

## Un Analyseur Sémantique pour le DHM Modélisation – Réalisation – Évaluation

Jérôme Lehuen (1), Thierry Lemeunier (2)

LIUM - Université du Maine

Avenue Laënnec, 72085 Le Mans Cedex 9

(1) Jerome.Lehuen@lium.univ-lemans.fr

(2) Thierry.Lemeunier@lium.univ-lemans.fr

**Résumé** Cet article décrit un modèle de langage dédié au dialogue homme-machine, son implémentation en CLIPS, ainsi qu'une évaluation comparative. Notre problématique n'est ni d'analyser des grands corpus, ni de proposer une grammaire à grande couverture. Notre objectif est de produire des représentations sémantiques utilisables par un module de dialogue à partir d'énoncés oraux courts, le plus souvent agrammaticaux. Une démarche pragmatique nous a conduit à fonder l'analyse sur des principes simples mais efficaces dans le cadre que nous nous sommes fixé. L'algorithme retenu s'inspire de l'analyse tabulaire. L'évaluation que nous présentons repose sur le corpus MEDIA qui a fait l'objet d'une annotation sémantique manuelle de référence pour une campagne d'évaluation d'analyseurs sémantiques pour le dialogue. Les résultats que nous obtenons place notre analyseur dans le trio de tête des systèmes évalués lors de la campagne de juin 2005, et nous confortent dans nos choix d'algorithme et de représentation des connaissances.

**Abstract** This article describes a natural language model dedicated to man-machine dialogue, its implementation in CLIPS, as well as a comparative evaluation. Our problematic is neither to analyze large corpora nor to propose a large-coverage grammar. Our objective is to produce semantic representations usable for a dialog module from short oral utterances that are rather often ungrammatical. A pragmatic approach leads us to base parsing on simple but efficient principles within the man-machine dialog framework. Chart parsing influences the algorithm we have chosen. The evaluation that we present here uses the MEDIA corpora. It has been manually annotated and represents a standard usable in an evaluation campaign for semantic parsers dedicated to the dialog. With the results that we obtain our parser is in the three bests of the systems evaluated in the June 2005 campaign. It confirms our choices of algorithm and of knowledge representation.

**Mots-clés :** Analyse sémantique tabulaire, contexte dialogique, évaluation.

**Keywords:** Semantic chart parsing, dialogue context, evaluation.

## 1 Introduction

Cet article décrit un analyseur sémantique destiné à être intégré dans un système de dialogue homme-machine avec entrée vocale. La langue orale spontanée présente un certain nombre de spécificités que l'on ne trouve pas dans les textes écrits (Blanche-Benveniste et al., 1990). Elle se distingue d'abord par sa linéarité : toute correction ou reprise perturbe l'énonciation. Elle admet une plus grande souplesse syntaxique : elle comporte ellipses, structures clivées et autres disfluences qui perturbent les outils d'analyse sensibles à l'agrammaticalité. Une démarche pragmatique nous a conduit à fonder l'analyse sur des principes simples mais efficaces, capables d'appréhender ces types d'énoncés oraux. Afin de ne pas faire reposer l'interprétation sémantique sur le succès d'une analyse syntaxique, nous avons cherché à intégrer des aspects syntaxiques et sémantiques au sein d'un même modèle, ce qui permet de faire collaborer harmonieusement ces deux dimensions. De plus, notre modèle permet des analyses partielles ainsi que la génération de certaines hypothèses de rattachement entre constituants.

Nous justifions aussi nos choix par rapport à la suite du projet, à savoir l'intégration avec un module de reconnaissance de la parole. L'entrée de notre analyseur pourrait être, non plus une chaîne de caractères, mais un treillis de mots ou un réseau de confusion, qui présente un ensemble de possibilités. La problématique sera alors d'utiliser les contraintes sémantiques afin de départager deux mots de scores équivalents.

Le contexte de notre recherche et les contraintes que nous avons évoqué dans cette introduction nous ont conduit à développer d'une part une grammaire syntaxico-sémantique et d'autre part un processus de type analyse tabulaire. Dans la suite de l'article nous présentons tout d'abord notre modèle des connaissances syntaxico-sémantiques. Dans la deuxième partie nous présentons une implémentation de ce modèle. Dans une troisième partie, nous présentons les résultats obtenus suite à une évaluation fondée sur le corpus MEDIA (Maynard et al., 2005).

## 2 Modélisation

Les connaissances se présentent sous la forme d'un ensemble d'unités lexicales définies à la fois syntaxiquement et sémantiquement. Chaque lexie est caractérisée par un nom, par un système d'offres et d'attentes, et par au moins une forme d'usage langagière (Figure 1). Cela forme une association signifié/signifiant. Les offres et les attentes sont des traits sémantiques. Ils permettent de lier sémantiquement les lexies entre elles. Les formes d'usage sont des structures syntaxiques composées de mots (symboles terminaux) et/ou de relations de dépendance faisant référence aux attentes de la lexie (symboles non-terminaux), comme par exemple « *A1 chambre A2* » où A1 et A2 font référence à des attentes de la lexie [chambre]. On obtient ainsi une sorte de grammaire non contextuelle utilisable pour l'analyse et pour la génération (Lehuen, 2008).

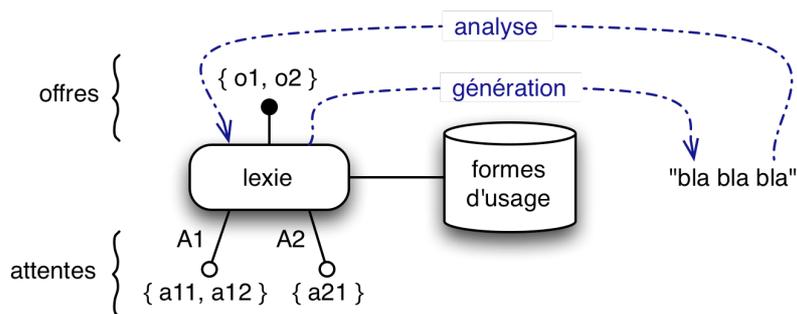


Figure 1 : Modèle de lexie à deux attentes

Les exemples de cet article sont extraits de la grammaire écrite à partir du corpus MEDIA. La sémantique étant imposée par l’annotation manuelle, les offres et les attentes des lexies sont généralement réduites à une catégorie de l’ontologie MEDIA. Il est néanmoins possible de décrire les lexies de façon plus fine, en remplaçant la catégorie par plusieurs traits, ce qui permet à l’algorithme d’analyse d’utiliser le principe *d’isotopie* (Rastier, 1987). Dans le texte qui suit, les lexies sont notées entre crochet et les traits, qui constituent les offres et les attentes, sont notés entre accolades.

## 2.1 De la lexie au granule

La représentation sémantique d’un énoncé consiste en une structure de dépendances (Tesnière, 1959) qui couvre les mots de l’énoncé correspondant à des formes d’usage reconnues. Les nœuds du graphe – les *granules* – correspondent à des lexies instanciées dans le contexte de l’énoncé (Figure 2). Cette représentation est destinée à être réinterprétée dans le double contexte du dialogue et de l’application. La connexité du graphe, qui dépend de la grammaire et de l’énoncé, peut être modifiée.

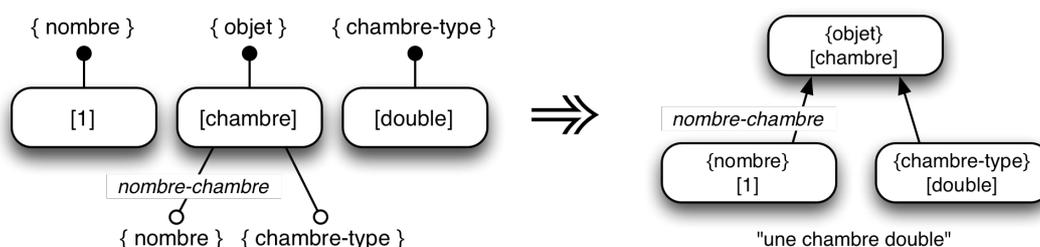


Figure 2 : Des lexies à la structure de granules

La lexie [chambre] comporte cinq attentes dont une associée à un rôle (Figure 3). Cet attribut rôle permet de contextualiser une lexie instanciée. Ainsi, une lexie ayant pour trait {nombre} dans le contexte d’une lexie [chambre] est interprétée comme un granule ayant un trait {nombre-chambre}. Ce mécanisme *d’actualisation* (Rastier, 1987) permet d’interpréter un énoncé comme « *je voudrais réserver du 15 au 18* » où les nombres 15 et 18 endossent respectivement les rôles *temps-jour-mois-debut* et *temps-jour-mois-fin*.

```
<lexie nom="chambre" offres="objet">
  <attente code="A1" traits="nombre" role="nombre-chambre"/>
  <attente code="A2" traits="chambre-type"/>
  <attente code="A3" traits="chambre-fumeur"/>
  <attente code="A4" traits="chambre-equipement"/>
  <attente code="A5" traits="chambre-standing"/>
</lexie>
```

```

<syntaxe pattern="chambre(s)"/>
<syntaxe pattern="A1 chambre(s)" exemple="deux chambres"/>
<syntaxe pattern="A1 chambre(s) A2" exemple="deux chambres simples"/>
<syntaxe pattern="A1 chambre(s) A3" exemple="deux chambres non fumeur"/>
<syntaxe pattern="A1 chambre(s) A2 A3" exemple="deux chambres simples non fumeur"/>
<syntaxe pattern="A1 A2" exemple="deux singles"/>
</lexie>

```

Figure 3 : Lexie [chambre] et quelques formes d'usage

Certaines lexies « anonymes » n'existent que pour permettre de contextualiser des lexies plus générales grâce à des formes d'usages particulières. Ces lexies peuvent ou non avoir des offres selon que l'on souhaite ou non les rattacher à d'autres lexies. Par exemple, la lexie décrite Figure 4 permet de contextualiser une lexie ayant pour trait {nombre} grâce au rôle *sejour-nbPersonne*, sans avoir à définir une lexie [personne].

```

<lexie offres="infosejour">
  <attente code="A1" traits="nombre" role="sejour-nbPersonne"/>
  <syntaxe pattern="A1 personne(s)"/>
  <syntaxe pattern="pour A1 personne(s)"/>
</lexie>

```

Figure 4 : Exemple de lexie anonyme

## 2.2 Interrogation et négation

Les questions en tant qu'actes de langage ne sont pas toujours exprimées par des phrases interrogatives. Certaines assertions peuvent même être interprétées comme des questions (Kerbrat-Orecchioni, 1991). De plus, comme les énoncés oraux qui nous intéressent ne sont pas toujours analysables de façon globale, il faut plutôt chercher des marqueurs interrogatifs plutôt que des structures grammaticalement correctes. Les questions peuvent être marquées par des séquences préliminaires comme « *(je peux) (vous) poser une question* ». Certaines expressions comme « *(je voudrais) savoir si...* », « *(pouvez-vous) me dire si...* » ainsi que des assertions modalisées comme « *je suppose que...* », « *il paraît que...* » dénotent également des questions. Même des assertions comme « *je peux amener mon chien* » ou « *les chiens ne sont pas acceptés* » peuvent être interprétées comme des questions selon le contexte.

```

<lexie nom="questionner" offres="acte">
  <attente code="A1" traits="questionnable"/>
  <syntaxe pattern="savoir si"/>
  <syntaxe pattern="est ce que"/>
  <syntaxe pattern="savoir si A1" exemple="savoir si les chiens sont acceptés"/>
  <syntaxe pattern="est ce que A1" exemple="est ce que les chiens sont acceptés"/>
</lexie>

<lexie nom="service-animaux" offres="hotel-service questionnable">
  <attente code="A1" traits="animaux"/>
  <syntaxe pattern="A1" exemple="les chiens"/>
  <syntaxe pattern="A1 sont acceptés" exemple="les chiens sont acceptés"/>
  <syntaxe pattern="A1 sont ils acceptés" traits="inter"/>
  <syntaxe pattern="A1 ne sont pas acceptés" traits="neg"/>
</lexie>

<lexie offres="animaux">
  <attente code="A1" traits="determinant"/>
  <syntaxe pattern="A1 animal/animaux"/>
  <syntaxe pattern="A1 chien(s)"/>
  <syntaxe pattern="A1 chat(s)"/>
</lexie>

```

Figure 5 : Fragment de grammaire traitant quelques formes interrogatives

Avec notre modèle, les questions sont analysables de plusieurs façons. La première consiste à utiliser des lexies dédiées comme [questionner] qui offre un trait {acte} (de langage) et qui attend une lexie optionnelle avec le trait {questionnable} (Figure 5). Ainsi, même si la lexie sur laquelle porte la question n'est pas repérable, le module de dialogue sait qu'une question a été posée. Si une lexie [questionner] précède une lexie avec le trait {questionnable} alors qu'elles ne sont pas rattachées, le module de dialogue peut inférer un rattachement hypothétique et agir en conséquence.

Une deuxième façon de traiter les questions est d'utiliser des formes d'usage interrogatives (exemple de la lexie [service-animaux] de la Figure 5). En effet, le modèle permet d'associer à chaque forme d'usage des traits spécifiques à celle-ci qui viennent s'ajouter aux traits de la lexie. Ainsi, un énoncé comme « *les chiens sont-ils acceptés* » est analysé comme une question au travers du trait **inter** sur le granule [service-animaux] (Figure 6).

Ces deux façons de prendre en compte l'interrogation peuvent se superposer si la grammaire le permet. Certaines formes négatives, qui sont aussi prises en compte au niveau des formes d'usage, peuvent également être interprétées comme des questions, selon le contexte. La méthode utilisée pour identifier certaines formes interrogatives et négatives est facilement réutilisable pour identifier et représenter d'autre forme morphosyntaxique et/ou illocutoire, comme le niveau de langage ou de politesse par exemple.

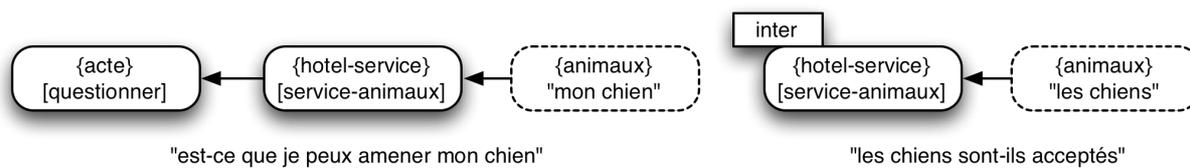


Figure 6 : Deux façons de prendre en compte l'interrogation

### 3 Réalisation

L'architecture de l'analyseur que nous avons développé est fondée sur le principe de l'analyse tabulaire ou *chart parsing* (Kay, 1986). L'idée est de développer et de garder en même temps plusieurs hypothèses ainsi que les conclusions partielles de l'analyse en utilisant la technique de mémorisation. Plusieurs algorithmes et améliorations visant à améliorer leur efficacité ont été proposés (Younger, 1967), (Earley, 1970). En ce qui nous concerne, le choix de l'analyse tabulaire a été motivé, d'une part par les caractéristiques des données (énoncés oraux), d'autre part par notre modèle de connaissances (grammaire non contextuelle). En effet, la structure de *chart* permet d'appliquer efficacement une grammaire de ce type sur un énoncé partiellement grammatical, et cela sans *backtracking*, puisque les possibilités d'analyse sont construites et conservées jusqu'à ce qu'elles soient supprimées.

Le choix de l'analyse tabulaire est aussi justifié par la poursuite de notre projet, puisque notre objectif est d'insérer notre module d'analyse sémantique dans un système de dialogue oral comprenant un module de reconnaissance de la parole qui produit, non pas une chaîne de caractères, mais un treillis de mots, un réseau de confusion, ou une liste des *n-best* hypothèses. La caractéristique de ces sorties est de présenter plusieurs possibilités où chaque possibilité est associée à un score. Le module d'analyse devra être en mesure de renforcer ou de diminuer un score en fonction de contraintes sémantiques non prises en compte au niveau de la reconnaissance de parole.

### 3.1 Analyse tabulaire

La première étape consiste à instancier des granules à partir des mots (symboles terminaux) et des granules (symboles non-terminaux) présents dans la mémoire de travail. Cette étape peut donner lieu à des conflits. Pour chaque granule, un score est calculé en fonction du nombre de mots et de ses éventuels constituants. La deuxième étape – la résolution de conflits – consiste à départager les granules conflictuels sur la base de leurs scores respectifs (certains conflits peuvent être résolus dès la première étape afin de limiter la combinatoire). La troisième étape consiste à rattacher certains granules isolés sur la base des offres et des attentes. Cette étape permet de contourner certaines structures agrammaticales, ou non décrites dans la grammaire.

La figure suivante (Figure 7) représente l’analyse de l’énoncé « *alors il me faudrait deux chambres doubles et une simple* ». Les granules supprimés sont marqués d’une croix et les granules-racines d’un point. Ceux en pointillés seront omis lors de la projection (cf. § 4.2).

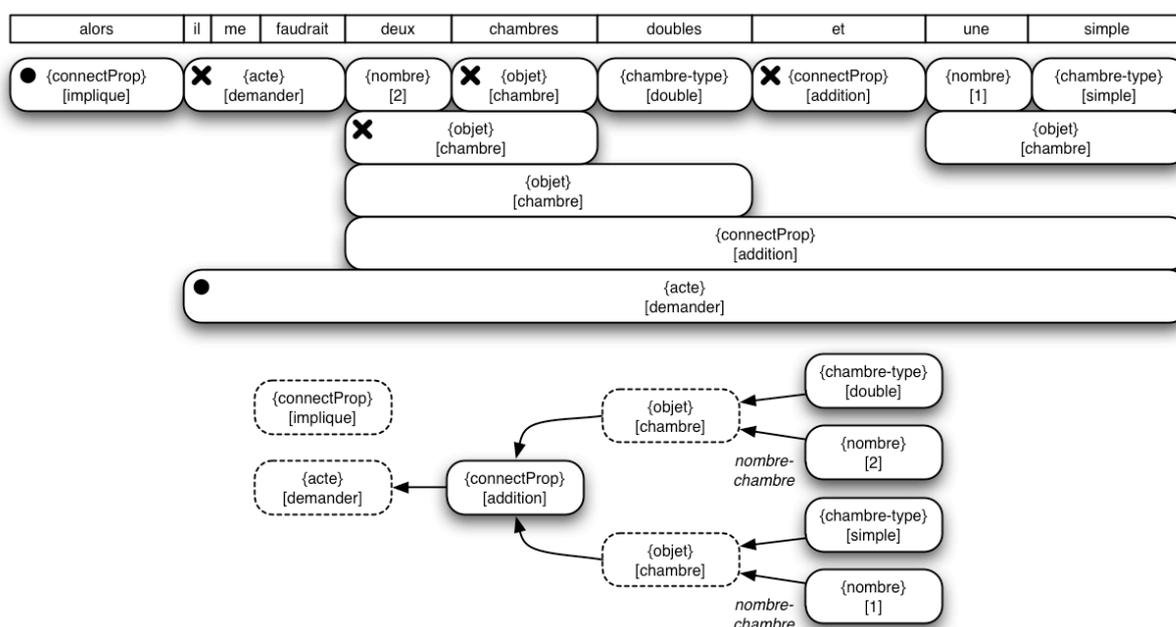


Figure 7 : Exemples d’analyse tabulaire et de sa structure de granules

### 3.2 Plateforme de programmation

Notre algorithme d’analyse tabulaire repose en réalité sur l’algorithme RETE (Forgy, 1982) qui intervient dans l’implémentation de systèmes à base de faits et de règles de production, tel que CLIPS. L’algorithme RETE compile les règles sous la forme d’un réseau de nœuds que les faits vont devoir parcourir. Le principe des « règles pointées » de l’analyse tabulaire est directement pris en compte par le moteur d’inférence : à chaque fait ajouté dans la base, c’est-à-dire à chaque reconnaissance d’un symbole non-terminal ou terminal, l’ensemble des règles qui peuvent à nouveau s’appliquer sont activées dans l’agenda du moteur d’inférence. Cet algorithme, qui tend à sacrifier la mémoire au profit de la vitesse, s’est révélé particulièrement efficace pour implémenter notre grammaire. Concrètement, les formes d’usage codées en XML sont d’abord traduites sous la forme de règles CLIPS à l’aide de transformations XSLT, puis compilées par l’algorithme RETE. Il y a une règle CLIPS par forme d’usage, soit plus de 1000 règles pour la grammaire du corpus MEDIA.

## 4 Évaluation

Afin d'évaluer les performances de notre analyseur, nous avons écrit une grammaire à partir des données du corpus MEDIA, ce qui nous a permis de positionner notre système par rapport aux cinq autres ayant participé à la campagne d'évaluation MEDIA-EVALDA de 2005.

La méthode d'évaluation s'appuie, d'une part sur une représentation sémantique de référence, d'autre part sur un corpus annoté manuellement selon cette représentation (Maynard et al., 2004, 2005). L'évaluation d'un analyseur consiste alors à comparer l'analyse d'un corpus de test (utilisé uniquement pour l'évaluation) aux résultats annotés manuellement.

Le corpus se présente sous la forme de fichiers XML regroupant 1257 dialogues, soit plus de 15000 énoncés, dont 3000 dédiés à la phase d'évaluation. Si nous avons pris le temps d'améliorer notre approche, nous avons respecté le protocole de la campagne d'évaluation afin d'obtenir des résultats comparables à ceux des participants. Comme ces derniers, nous avons développé un module de projection de notre propre représentation vers la représentation de référence. Dans les paragraphes suivants, nous décrivons le modèle de référence, notre méthode de projection, et les résultats que nous obtenons.

### 4.1 Le modèle des triplets de MEDIA

Le choix de la représentation de référence a été guidé par des contraintes liées à l'annotation manuelle. Premièrement, les représentations hiérarchiques ont été écartées au bénéfice d'une représentation plate qui suit l'ordre d'apparition des mots. Deuxièmement, il a été décidé de ne représenter que les segments porteurs de sens pour l'application, et non tous les mots de l'énoncé.

Chaque segment est représenté par un triplet <mode:attribut:valeur>. Le mode indique si le segment correspond à une affirmation, à une négation, à une interrogation ou à une possibilité. L'attribut correspond à un concept simple ou composé de l'ontologie MEDIA. La valeur correspond à une instanciation du concept dans l'énoncé. L'ontologie contient 83 concepts simples, 19 spécifieurs et au total 1121 concepts simples et composés (c'est-à-dire avec spécifieurs). Le tableau suivant contient la représentation sous forme de triplets de l'énoncé « *alors il me faudrait deux chambres doubles et une simple* ».

Mots	Mode	Attribut	Valeur
deux chambres	+	nombre-chambre	2
doubles	+	chambre-type	double
et	+	connectProp	addition
une	+	nombre-chambre	1
simple	+	chambre-type	simple

Tableau 1 : Exemple de représentation selon le modèle des triplets de MEDIA

## 4.2 Projection sur le modèle des triplets

La projection est l'opération qui consiste à instancier une liste ordonnée de triplets à partir d'une structure de granules, connexe ou non. Nous avons identifié deux cas de projection (Figure 8) et quelques transformations.

Le cas général consiste à transformer les traits de la lexie en attributs, les noms des lexies en valeurs normalisées, et les traits spécifiques aux formes d'usage (f-traits) en modes. Par exemple, la projection des granules de l'énoncé « *les chiens sont-ils acceptés* » (Figure 6) produit le triplet <?:hotel-service:service-animaux>, le granule anonyme de trait {animaux} n'étant pas projeté.

Les granules dotés d'un rôle sont projetés différemment, le rôle étant prédominant sur le trait. Par exemple, la projection des granules de l'énoncé « *une chambre double* » (Figure 2) produit les triplets <+:nombre-chambre:1> et <+:chambre-type:double>, le triplet <+:objet:chambre> étant supprimé par respect des consignes d'annotation.

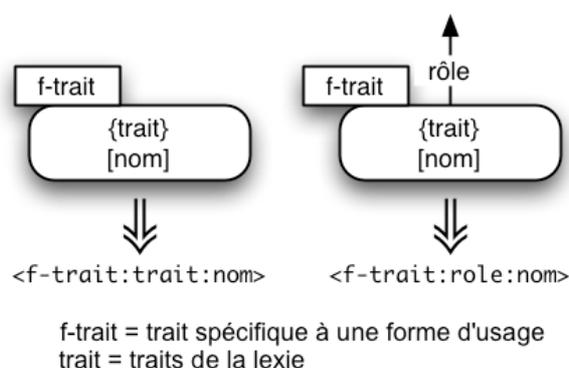


Figure 8 : Deux cas de projection des granules vers les triplets

Après la première étape de passage des granules vers le modèle des triplets, une seconde étape de transformation des triplets finalise le résultat, selon les consignes de l'annotation manuelle. Nous décrivons ci-dessous les transformations les plus significatives : l'inférence de mode et l'inférence de spécifieur.

Le cas de l'énoncé « *est ce que je peux amener mon chien* » est un exemple de modification d'un mode par un granule père. Les granules produits par notre analyse (Figure 6) sont projetés dans un premier temps en deux triplets : <+:acte:questionner> et <+:hotel-services:service-animaux>, le granule anonyme de trait {animaux} n'étant pas projeté. La lexie [questionner] ne faisant pas partie de l'ontologie MEDIA, le triplet correspondant ne sera pas conservé. Mais avant de disparaître, il agira sur le mode du triplet correspondant au granule fils [service-animaux]. Il ne restera que le triplet <?:hotel-services:service-animaux>.

Le cas de l'énoncé « *je voudrais réserver une chambre double du premier au 6 décembre* » illustre le principe de propagation de spécifieur. Les triplets produits lors de l'étape n°1 sont : <+:command-tache:reservation>, <+:nombre-chambre:1>, <+:chambre-type:double>, <+:temps-date-debut:01/12> et <+:temps-date-fin:06/12>. Le manuel d'annotation<sup>1</sup> stipule

<sup>1</sup> [http://www.limsi.fr/Individu/hbm/manuel\\_annotation\\_11\\_04.html](http://www.limsi.fr/Individu/hbm/manuel_annotation_11_04.html) (accédé le 06/03/09).

que si l'action de *réserver* est exprimée, on propage un spécifieur « reservation » sur certains attributs (dont nombre-chambre, temps-date-debut et temps-date-fin). Le triplet <+:nombre-chambre:1> devient alors <+:nombre-chambre-reservation:1>, le triplet <+:temps-date-debut:01/12> devient alors <+:temps-date-debut-reservation:01/12>, etc.

### 4.3 Résultats

L'évaluation consiste en un taux d'erreur calculé à partir de l'alignement du résultat obtenu et du résultat de référence. Ce taux est égal à  $SUB + DEL + INS / TOT$  où SUB est le nombre de triplets modifiés, DEL le nombre de triplets supprimés, INS le nombre de triplets ajoutés, et TOT le nombre de triplets à identifier. Quatre taux sont calculés selon que l'on tient compte des quatre modes ou seulement des modes affirmatif et négatif, et selon que l'on n'utilise que les concepts de base (*relax scoring*) ou si l'on fait usage des spécifieurs (*full scoring*).

Full 4 modes	Full 2 modes	Relax 4 modes	Relax 2 modes
LIMSI-1 : 29,0	LIMSI-2 : 23,2	LIMSI-1 : 27,0	LIMSI-2 : 19,6
LIMSI-2 : 30,3	LIMSI-1 : 23,8	LIMSI-2 : 27,2	LIMSI-1 : 21,6
<b>LIUM : 34,7</b>	LORIA : 28,9	LIA : 29,8	LIA : 24,1
LORIA : 36,3	<b>LIUM : 30,0</b>	<b>LIUM : 31,9</b>	LORIA : 24,6
VALORIA : 37,8	VOLORIA : 30,6	LORIA : 32,3	<b>LIUM : 26,9</b>
LIA : 41,3	LIA : 36,4	VALORIA : 35,1	VALORIA : 27,6

Tableau 2 : Tableau comparatif des taux d'erreurs

Nous obtenons un score relativement bon dans le mode le plus exigeant (full 4 modes). Les écarts de 4,7% et 5% constatés entre les taux « 2 modes » et « 4 modes » correspondent à l'identification des interrogations et des demandes de possibilité, difficiles à traiter sans prise en compte du contexte de l'application. Cet écart étant en moyenne de 6,57% pour les autres systèmes, notre analyseur serait plus précis sur cet aspect. Nous constatons un écart moyen de 2,9% entre les taux « relax » et « full ». Cela peut s'expliquer par la relative proximité de notre représentation sémantique avec celle de MEDIA, ce qui permet d'éviter certaines erreurs de projection. En tenant compte des deux écarts précédents, on peut dire que sur les 34,7% du mode « full 4 modes », il reste environ 27% d'erreurs dont les causes ont principalement pour origine des problèmes d'identification des expressions référentielles et des connecteurs (également une source d'erreurs pour les autres systèmes). Enfin, nous avons retrouvé une source d'erreurs identifiée dans (Villaneau, Lamprier, 2005) liée à la projection d'une représentation hiérarchique vers une représentation linéaire. En effet, les triplets suivent l'ordre d'apparition des mots, alors que nos structures de granules ne sont pas dépendantes de cet ordre. Quelque soit la méthode de projection, nous obtenons des inversions par rapport à l'ordre attendu, qui augmentent conjointement les taux de substitutions et d'insertions.

## 5 Conclusion

Le modèle sémantique, ainsi que le principe d'analyse pour le dialogue homme-machine que nous présentons dans cet article, ont été développés avant le choix du corpus MEDIA pour leur évaluation. Ils ne sont donc pas *ad hoc*. Nous retenons de cette expérience qu'avec un modèle linguistique simple (une grammaire sémantique non contextuelle), mais adapté à notre

problématique, un algorithme et une architecture minimalistes (nous ne faisons reposer notre interprétation sur aucun dictionnaire, ni segmentation morphosyntaxique), nous obtenons des résultats comparables à ceux des autres systèmes, en particulier sur les modes d'évaluation les plus exigeants. Ceci dit, les problèmes qui ont contribué à faire monter nos taux d'erreurs semblent précisément liés à un manque de pré-segmentation de l'énoncé (identification des groupes nominaux, des relatives, *etc.*). Comme d'autres participants, nous avons regretté les limites du modèle sémantique de référence, qui nous a contraint à perdre la dimension hiérarchique lors de l'étape de projection, tout en induisant une source importante d'erreurs.

## Références

BLANCHE-BENVENISTE C., BILGER M., ROUGET C., VAN DEN EYNDE K. (1990). *Le Français parlé : études grammaticales*. Paris : CNRS Editions.

BONNEAU-MAYNARD H., ROSSET S., AYACHE C., KUHN A., MOSTEFA D. (2005). Semantic Annotation of the French MEDIA dialog corpus. *Proceedings of Eurospeech 05*, Lisboa.

BONNEAU-MAYNARD H., AYACHE C., BECHET F., DENIS A., KUHN A., LEFEVRE F., MOSTEFA D., QUIGNARD M., ROSSET S., SEVRAN C., VILLANEAU J. (2006). *Results of the french EVALDA-MEDIA evaluation campaign for literal understanding*. Actes de LREC 2006, Gênes, Italie, 2054–2059.

EARLEY J. (1970). An efficient context-free parsing algorithm. *Communications of the Association for Computing Machinery* 13(2), 94-102.

FORGY, C.L. (1982). Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem. *Artificial Intelligence* (19), 17-37.

KAY M. (1986). Algorithms chemata and data structures in syntactic processing. *Readings in natural language processing*. San Francisco, CA, USA : Morgan Kaufmann Publishers Inc.

KERBRAT-ORECCHIONI C. (1991). L'acte de question et l'acte d'assertion : opposition discrète ou continuum ?, *La Question*. Presses Universitaires de Lyon, 87-111.

LEHUEN J. (2008). Un modèle de langage pour le DHM : la Grammaire Sémantique Réversible, *Actes de TALN'08*, 9-13 juin 2008, Avignon (France).

RASTIER F. (1987). *Sémantique interprétative*. Paris : PUF.

TESNIÈRE L. (1988). *Éléments de syntaxe structurale*. Paris : Editions Klincksieck.

VILLANEAU J., LAMPRIER S. (2005). Corpus de dialogue homme/machine : annotation sémantique et compréhension. Actes des Journées Internationales de Linguistique de Corpus 2005, Lorient, France, 221-229.

YOUNGER D. (1967). Recognition and parsing of context-free languages in time  $n^3$ . *Information and Control* 10(2), 189–208.