Dans notre modèle, les pondérations des arêtes dépendent uniquement du type de relation de l'arête – et non de l'arête elle-même, comme envisagé initialement par Jiang & Conrath (1997).

alors celle qui donne le plus faible « degré de non-relation ». Nous avons alors la mesure de *non-relation sémantique* suivante :

$$dist_{ONT}(c_1, c_2) = \min_{t \in \mathcal{C}, X \in \mathcal{R}} \left\{ TC_X \times \left( \frac{|sp_X(c_1, t)| - 1}{|sp_X(c_1, t)|} \right) + dist_{JCsimple}(t, c_2) \right\}$$

où  $\mathcal{C}$  est l'ensemble des concepts et  $\mathcal{R}$  l'ensemble des relations dans l'ontologie. Autrement dit, notre mesure recherche pour chaque relation  $X$  le plus court chemin entre  $c_1$  et tous les concepts  $t$  atteignables suivant  $X$  depuis  $c_1$  et pouvant eux-même atteindre  $c_2$  uniquement par la relation hiérarchique.<sup>7</sup>

Notre mesure  $dist_{ONT}$  étant bornée sur  $[0, 2 + \max_{X \in \mathcal{R}} TC_X]$ , nous pouvons définir notre mesure de degré de relation sémantique :

$$sim_{ONT}(c_1, c_2) = 1 - \frac{dist_{ONT}(c_1, c_2)}{2 + \max_{X \in \mathcal{R}} TC_X}$$

### 3.3 Interprétation sémantique

Nous utilisons notre mesure de degré de relation pour calculer un score d'appariement entre un évènement formel de l'agent (candidat au titre de la meilleure interprétation) et la commande rentrée par l'utilisateur. Nous pouvons ainsi choisir, parmi l'ensemble des évènements formels construit celui qui possède le meilleur score d'interprétation.

Notre calcul du score d'appariement repose sur un découpage en trois composantes d'interprétation, selon une idée proche de celle de (Ehrig *et al.*, 2004)<sup>8</sup>. Nous allons décrire ces trois niveaux de compréhension dans les trois sous-sections suivantes.

#### 3.3.1 Le niveau contextuel

Ce niveau mesure la proximité sémantique des termes dans l'environnement de l'agent (*i.e.* dans l'environnement réel de l'application). Nous utilisons, pour récupérer un modèle de l'environnement à l'état courant, les capacités d'introspection de nos agents propres au modèle d'agent que nous utilisons (Sabouret & Sansonnet, 2001). Nous modélisons ce code sous la forme de graphe de dépendance non-orienté. L'algorithme consiste alors à rechercher parmi les stables maximaux<sup>9</sup> de ce graphe celles qui possède le minimum de sommets. Le nombre de sommets permet de discriminer si un ensemble de termes donné du code contient des éléments très proches contextuellement ou non. Par exemple, si l'objet *cube* est positionné actuellement dans la ligne *upper* et la colonne *left*, notre système pourra déduire par cet algorithme que les termes  $\{upper, left, cube\}$  sont plus proches du contexte actuel que les termes  $\{lower, right, cube\}$ .

<sup>7</sup>Pour que la composante relationnelle soit effectivement nulle lorsqu'aucune relation ne peut être utilisée, nous ajoutons dans l'ontologie une relation *vide* reliant chaque noeud à lui-même et telle que  $TC_{vide} = 0$ .

<sup>8</sup>A la différence de Ehrig, nous n'avons pas de fonction *amalgame* pour pondérer chaque niveau. Chaque niveau est dépendant du niveau précédent.

<sup>9</sup>Un stable maximal est un sous-graphe sans arête tel que tout ajout de sommet à ce sous-graphe ajoute une arête du graphe initial.

Cette mesure conceptuelle, fortement liée au langage d'agent que nous utilisons, est utilisée uniquement pour comparer différents ensemble de terme de l'environnement entre eux et non pas pour les évaluer individuellement par rapport à la commande utilisateur.

### 3.3.2 Le niveau ontologique

Le niveau ontologique utilise la mesure  $sim_{ONT}$  définie précédemment. Ce niveau correspond au calcul des degrés de relation entre les termes fournis par le niveau des évènements. Dans le cas de commandes contenant des références (*p. ex.* « à gauche de la boulangerie », etc.), nous utilisons l'algorithme du niveau contextuel pour décider comment interpréter cette référence. Dans l'exemple précédent, nous pouvons ainsi traduire « à gauche de la boulangerie » par une expression hors-contexte du type « le 32 avenue Guy Moquet ».

### 3.3.3 Le niveau des évènements

Le dernier niveau calcule le score sémantique entre un évènement formel généré (noté  $evt$ ) et la modélisation de la commande utilisateur (noté  $user$ ). Nous ne donnerons ici qu'une version simplifiée de notre formule, sans prendre en compte la structure des évènements et des commandes utilisateur pour plus de clarté. Dans ce cas, le calcul du score correspond à calculer le degré de relation sémantique global entre  $evt$  et  $user$ . Pour cela, nous faisons la moyenne des scores de degré de relation entre chaque terme en tenant compte de toutes les permutations possibles. Nous pouvons synthétiser ce score de la façon suivante :

$$score_{EVENT}(evt, user) = \frac{\sum_{d \in user} \left( \max_{e \in flat(evt)} sim_{ONT}(d, e) \right)}{|user|}$$

où la fonction  $flat$  est une méthode d'énumération de l'évènement compatible avec l'énumération des concepts  $d$  de la modélisation.

## 4 Conclusion

Dans cet article, nous avons montré comment l'analyse sémantique dans un système d'interprétation de commande pouvait tirer parti d'une mesure de degré de relation sémantique. notre interprétation sémantique est basé sur un modèle en trois couches sémantique lié au contexte courant, à l'ontologie du domaine et des donnée d'interaction. Cette mesure est construite autour de travaux sur les mesures de similarité sémantique et l'alignement d'ontologie.

L'évaluation en cours devrait nous permettre d'affiner les pondérations en hiérarchie et autres relations. Cette mesure étant faite pour les systèmes possédant des ontologies spécifiques du domaine, il ne nous semble pas possible d'en faire une évaluation générale. C'est pourquoi nous cherchons actuellement à implémenter d'autres agents pour

montrer sa généricité entre différents domaines en plus de son efficacité à traiter correctement le problème. De même, nous avons envisagé de considérer d'autres critères pour cette pondération, comme la densité de relation pour un concept donné (*i.e.* plus un concept  $c$  est relié à d'autres concepts par une relation de type  $X$ , moins le type  $X$  de relation est informatif sur la description de  $c$ ).

De plus, nous ne pouvons pas savoir à l'avance si un type de relation représente, pour le développeur de l'ontologie, une relation de rapprochement sémantique. Dans notre modèle actuel, un utilisateur doit donc confirmer au système la validité des déductions sémantiques faite grâce à notre mesure de degré de relation, lorsqu'elle utilise des relations non-hiérarchiques (*p. ex.* « Par "cube", voulez-vous désigner le terme "carré" ? »). Nous pensons que les confirmations de l'utilisateur peuvent servir de base pour un algorithme d'apprentissage par renforcement dont le but serait de comprendre quels relations amènent à des résultats sémantiques cohérents si elles sont utilisées dans la mesure. Cet algorithme aurait ainsi en charge d'apprendre les pondérations optimales de la fonction  $TC_X$ . Cet algorithme est actuellement à l'étude.

## Références

- ALEKSOVSKI Z., TEN KATE W. & VAN HARMELEN F. (2006). Exploiting the structure of background knowledge used in ontology matching. In *Proc. Workshop on Ontology Matching in ISWC2006 : CEUR Workshop Proceedings*.
- BUDANITSKY A. & HIRST G. (2006). Evaluating wordnet-based measures of semantic distance. *Computational Linguistics*, **32**(1), 13–47.
- CORBY O., DIENG-KUNTZ R. & FARON-ZUCKER C. (2004). Querying the Semantic Web with the CORESE search engine. In I. PRESS, Ed., *Proc. of the ECAI'2004*, p. 705–709, Valencia.
- DZIKOVSKA M. O., ALLEN J. F. & SWIFT M. D. (2003). Integrating linguistic and domain knowledge for spoken dialogue systems in multiple domains. In *Proc. of IJCAI-03 Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.
- EHRIG M., HAASE P., STOJANOVIC N. & HEFKE M. (2004). Similarity for ontologies—a comprehensive framework. *Workshop Enterprise Modelling and Ontology : Ingredients for Interoperability, (PAKM 2004)*.
- C. FELLBAUM, Ed. (1998). *WordNet, An Electronic Lexical Database*. MIT-Press.
- GABRILOVICH E. & MARKOVITCH S. (2007). Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. In *Proc. IJCAI07*, p. 1606–1611.
- HOLLINK L. (2006). *Semantic Annotation for Retrieval of Visual Resources*. PhD thesis, Vrije Universiteit Amsterdam.
- JIANG J. & CONRATH D. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. on International Conference on Research in Computational Linguistics*, p. 19–33, Taiwan.
- MAES P. (1994). Agents that reduce workload and information overload. *Communications of the ACM*, **37**(7), 30–40.
- MAZUEL L. & SABOURET N. (2006). Generic command interpretation algorithms for conversational agents. In *Proc. Intelligent Agent Technology (IAT'06)*, p. 146–153 : IEEE Computer Society.

- MILWARD D. & BEVERIDGE M. (2003). Ontology-based dialogue systems. In *Proc. 3rd Workshop on Knowledge and reasoning in practical dialogue systems (IJCAI03)*, p. 9–18.
- PARAISO E. & BARTHÈS J. (2004). Architecture d'une interface conversationnelle pour les agents assistants personnels. In P. PAROUBECK & J.-P. SANSONNET, Eds., *Actes de la Journée d'Etude ATALA Agental « Agents et Langue »*, p. 83–90, Paris, France : ATALA ATALA.
- RADA R., MILI H., BICKNELL E. & BLETTNER M. (1989). Development and Application of a Metric on Semantic Nets. *IEEE Transactions on Systems, Man, and Cybernetics*, **19**(1), 17–30.
- RESNIK P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *IJCAI*, p. 448–453.
- SABOURET N. & SANSONNET J. (2001). Automated Answers to Questions about a Running Process. In *Proc. CommonSense 2001*, p. 217–227.
- SADEK D., BRETIER P. & PANAGET E. (1997). Artimis : Natural dialogue meets rational agency. In *IJCAI (2)*, p. 1030–1035.
- SECO N., VEALE T. & HAYES J. (2004). An Intrinsic Information Content Metric for Semantic Similarity in WordNet. In *Proc. ECAI'2004, the 16th European Conference on Artificial Intelligence*, p. 1089–1090.
- ZHONG J., ZHU H., LI J. & YU Y. (2002). Conceptual graph matching for semantic search. In *ICCS '02 : Proceedings of the 10th International Conference on Conceptual Structures*, p. 92–196, London, UK : Springer-Verlag.