
Approche bayésienne de la composition sémantique dans les systèmes de dialogue oral

Marie-Jean Meurs, Fabrice Lefèvre et Renato de Mori

*Université d'Avignon et des Pays de Vaucluse
Laboratoire Informatique d'Avignon (EA 931), F-84911 Avignon, France.
{marie-jean.meurs,fabrice.lefevre,renato.demori}@univ-avignon.fr*

RÉSUMÉ. Cet article présente un processus stochastique pour la composition de structures sémantiques, dédié à l'interprétation de dialogues oraux homme-machine. Ce processus permet d'inférer des structures hiérarchiques de cadres sémantiques (CS) à partir des séquences de mots et de concepts de base issues de l'énoncé utilisateur. Un processus logique est d'abord mis en œuvre pour obtenir une annotation de référence de données d'entraînement. Il est alors possible d'apprendre les paramètres de réseaux bayésiens dynamiques qui sont utilisés pour produire les hypothèses de structures de CS sur les données de test. Des règles logiques viennent finalement compléter les liaisons manquantes entre les CS obtenus lors du décodage séquentiel. Cette approche probabiliste présente l'intérêt de produire des listes scorées d'hypothèses d'arbres sémantiques. Les expériences, menées sur le corpus de dialogues en français MEDIA, montrent que l'approche probabiliste proposée permet de générer une annotation en CS suffisamment fiable et robuste pour être utilisée dans un système de dialogue réel.

ABSTRACT. Focusing on the interpretation component of spoken dialog systems, this paper introduces a stochastic approach based on dynamic Bayesian networks to infer and compose semantic structures from speech. Word strings, basic concept sequences and composed semantic frames (as defined in the Berkeley FrameNet paradigm) are derived sequentially from the users' inputs. A semi-automatic process provides a reference frame annotation of the speech training data. Then the DBN trained on these data are used to hypothesize the frames and their constituents from the test data. Eventually a rule-based process produces the final composed frame annotation. Experimental results on the French MEDIA dialog corpus show the appropriateness of the technique which both lead to good semantic tree identification performance and can provide the dialog manager with n-best lists of scored hypotheses.

MOTS-CLÉS : système de dialogue oral, composition sémantique, réseaux bayésiens dynamiques.

KEYWORDS: spoken dialog system, spoken language understanding, semantic frame composition, dynamic Bayesian networks.

1. Introduction

La complexité des systèmes de dialogue oral homme-machine dépend essentiellement de la complexité de la tâche visée. Ainsi les systèmes dédiés au routage d'appels téléphoniques ou encore à la recherche d'information pouvant se présenter sous forme de formulaires (répertoires téléphoniques ou recherche d'horaires entre autres exemples) peuvent reposer sur une représentation des connaissances sémantiques relativement simple. *A contrario* dès lors qu'un système doit pouvoir gérer plusieurs demandes conjointes ou intégrer des phases de négociation, une représentation sémantique de haut niveau est requise. De tels systèmes doivent disposer de connaissances sémantiques permettant d'interagir avec des informations complexes sur l'état courant du dialogue tout en supportant les erreurs introduites dans la chaîne de traitement du signal de parole par les modules de reconnaissance et de compréhension automatiques. La difficulté à établir une telle représentation de haut niveau, qui soit à la fois fonctionnelle et robuste, explique que les systèmes de dialogue actuels soient limités à des univers sémantiques assez restreints.

Les domaines d'application des systèmes de dialogue sont variés. Depuis le système ATIS (*Air Travel Information Service*, projet ARPA) (Price, 1990), dédié à la consultation d'horaires d'avions aux Etats-Unis, l'essor des services basés sur la téléphonie interactive a motivé le développement de nombreux systèmes parmi lesquels "How May I Help You" (Gorin *et al.*, 1997) d'AT&T qui s'intéresse au routage d'appels téléphoniques, "ARISE" (*Automatic Railway Information Systems for Europe*) développé au LIMSI (Lamel *et al.*, 2000), dédié à la recherche d'information et d'horaires ferroviaires ou encore le service "3000", évolution du serveur "AGS" (Sadek *et al.*, 1995), réalisé et exploité commercialement par France Télécom (Damnati *et al.*, 2007) qui permet aux utilisateurs de gérer leur compte client, de contrôler leur consommation ou même de payer leurs factures.

Les applications dédiées au tutorat, notamment dans le contexte de l'apprentissage des langues, sont nombreuses à utiliser des systèmes de dialogue. Développées dans l'esprit des tuteurs intelligents des années 1970-80, les applications du domaine sont encore souvent basées sur le dialogue textuel mais de nombreux projets académiques récents intègrent avec succès l'oral comme modalité d'échange (ITSPOKE (Litman *et al.*, 2004), SCoT (Pon-Barry *et al.*, 2006)...).

Les systèmes de dialogue peuvent également être utilisés dans le cadre de l'aide aux personnes handicapées (voir (Hockey *et al.*, 2007) qui présentent le pilotage d'un fauteuil électrique par le biais d'un dialogue oral entre l'utilisateur et la machine) ou aux personnes âgées (voir l'adaptation du système d'information LET'S GO ! aux personnes âgées (Raux *et al.*, 2003)).

Développer un système de dialogue capable de se remettre en cause, de proposer des alternatives et de s'adapter aux phases de négociation présentes dans le dialogue "naturel" suppose donc la construction d'une nouvelle représentation des connaissances. Elle doit permettre la composition sémantique au sein des tours de parole comme au cours du déroulement du dialogue. L'objectif de notre travail est de pro-

duire une information sémantique plus riche à fournir au module de gestion du dialogue, d'une façon pertinente et facilement généralisable. Pour ce faire, une étape de composition sémantique est considérée afin de capturer les notions sémantiques abstraites sous-tendues par la représentation conceptuelle de base, généralement utilisée dans de tels systèmes (voir par exemple (Lefèvre, 2007)).

En l'absence de consensus sur la définition de structures de représentation sémantique adaptées à un système de dialogue oral, nous avons choisi d'utiliser le formalisme des cadres sémantiques (CS) et de les définir en accord avec le paradigme du projet FrameNet de Berkeley (Fillmore *et al.*, 2003). Les structures de CS ont été retenues en raison de leur capacité à représenter les dialogues de négociation et à s'adapter aux actions complexes du gestionnaire de dialogue. Décrivant une situation abstraite ou concrète impliquant des notions prédéfinies, un CS comporte des éléments (CSE), ou rôles sémantiques, qui lui sont propres. Les CS définis dans le projet FrameNet étant pour la plupart très généralistes, des CS adaptés à la tâche visée ont été élaborés.

Un processus semi-manuel à base de règles a été développé pour fournir une annotation en CS et CSE de dialogues manuellement transcrits et annotés en concepts de base. Cette annotation, bien qu'imparfaite, est assez fiable. Cependant, dans le cadre d'une utilisation au sein d'un système de dialogue réel, les erreurs commises sur les mots et les concepts par les systèmes automatiques de transcription et d'annotation conceptuelle doivent être prises en compte. Il est donc nécessaire de concevoir un système capable de produire des listes de n -meilleures hypothèses utilisables lors des étapes de validation ultérieures, ce que ne permet pas une approche déterministe à base de règles.

Dans un système de dialogue oral, le module de compréhension relie le système de reconnaissance automatique de la parole (RAP) au gestionnaire de dialogue. Après analyse de la proposition de l'utilisateur, il construit une représentation de son contenu sémantique. A partir de cette représentation, le gestionnaire de dialogue peut choisir la meilleure action à réaliser, considérant le contexte courant du dialogue. Les approches stochastiques limitent le besoin d'expertise humaine et réduisent en cela les coûts de développement. De plus, ces approches sont capables de produire des réseaux d'hypothèses (ou les listes des n -meilleures hypothèses) avec leurs indices de confiance afin de propager l'incertitude jusqu'au module de décision.

De façon générale, l'introduction de processus stochastiques dans chaque composant des systèmes de dialogue permet d'améliorer leurs performances en augmentant leur robustesse aux variabilités de la parole. Ainsi, il est maintenant admis que les méthodes probabilistes sont d'efficaces alternatives aux méthodes à base de règles pour la compréhension littérale de l'oral (Levin *et al.*, 1995, He *et al.*, 2005, Lefèvre, 2007). Des systèmes de compréhension automatique de la parole dans lesquels la totalité du processus de compréhension est stochastique ont été présentés dans de précédents travaux (Bonneau-Maynard *et al.*, 2005a, Raymond *et al.*, 2006, Lefèvre, 2007). Généralement conçus pour gagner en robustesse, ces systèmes fonctionnent par raffinement progressif des hypothèses de concepts produites.

Dans cette optique, nous proposons un système de compréhension automatique de la parole basé sur deux étapes de décodage utilisant des réseaux bayésiens dynamiques (RBD). La première étape déduit les concepts de base de la transcription des propos du locuteur (Lefèvre, 2007). La seconde étape réalise des inférences séquentielles sur des CS et des CSE, en tenant compte de tous les niveaux d'annotation disponibles (mots et concepts). Trois variantes du système sont présentées et évaluées. Dans le premier modèle, CS et CSE sont décodés conjointement comme une seule variable. Dans le second modèle, CS et CSE sont séparés en deux variables distinctes décodées simultanément. Enfin, le dernier modèle comporte deux niveaux : les CS sont décodés puis considérés comme des observations utilisées pour le décodage des CSE.

Les premiers résultats obtenus par nos systèmes basés sur les RBD ont été récemment présentés (Meurs *et al.*, 2008b, Meurs *et al.*, 2009). Les bonnes performances observées montrent qu'à partir d'un corpus d'entraînement annoté en CS et CSE, un décodage séquentiel permet de découvrir et de composer les CS et CSE associés aux propos d'un locuteur, et cela malgré la prise en compte de dépendances éloignées lors de l'étape de production des données d'entraînement.

Le corpus de dialogues en français MEDIA, utilisé dans nos expériences, est présenté dans la section 2. La section 3 rappelle ensuite les notions fondamentales qui sous-tendent les définitions des CS et CSE et décrit le processus en deux étapes à base de règles utilisé pour fournir l'annotation de référence en CS et CSE du corpus MEDIA. La section 4 expose les modèles à base de RBD employés pour la composition des CS et CSE. Enfin, les résultats expérimentaux sont détaillés dans la section 5.

2. Le corpus MEDIA

Le corpus MEDIA est un corpus de dialogues en français issu de la simulation d'un serveur téléphonique d'informations touristiques et de réservation d'hôtels (Bonneau-Maynard *et al.*, 2005b). Il a été enregistré selon le protocole du *Magicien d'Oz* (système de dialogue simulé par un opérateur humain) et contient 1 257 dialogues produits par 250 locuteurs. La durée totale d'enregistrement audio est d'environ 70 heures. Huit catégories de scénarii ont été définies avec différents niveaux de complexité. Chaque locuteur a enregistré cinq scénarii différents.

Une description détaillée des caractéristiques techniques du corpus est donnée dans le tableau 1.

La collecte du corpus et la gestion technique du projet MEDIA ont été prises en charge par ELDA/ELRA¹. La société VECSYS² a mis en place la plate-forme d'enregistrement du corpus (matériel et outil WoZ).

Le corpus est transcrit manuellement et enrichi par une annotation conceptuelle également manuelle utilisant 83 concepts de base rassemblés dans un dictionnaire sé-

1. <http://www.elda.org>

2. <http://www.vecsys.fr>

nombre de phrases utilisateur	18.801
nombre de phrases système	19.633
nombre de mots moyen par phrase utilisateur	8,3
nombre de mots moyen par phrase système	14,4
taille du vocabulaire utilisateur	2.715 mots
taille du vocabulaire système	1.932 mots
durée moyenne d'un dialogue	3 min et 30 s

Tableau 1. *Caractéristiques du corpus de dialogues MEDIA*

mantique de concepts. La mesure de l'accord inter-annotateur est toujours supérieure à 80 % et atteint jusqu'à 90 % dans le meilleur cas (Bonneau-Maynard *et al.*, 2005b).

Le dictionnaire sémantique utilisé pour l'annotation du corpus MEDIA associe à un mot (ou une séquence de mots) une paire *concept-valeur* puis un *spécifieur* définissant des relations entre concepts et enfin un *mode* (affirmatif, négatif, interrogatif ou optionnel) attaché au concept. Avec 19 spécifieurs pouvant être associés aux 83 concepts de base, le schéma d'annotation MEDIA offre un mécanisme simple permettant de préserver certaines relations élémentaires entre les concepts au sein de la phrase.

Un exemple de message annoté du corpus MEDIA est donné dans le tableau 2. La première colonne contient les séquences de mots W^c , supports de chaque concept présent dans la seconde colonne. La troisième colonne indique le mode et la quatrième colonne fournit les spécifieurs associés aux concepts. La dernière colonne présente les valeurs normalisées des concepts c associés aux séquences W^c .

W^c	<i>concept c</i>	<i>mode</i>	<i>spécifieur</i>	<i>valeur</i>
<i>Je voudrais réserver une chambre pour deux nuits à Marseille</i>	commande	+		réservation
	chambre-quantité	+	réservation	1
	séjour-nbNuit	+	réservation	2
	localisation-ville	+	hôtel	Marseille

Tableau 2. *Exemple d'annotation sémantique du corpus MEDIA*

Dans cet exemple, le spécifieur *réservation* est lié aux concepts *chambre-quantité* et *séjour-nbNuit*. On obtient ainsi une structure hiérarchique représentant une réservation associée au concept *commande* et développée grâce aux valeurs des concepts *chambre-quantité* et *séjour-nbNuit*. Le spécifieur *hôtel* adjoint au concept *localisation-ville* permet de relier le lieu évoqué dans le segment "à Marseille" à la partie précédente de l'énoncé.

La combinaison des spécifieurs et des concepts permet de recomposer une représentation hiérarchique de la requête du locuteur à partir de l'annotation à *plat*. Cette annotation fournit des étiquettes comparables aux constituants proposés par un analyseur sémantique de surface. Cependant, pour obtenir une représentation complète de la composition sémantique d'une proposition, l'utilisation de structures plus riches et plus complexes est nécessaire.

3. Annotation en *cadres sémantiques*

Les connaissances sémantiques théoriques permettent de définir des structures sémantiques adaptées à divers domaines ainsi que l'ont montré (Woods, 1975) avec les réseaux sémantiques représentant des entités/rerelations ou (Jackendoff, 1990) avec les structures de type fonction/argument. Une façon efficace de modéliser les connaissances sémantiques est de les représenter comme un ensemble de formules logiques sur lequel repose le processus de compréhension. Dans ce contexte, Fillmore (1982, 1985) définit les CS comme des structures cognitives associées au processus de compréhension. Des éléments (CSE), ou rôles sémantiques, sont associés en propre à chaque CS. Un CS est donc un modèle représentant des entités sémantiques et leurs propriétés (Petrucci, 1996, Lowe *et al.*, 1997).

Le choix d'une annotation en CS et CSE dans ce travail est motivé par leur capacité à représenter des dialogues intégrant des phases de négociation et à s'adapter aux actions complexes du gestionnaire de dialogue dans ce contexte. Un CS décrit une situation concrète ou abstraite impliquant ses CSE. Les mots ou groupes de mots déclenchant l'instanciation d'un CS ou d'un CSE sont ses *unités lexicales* (LU). Ces LU associent un mot (ou une séquence de mots) à un sens.

Le projet FrameNet de l'Université de Berkeley fournit une base de données de CS pour la langue anglaise (Fillmore *et al.*, 2003). Actuellement, celle-ci contient environ 825 CS reliés hiérarchiquement et plus de 10 000 LU. Une ressource de 135 000 propositions annotées à l'aide de ces CS et de leurs CSE est également disponible dans le cadre du projet.

Les premières bases de données comparables disponibles pour la langue française (Pado *et al.*, 2007) ne comportant pas les objets sémantiques indispensables à la couverture de notre domaine, nous avons défini manuellement un ensemble de CS et de CSE pour décrire nos connaissances en termes de composition sémantique adaptée au corpus MEDIA (Meurs *et al.*, 2008a). Cette base de connaissances contient 21 CS et 86 CSE. Elle est présentée graphiquement en Annexe 1. Elle est donc de taille très modeste en comparaison des versions française de FrameNet (Pado *et al.*, 2007) ou anglaise. Le tableau 3 reprend les informations sur le dimensionnement des trois ressources (base de connaissances MEDIA, FrameNet français et FrameNet). A titre de comparaison, le CS *REQUEST*, commun aux trois ressources, est présenté avec ses CSE selon les différentes versions.

	base MEDIA	FrameNet français	FrameNet
CS	21	138	> 825
CSE	86	1371	6800
CS	REQUEST	REQUEST	REQUEST
CSE	<i>agent</i> <i>recipient</i> <i>theme</i>	<i>speaker</i> <i>medium</i> <i>adresse</i> <i>manner</i> <i>message</i> <i>means</i> <i>topic</i>	<i>speaker</i> <i>medium</i> <i>adresse</i> <i>manner</i> <i>message</i> <i>means</i> <i>topic</i> <i>beneficiary</i> <i>time</i>

Tableau 3. Comparaison des 3 bases de connaissances : base MEDIA - FrameNet français - FrameNet. Exemple du cadre sémantique *REQUEST* et de ses éléments dans chacune des bases

Les CS et CSE sont définis par des modèles composés de LU et de concepts de base (*unités conceptuelles*, CU). Ces CU sont issues du dictionnaire sémantique de concepts MEDIA, dès lors qu'ils peuvent être associés à un CS ou à un CSE. Dans le cas contraire, des CU adaptées ont été définies. La base de connaissances MEDIA contient actuellement 106 CU et plus de 1 000 LU. Le tableau 4 présente un extrait de la définition du CS *LOCATION* avec l'un de ses CSE, *location_town*.

```
<frame csname="LOCATION">
  <concept value="localisation" />
  <lexical_units value="lieu,endroit" />
  <framelement csname="location_town">
    <concept value="localisation_ville" />
    <specific_lexical_units value="paris,marseille..." />
  </framelement> ...
</frame>
```

Tableau 4. Extrait de la définition du CS MEDIA *LOCATION* et de l'un de ses CSE *location_town*

La principale différence structurelle entre la base de connaissances MEDIA et les bases de type FrameNet française ou anglaise réside dans sa construction hiérarchique. En effet, un CSE de cette base peut prendre pour valeur un CS. Cela permet une représentation par *arbres sémantiques* des propositions du locuteur. Cette caractéristique n'est pas proposée pour les objets sémantiques de type strictement FrameNet. Un exemple d'arbre sémantique est présenté dans la figure 1.

Pour obtenir des annotations initiales en CS et CSE sur les données d'entraînement, un processus d'annotation en deux étapes à base de règles a été développé :

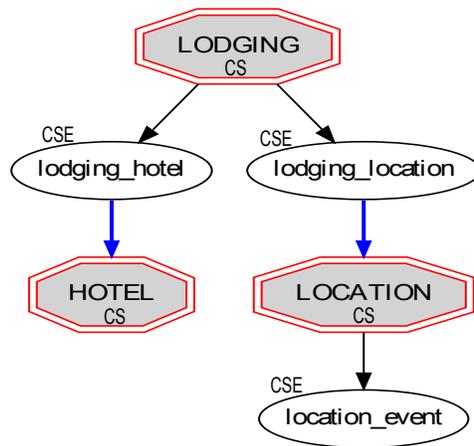


Figure 1. CS, CSE et relations associés à la séquence de mots “séjourner dans un hôtel proche du Festival de Cannes”

1) la première étape utilise les modèles définissant CS et CSE pour déclencher l’instanciation de CS et de leurs CSE selon que LU et CU sont rencontrés dans les données à annoter ;

2) la seconde étape compose les CS et CSE proposés durant l’étape précédente ou en infère de nouveaux par le biais de l’application d’une série de règles logiques. CS et CSE déterminent les valeurs de vérité des prédicats des règles. Selon ces valeurs de vérité, des CS et CSE peuvent être créés, supprimés, modifiés ou reliés. Ce dernier cas est justifié par la hiérarchie présente au sein des CS, traduite par l’aptitude de certains CSE à prendre des CS pour valeurs.

Le langage de programmation logique Prolog (Colmerauer *et al.*, 1993) a été utilisé pour réaliser les inférences logiques. Un programme Prolog se compose d’une base de faits et d’un ensemble de règles logiques décrivant les relations entre des faits potentiels. L’implémentation des règles est réalisée sous SWI-Prolog (Wielemaker, 2003). Une règle Prolog est de la forme :

```

do_lien(LODH, H) :-
    is_cse(lodging_hotel, LODH),
    is_concept_of(hotel, LODH),
    is_cs(HOTEL, H) .
  
```

Dans cet exemple, la règle crée un lien entre le CSE `lodging_hotel` et le CS `HOTEL` sous la condition que le CSE `lodging_hotel` ait été déclenché par la présence du concept `hotel`. Le CSE `lodging_hotel` prend ainsi le CS `HOTEL` pour valeur.

Environ 70 règles sont actuellement appliquées. Elles ne prennent en compte ni les mots, ni l'ordre d'instanciation des CS et CSE lors de l'étape 1. La plupart de ces règles relient CS et CSE, instancient ceux non découverts par l'étape 1 et suppriment les redondances. L'inférence logique est appliquée itérativement, chaque sortie fournissant les faits soumis à l'étape de résolution suivante. Un exemple de tour de parole annoté est donné dans l'Annexe 2.

Ce processus, dédié à la tâche d'annotation des données d'apprentissage, est progressivement enrichi pour améliorer ses performances. Il permet d'obtenir une annotation en CS et CSE de référence pour le corpus d'entraînement sur lequel les modèles stochastiques peuvent être appris.

4. Modèles RBD pour la composition des cadres sémantiques

Les réseaux bayésiens dynamiques sont des modèles d'une grande flexibilité permettant de représenter des systèmes stochastiques complexes. Leur adaptabilité autorise des modélisations variées qu'il est peu coûteux de faire évoluer. Les RBD, utilisés dans de nombreuses tâches de modélisation de données séquentielles, obtiennent des résultats au niveau de l'état de l'art (voir (Lefèvre, 2007) pour un exemple d'application des RBD au problème de la compréhension automatique de la parole).

4.1. Structure des modèles

Les modèles génératifs à base de RBD utilisés pour la composition des CS dans notre système de compréhension sont présentés dans les figures 2, 3 et 4. Les sommets symbolisent les variables et les arcs matérialisent les dépendances conditionnelles. Par souci de lisibilité, certains sommets et certains arcs ne sont pas représentés.

Dans les figures, seuls deux événements temporels (soit deux mots) sont indiqués. En pratique, ce schéma est répété autant que nécessaire au long de la séquence de mots étudiée. Les sommets sont grisés lorsque les variables sont observées tandis qu'ils restent clairs pour les variables cachées du modèle. Les arcs pleins illustrent les dépendances conditionnelles entre les variables, les arcs en traits pointillés indiquent les relations aux variables transitionnelles (modifiant les relations conditionnelles entre les variables qu'elles désignent).

Un exemple de variable transitionnelle est donné par le sommet clair CS_{trans} qui influence le sommet CS. Ainsi, si CS_{trans} a une valeur nulle, le CS est identique à son prédécesseur. Lorsque CS_{trans} est égal à un, la nouvelle valeur du CS est déterminée en fonction de la probabilité $P(cs|cs_{-1})$ du CS cs connaissant le CS précédent cs_{-1} .

La figure 2 décrit le modèle génératif *compact* à base de RBD dans lequel CS et CSE sont représentés par une seule variable composite. Dans la figure 3, le modèle *factorisé* considère des CS et CSE représentés par deux variables distinctes mais si-

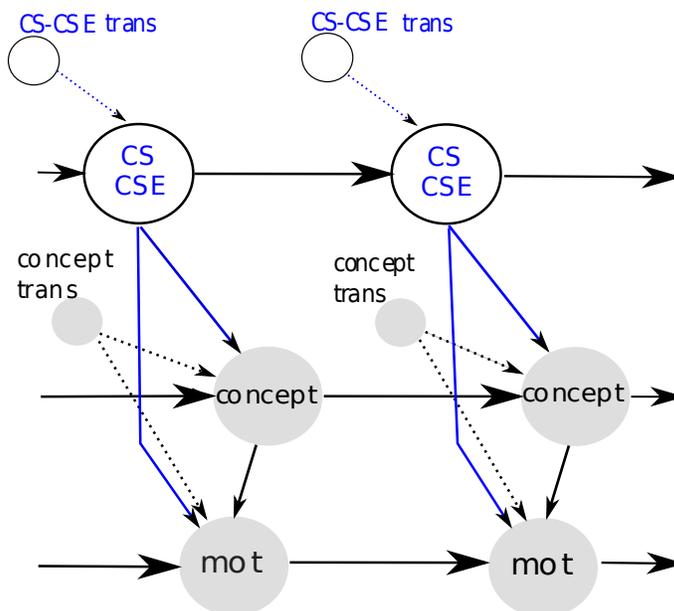


Figure 2. CS et CSE considérés comme une seule variable non observée

multanément décodées. Le modèle à *deux niveaux* dans lequel les CS sont décodés en premier lieu, puis utilisés comme des valeurs observées lors du décodage des CSE est illustré par la figure 4.

Le choix de représenter CS et CSE par une seule variable est essentiellement motivé par la réduction de la complexité du décodage : la cardinalité de l'ensemble des valeurs possibles de CS/CSE est limitée au nombre de combinaisons de CS/CSE observées dans les données d'apprentissage. Cependant, cette approche conduit à l'utilisation de liens déterministes et fixés entre CS et CSE.

Le modèle factorisé permet en revanche de considérer les ambiguïtés des liens entre CS et CSE en leur attribuant des probabilités et testant chaque combinaison au cours du décodage. Les combinaisons non rencontrées dans les données d'apprentissage sont évaluées grâce à des techniques de repli. Évidemment, cette approche est coûteuse et la complexité du modèle est telle qu'il est nécessaire d'utiliser un algorithme sous-optimal de recherche en largeur (de type recherche en faisceau, *beam search*) pendant le décodage.

Enfin, bien qu'également sous-optimale, l'approche à deux niveaux de décodage propose un modèle de complexité moindre que celle du modèle factorisé tout en conservant l'aptitude à décodifier des CS et CSE reliés par des liens non déterministes.

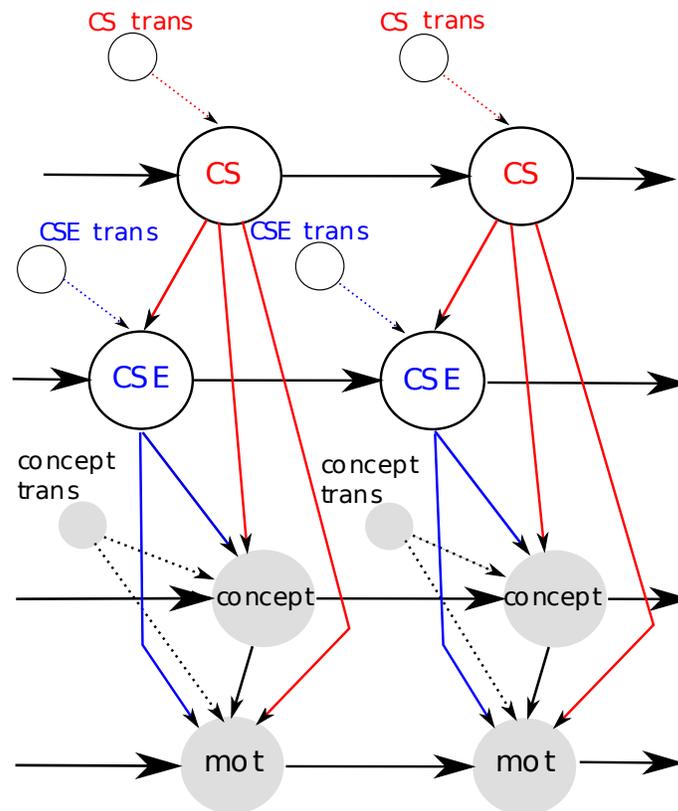


Figure 3. CS et CSE considérés comme deux variables non observées

4.2. Paramètres des modèles

Toutes les variables sont observées pendant l'entraînement du modèle. Ainsi, les tables de probabilités conditionnelles associées aux arcs sont directement obtenues à partir des observations.

Le calcul des probabilités de ces tables est réalisé grâce à des modèles de langage factorisés (MLF) utilisant des techniques de repli parallèle généralisé (RPG) (Bilmes *et al.*, 2003, Kirchhoff *et al.*, 2007). Les MLF sont une extension des modèles de langage classiques dans laquelle les prédictions sont basées sur un ensemble de caractéristiques et non plus seulement sur les précédentes occurrences de la variable. Le RPG permet d'étendre les procédures de repli standard au cas où des paramètres de différents types sont considérés, sans contrainte temporelle imposée (contrairement aux modèles de langage classiques, dans un MLF les paramètres peuvent apparaître au moment de la prédiction).

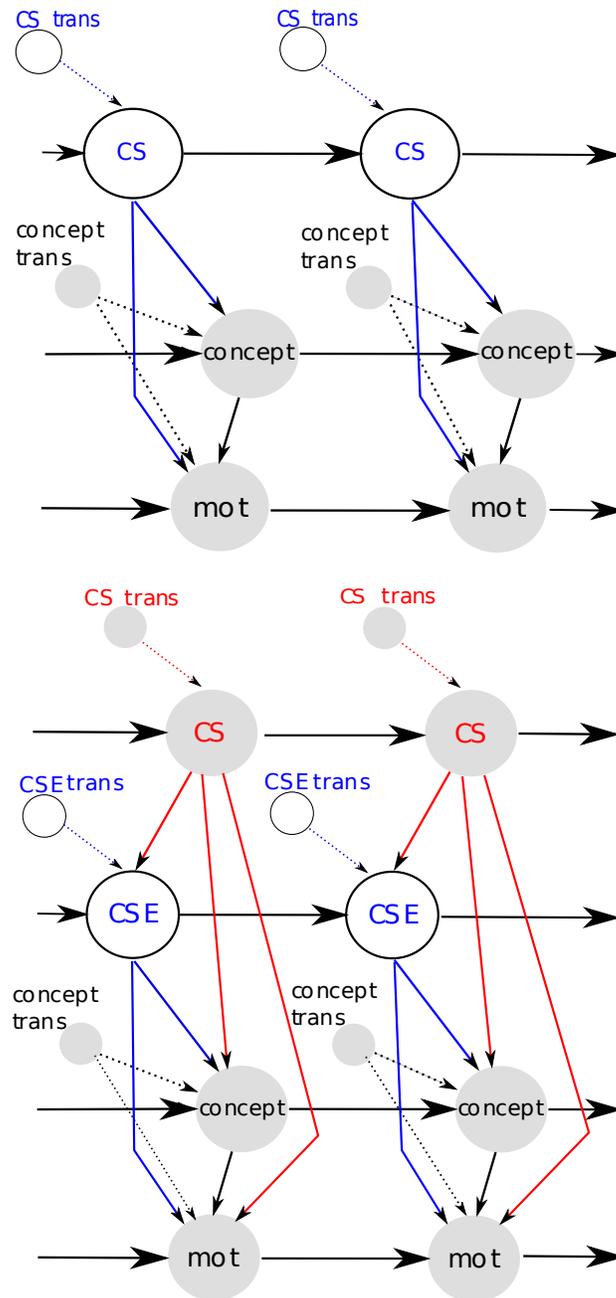


Figure 4. *Décodage à deux niveaux des CS et CSE*

Dans nos modèles de composition à base de RBD, plusieurs implémentations de MLF sont utilisées, correspondant aux arcs des graphes des figures 2, 3 et 4. Dans ces implémentations, h représente la longueur de l'historique ($h = -1$ pour un bigramme), $CSCSE$, CS , CSE , C et M représentent respectivement un CS/CSE (une seule variable), un CS, un CSE, un concept et un mot.

Les séquences de mots, de concepts et de transitions sont des variables observées pour le décodage des CS et des CSE : elles ont été décodées par les modules de reconnaissance de la parole et de compréhension. Étant donné la faible densité des données d'apprentissage, les probabilités conditionnelles utilisées dans les modèles sont limitées à des MLF bigrammes ($h = -1$).

Les tables de probabilités conditionnelles utilisées par les modèles RBD sont produites par les implémentations des MLF suivantes :

- modèle *compact*, une seule variable CS/CSE (figure 2) :

$P(CSCSE) \simeq \prod P(cscse|cscse_h)$: séquences de CS/CSE ;

$P(C|CSCSE) \simeq \prod P(c|c_h, cscse)$, RPG dans l'ordre $\{c_h, cscse\}$:
séquences de concepts conditionnées par les CS/CSE ;

$P(M|C, CSCSE) \simeq \prod P(m|m_h, c, cscse)$, RPG dans l'ordre $\{m_h, c, cscse\}$:
séquences de mots conditionnées par les concepts et les CS/CSE.

- modèle *factorisé*, une variable CS, une variable CSE, décodage simultané (figure 3) :

$P(CS) \simeq \prod P(cs|cs_h)$: séquences de CS ;

$P(CSE|CS) \simeq \prod P(cse|cse_h, cs)$, RPG dans l'ordre $\{cse_h, cs\}$:
séquences de CSE conditionnées par les CS ;

$P(C|CSE, CS) \simeq \prod P(c|c_h, cse, cs)$, RPG dans l'ordre $\{c_h, cse, cs\}$:
séquences de concepts conditionnées par les CSE et les CS ;

$P(M|C, CSE, CS) \simeq \prod P(m|m_h, c, cse, cs)$, RPG dans l'ordre $\{m_h, c, cse, cs\}$:
séquences de mots conditionnées par les concepts, CSE et CS.

- modèle à *deux niveaux*, une variable CS, une variable CSE, décodage à deux niveaux (figure 4) :

– *Premier niveau* :

$P(CS) \simeq \prod P(cs|cs_h)$: séquences de CS ;

$P(C|CS) \simeq \prod P(c|c_h, cs)$, RPG dans l'ordre $\{c_h, cs\}$:
séquences de concepts conditionnées par les CS ;

$P(M|C, CS) \simeq \prod P(m|m_h, c, cs)$, RPG dans l'ordre $\{m_h, c, cs\}$:
séquences de mots conditionnées par les concepts et les CS.

– *Second niveau* :

$P(\hat{CS}) \simeq \prod P(\hat{cs}|\hat{cs}_h)$: séquences de CS observés ;

$P(CSE|\hat{CS}) \simeq \prod P(cse|cse_h, \hat{cs})$, RPG dans l'ordre $\{cse_h, \hat{cs}\}$:
séquences de CSE conditionnées par les CS observés ;

$P(C|\hat{CS}, CSE) \simeq \prod P(c|c_h, \hat{cs}, cse)$, RPG dans l'ordre $\{c_h, \hat{cs}, cse\}$:
séquences de concepts conditionnées par les CSE et les CS observés ;

$P(M|C, \hat{CS}, CSE) \simeq \prod P(m|m_h, c, \hat{cs}, cse)$, RPG dans l'ordre
 $\{m_h, c, \hat{cs}, cse\}$: séquences de mots conditionnées par les concepts,
CSE et CS.

où le chapeau indique les variables ayant des valeurs fixées.

4.3. Composition des CS

La représentation des CS et CSE étant hiérarchique, des situations de recouvrement peuvent se produire lors de la détermination des CS et CSE associés à un concept. Cela arrive principalement lorsque plusieurs CS ou CSE ont été déclenchés par le même concept mais également lorsque les processus d'inférence et de composition ont créé des structures de CS et CSE imbriquées reliées au même concept. Pour résoudre ce problème, nous avons développé et appliqué un algorithme de projection d'arbre dont le pseudo-code est présenté en Annexe 3.

La projection est réalisée sur l'annotation en CS et CSE de la phrase complète structurée en arbre. Elle permet de définir des sous-branches de l'arbre associées à un seul concept. Partant d'une feuille de l'arbre, une classe de CS/CSE est obtenue en agrégeant les valeurs de nœuds pères (CS ou CSE) aussi longtemps qu'ils sont associés au même concept (ou à aucun). Les arêtes des branches sont définies par les liens d'appartenance des CSE aux CS et par les liens de typage des CSE en CS (un CSE prenant un CS pour valeur). Par exemple, la séquence de mots " près du Festival de Cannes", proposée dans la figure 1, entraîne la création de la branche projetée `location_event-LOCATION-lodging_location-LODGING`.

Ensuite, pour le modèle dans lequel CS et CSE sont représentés par une unique variable, les branches sont conservées telles qu'obtenues et considérées directement comme des classes de CS et CSE associés. Pour les deux autres modèles, les CS et CSE de chaque branche sont considérés séparément pour produire deux sous-ensembles distincts de classes.

Les classes composées du premier modèle sont utilisées lors du décodage puis les CS et les CSE sont séparés. Durant cette dernière étape, les deux types de liens entre CS et CSE (appartenance et typage) peuvent être retrouvés. Cependant, la séquentialité du décodage entraîne la perte de certains liens de typage. Un ensemble de règles logiques manuellement définies est alors appliqué pour retrouver ces liens à partir de l'ensemble des branches décodées. Ces règles sont comparables à celles utilisées pour l'annotation semi-automatique des données d'apprentissage mais ne prennent pas en compte les mêmes informations. En effet, les informations disponibles sont différentes

de celles fournies par les données d'apprentissage : par exemple, après décodage, un CS n'est plus associé à un mot particulier mais à une unité conceptuelle. Les données d'apprentissage annotées en CS et CSE fournissent les ensembles de classes de CS et CSE à partir desquels les paramètres des réseaux bayésiens dynamiques sont estimés.

5. Expériences et résultats

Pour évaluer les performances des systèmes de composition des CS utilisant les RBD, un ensemble de données de test est défini. Quinze dialogues contenant au total 225 tours de parole utilisateur ont été manuellement annotés en CS et CSE par un expert. Le système d'annotation en deux étapes à base de règles (décrit en section 3) a été utilisé pour produire une annotation en CS et CSE sur le corpus MEDIA complet (transcriptions et annotation conceptuelle manuelles), les données de test étant exclues. La qualité de cette annotation a été évaluée sur les données de test : une F-mesure de 0,95 pour l'identification des CS confirme la fiabilité du système et la consistance des données d'apprentissage.

Les expériences sont menées sur l'ensemble de test dans trois conditions différentes, fonctions de la qualité des données initiales :

- référence (REF) : les tours de parole du locuteur sont manuellement transcrits et annotés en concepts et CS ;
- SLU : les concepts de base sont décodés à partir des transcriptions manuelles des tours de parole locuteur, en utilisant le modèle de compréhension (SLU) à base de RBD décrit dans (Lefèvre, 2006) ;
- ASR+SLU : les concepts sont décodés par le modèle de compréhension en utilisant la meilleure hypothèse (1-best) de séquence de mots générée par le système de reconnaissance automatique de la parole (ASR). Le système ASR utilisé est conforme à (Barrault *et al.*, 2008).

Les taux d'erreurs observés sur ces trois types de données sont présentés dans le tableau 5.

Type de données	REF	SLU	ASR + SLU
Taux d'erreurs mots (%)	0,0	0,0	14,8
Taux d'erreurs concepts (%)	0,0	10,6	24,3

Tableau 5. Taux d'erreurs en mots et en concepts observés sur les trois types de données

Pour indiquer le dimensionnement des modèles RBD, le nombre de mots, concepts et combinaisons de CS-CSE, CS et CSE distincts utilisés pour leur entraînement est donné dans le tableau 6.

Toutes les expériences présentées ont été réalisées avec GMTK (Bilmes *et al.*, 2002), outil logiciel de calcul et de manipulation de modèles graphiques, et

Modèle RBD	Mots	Concepts	Combin. CS-CSE	CS	CSE
compact	2201	78	636	x	x
factorisé	2201	78	x	234	339
2-niveaux	2201	78	x	234	339

Tableau 6. Cardinalités des variables de mots, concepts et des classes de combinaisons de CS-CSE, CS et CSE distinctes utilisées dans les 3 types de modèles RBD (compact, factorisé et 2-niveaux)

SRILM (Stolcke, 2002), outil logiciel pour les modèles de langage de type n -grammes.

Le tableau 7 regroupe les résultats des trois systèmes de composition en termes de précision, rappel et F-mesure. La précision est le nombre de CS, CSE ou liens corrects proposés par le système divisé par le nombre total de CS, CSE ou liens proposés par le système. Le rappel est le nombre de CS, CSE ou liens corrects proposés par le système divisé par le nombre total de CS, CSE ou liens contenus dans l'annotation de référence. La F-mesure est la moyenne harmonique standard de la précision et du rappel.

Les hypothèses de CS et CSE sont considérées comme correctes dès lors que les CS et CSE correspondants sont présents dans la référence : ni leurs liens ni l'ordre dans lequel ils apparaissent ne sont pris en compte. Lors de l'évaluation des CSE, seuls ceux associés à des CS corrects sont évalués. Enfin, les résultats obtenus sur les liens de typage entre CSE et CS sont présentés dans la dernière colonne du tableau 7.

Les résultats du tableau 7 montrent que les performances des trois systèmes à base de RBD sont très proches. Cependant, le système à deux niveaux peut être considéré comme le plus performant. En effet, au-delà des bons résultats qu'il produit, il utilise le modèle de plus faible complexité lors du décodage.

Les bons résultats obtenus sur les liens de typage confirment que les modèles à base de RBD, combinés à un ensemble de règles logiques de taille restreinte, peuvent être utilisés pour générer des structures sémantiques hiérarchiques consistantes. De plus, ces modèles ayant la capacité de produire des hypothèses avec leurs scores de confiance, ils peuvent être utilisés dans des contextes où les hypothèses sont multiples (réseaux de confusion, n -meilleures hypothèses) ou encore dans des protocoles d'évaluation en évaluant et classant les hypothèses proposées par d'autres systèmes.

6. Conclusion

Cet article présente un module stochastique d'analyse sémantique du dialogue oral. Ce module de compréhension s'appuie sur une représentation sémantique issue du formalisme FrameNet. Le choix de ce paradigme a été motivé par la capacité des cadres sémantiques à représenter des situations de négociation complexes et à s'adapter aux actions conséquentes du gestionnaire de dialogue.

Type de données		REF		
Modèles RBD		CS	CSE	Liens
Compact (CS/CSE)	\bar{p}	0.91	0.91	0.93
	\bar{r}	0.93	0.86	0.98
	$\bar{F}\text{-m}$	0.89	0.86	0.92
Factorisé (CS et CSE)	\bar{p}	0.92	0.92	0.94
	\bar{r}	0.92	0.85	0.98
	$\bar{F}\text{-m}$	0.90	0.86	0.94
Deux niveaux (CS puis CSE)	\bar{p}	0.92	0.91	0.92
	\bar{r}	0.94	0.82	0.98
	$\bar{F}\text{-m}$	0.91	0.83	0.93
SLU				
Compact (CS/CSE)	\bar{p}	0.87	0.91	0.93
	\bar{r}	0.82	0.83	0.98
	$\bar{F}\text{-m}$	0.81	0.82	0.92
Factorisé (CS et CSE)	\bar{p}	0.88	0.92	0.93
	\bar{r}	0.81	0.83	0.97
	$\bar{F}\text{-m}$	0.80	0.83	0.91
Deux niveaux (CS puis CSE)	\bar{p}	0.88	0.91	0.92
	\bar{r}	0.86	0.80	0.97
	$\bar{F}\text{-m}$	0.83	0.80	0.90
ASR + SLU				
Compact (CS/CSE)	\bar{p}	0.86	0.90	0.92
	\bar{r}	0.80	0.86	0.98
	$\bar{F}\text{-m}$	0.78	0.84	0.92
Factorisé (CS et CSE)	\bar{p}	0.87	0.90	0.94
	\bar{r}	0.79	0.86	0.97
	$\bar{F}\text{-m}$	0.78	0.84	0.93
Deux niveaux (CS puis CSE)	\bar{p}	0.87	0.89	0.93
	\bar{r}	0.81	0.82	0.98
	$\bar{F}\text{-m}$	0.79	0.80	0.92

Tableau 7. Précision (\bar{p}), Rappel (\bar{r}) et F-mesure ($\bar{F}\text{-m}$) sur les données de test MEDIA pour les trois systèmes de composition sémantique à base de RBD

Construite manuellement, la base de connaissances en cadres sémantiques (et leurs éléments) a été conçue pour pouvoir s'adapter au domaine du corpus d'expérimentation MEDIA tout en conservant la plus grande généralité possible. Sa structuration a été pensée pour permettre une représentation sémantique arborescente des messages de l'utilisateur du système de dialogue.

L'utilisation de modèles stochastiques dans un cadre applicatif nécessite la connaissance des distributions de probabilités conditionnelles liant les données auxquelles on s'intéresse. Dans notre cas, ces données sont les mots, les concepts de base, les CS et les CSE associés à chaque message du locuteur. Les distributions conditionnelles décrivant le conditionnement mutuel des mots, concepts, CS et CSE, peuvent être apprises sur des ensembles de données observées.

Le corpus d'expérimentation MEDIA n'étant pas annoté en CS et CSE, un système à base de règles a été développé pour permettre l'annotation initiale des données d'apprentissage. Ce système crée tout d'abord CS et CSE par reconnaissance de patrons lexicaux et conceptuels puis relie ces objets sémantiques lors d'une étape d'inférence logique. La qualité de l'annotation produite a été évaluée en la comparant à une annotation manuelle des données de test. Les résultats confirment la fiabilité du système et la consistance des données d'apprentissage.

Les modèles stochastiques que nous avons utilisés sont des réseaux bayésiens dynamiques (RBD). Ce sont des modèles d'une grande flexibilité permettant de représenter des systèmes stochastiques complexes. Trois modèles de structures et de complexités différentes sont évalués dans ce travail. L'apprentissage des paramètres de ces modèles s'appuie sur un algorithme de décomposition des arbres sémantiques en branches conceptuelles. L'algorithme est appliqué à l'annotation en CS et CSE de la phrase complète. Cette annotation étant structurée en arbre, l'algorithme permet de définir des sous-branches de l'arbre associées à un seul concept.

Les trois modèles présentés permettent de déduire des annotations sémantiques hiérarchiques en cadres sémantiques à partir des séquences de mots et de concepts extraites de la requête utilisateur. Les résultats expérimentaux tendent à montrer que les performances des modèles à base de RBD sont suffisamment bonnes pour permettre leur utilisation dans un système de dialogue. Ils pourront alors fournir au gestionnaire de dialogue une représentation sémantiquement riche et entièrement structurée de la requête de l'utilisateur. Bien que cet objectif puisse être également atteint par un système à base de règles, seuls les modèles à base de réseaux bayésiens dynamiques sont capables de produire des listes de n -meilleures hypothèses d'arbres sémantiques avec des scores de confiance associés, permettant au module de gestion du dialogue de choisir la meilleure interprétation parmi n en ayant recours à des informations contextuelles et liées au déroulement du dialogue dans son ensemble.

En outre, le module de compréhension proposé repose sur un système très ouvert dans sa définition. Il n'utilise pas de grammaire manuelle et peut être adapté au traitement de tout type de dialogue oral. Une fois l'annotation d'un corpus de référence réalisée, l'apprentissage des modèles est entièrement automatique. La base de connaissances sur laquelle repose la représentation sémantique est facilement modifiable pour s'adapter aux besoins de domaines variés. L'emploi des structures récursives de CS et CSE contribue à la richesse de cette représentation.

De même, les modèles stochastiques à base de RBD employés pour générer la représentation sémantique des messages n'imposent aucune contrainte quant à la na-

ture des informations observées. Il est aisé d'introduire de nouvelles variables dans ces modèles et donc de prendre en compte, sous réserve de calculabilité, de nouvelles informations pertinentes pour la compréhension (les informations liées au contexte de dialogue par exemple).

Dans la continuité directe des travaux présentés, nous avons pour objectif actuel d'évaluer en pratique le gain lié à l'utilisation d'une liste d'hypothèses d'interprétation au lieu d'une hypothèse unique. A chaque hypothèse sera associée une mesure de confiance tenant compte de l'ensemble des informations utilisées pour l'engendrer (par exemple le décodage acoustico-lexical). Parallèlement, nous travaillons à la prise en compte du contexte de dialogue lors de l'élaboration de l'interprétation sémantique. Enfin, une perspective importante de ce travail, en cours de réalisation, est l'intégration de notre module de compréhension dans un système de dialogue complet fonctionnel.

7. Bibliographie

- Barrault L., Servan C., Matrouf D., Linarès G., de Mori R., « Frame-Based Acoustic Feature Integration for Speech Understanding », *IEEE International Conference on Audio, Signal and Speech Processing*, Las Vegas, 2008.
- Bilmes J., Kirchoff K., « Factored language models and generalized parallel backoff », *NAACL Human Language Technology*, Edmonton, Canada, 2003.
- Bilmes J., Zweig G., « The graphical models toolkit : An open source software system for speech and time-series processing », *IEEE International Conference on Audio, Signal and Speech Processing*, Orlando, Florida, 2002.
- Bonneau-Maynard H., Lefèvre F., « A 2+1-level stochastic understanding model », *IEEE Workshop on Automatic Speech Recognition and Understanding*, San-Juan, Puerto-Rico, 2005a.
- Bonneau-Maynard H., Rosset S., Ayache C., Kuhn A., Mostefa D., the Media consortium, « Semantic annotation of the MEDIA corpus for spoken dialog », *ISCA Eurospeech*, Lisboa, Portugal, 2005b.
- Colmerauer A., Roussel P., « The birth of Prolog », *HOPL-II : The second ACM SIGPLAN conference on History of programming languages*, 1993.
- Damnati G., Béchet F., de Mori R., « Spoken Language Understanding strategies on the France Telecom 3000 Voice Agency corpus », *IEEE ICASSP*, 2007.
- Fillmore C., Johnson C., Petruck M., « Background to FrameNet », *International Journal of Lexicography*, vol. 16.3, p. 235-250, 2003.
- Gorin A. L., Riccardi G., Wright J. H., « How may I help you ? », *Speech Communication*, vol. 23, n° 1-2, p. 113-127, 1997.
- He Y., Young S., « Spoken Language Understanding using the Hidden Vector State Model », *Speech Communication*, vol. 48, n°3-4, p. 262-275, 2005.
- Hockey B. A., Miller D. P., « A Demonstration of a Conversationally Guided Smart Wheelchair », *Assets '07 : Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, October, 2007.
- Jackendoff R., « Semantic Structures », *The MIT Press, Cambridge Mass.*, 1990.

- Kirchhoff K., Bilmes J., Duh K., Factored Language Models Tutorial, Technical report, Dept. of EE, U. Washington, 2007.
- Lamel L., Rosset S., Gauvain J.-L., Bennacef S., Garnier-Rizet M., Prouts B., « The LIMSI ARISE System », *Speech Communication*, vol. 31, n°4, p. 339-354, 2000.
- Lefèvre F., « A DBN-based multi-level stochastic spoken language understanding system », *IEEE/ACL Workshop on Spoken Language Technology*, Aruba, 2006.
- Lefèvre F., « Dynamic Bayesian Networks and Discriminative Classifiers for Multi-Stage Semantic Interpretation », *IEEE International Conference on Audio, Signal and Speech Processing*, Honolulu, Hawaii, 2007.
- Levin E., Pieraccini R., « Concept-based Spontaneous Speech Understanding System », *ESCA Eurospeech*, Madrid, 1995.
- Litman D. J., Silliman S., « ITSPOKE : An Intelligent Tutoring Spoken Dialogue System », *HLT-NAACL*, 2004.
- Lowe J., Baker C., Fillmore C., « A frame-semantic approach to semantic annotation », *SIGLEX Workshop : Why, What, and How ?*, 1997.
- Meurs M.-J., Duvert F., Béchet F., Lefèvre F., de Mori R., « Semantic Frame Annotation on the French MEDIA corpus », *Language Resources and Evaluation Conference*, Marrakech, Morocco, 2008a.
- Meurs M.-J., Lefèvre F., de Mori R., « A Bayesian Approach to Semantic Composition for Spoken Language Interpretation », *ISCA Interspeech*, Brisbane, Australia, 2008b.
- Meurs M.-J., Lefèvre F., de Mori R., « Spoken Language Interpretation : On the Use of Dynamic Bayesian Networks for Semantic Composition », *IEEE International Conference on Audio, Signal and Speech Processing*, Taipei, Taiwan, 2009.
- Pado S., Pitel G., « Annotation précise du français en sémantique de rôles par projection cross-linguistique », *Conférence TALN*, Toulouse, France, 2007.
- Petruck M., « Frame semantics », *Handbook of Pragmatics*, 1996.
- Pon-Barry H., Schultz K., Owen Bratt E., Clark B., Peters S., « Responding to Student Uncertainty in Spoken Tutorial Dialogue Systems », *International Journal of Artificial Intelligence in Education*, vol. Volume 16, Number 2/2006, p. 171-194, 2006.
- Price P., « Evaluation of Spoken Language Systems : the ATIS Domain », *DARPA Workshop on Speech and Natural Language*, 1990.
- Raux A., Langner B., Black A. W., Eskenazi M., « LET'S GO : Improving Spoken Dialog Systems for the Elderly and Non-Natives », *EUROSPEECH - INTERSPEECH*, 2003.
- Raymond C., Béchet F., de Mori R., Damnati G., « On the use of finite state transducers for semantic interpretation », *Speech Communication*, vol. 48 :3-4, p. 288-304, 2006.
- Sadek M. D., Bretier P., Cadoret V., Cozannet A., Dupont P., Ferrieux A., Panaget F., « A Cooperative Spoken Dialogue System Based on a Rational Agent Model : A First Implementation on the AGS », *ESCA Workshop on Spoken Dialogue Systems*, 1995.
- Stolcke A., « SRILM an extensible language modeling toolkit », *IEEE International Conference on Audio, Signal and Speech Processing*, 2002.
- Wielemaker J., « An overview of the SWI-Prolog Programming Environment », *International Workshop on Logic Programming Environments*, 2003.
- Woods W., *What's in a Link : Foundations for Semantic Networks*, Bolt and Beranek and Newman, 1975.

Annexe 1

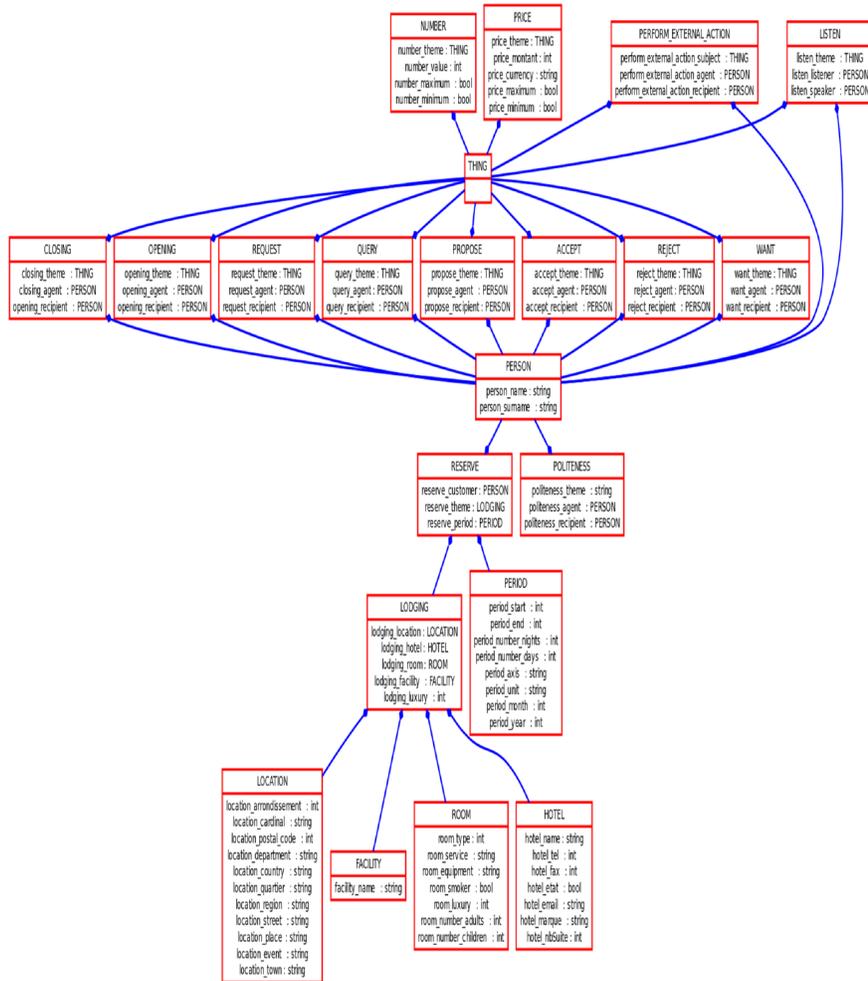


Figure 5. Base de connaissances associée au corpus MEDIA

Annexe 2

Un extrait de corpus annoté est donné ci-dessous :

```

<turn id="209_12_spk start="144" end="155" speaker="spk" audio="08_209.wav">
  <semAnnotation withContext="true" origin="ELDA" manual="true" tool="semantizer">
    <sem id="262" mode="+" concept="reponse" specif="" value="oui">
      <transcription origin="ELDA" manual="true" tool="transcriber">
        <Sync time="144"/>
        <Event desc="bb" extent="instantaneous" type="noise"/>
        oui oui
      </transcription>
    </sem>
    <sem id="263" mode="+" concept="command-tâche" specif="" value="reservation">
      <transcription origin="ELDA" manual="true" tool="transcriber">
        je souhaite réserver
        <Sync time="148"/>
        <Event desc="." extent="instantaneous" type="noise"/>
      </transcription>
    </sem>
  </semAnnotation>
  <frameAnnotation>
    <frame fname="ACCEPT" id="F_01" semid="262"/>
    <frame fname="PERSON" id="F_02" semid="263">
      <frlmt fename="person_name" id="FE_01" frid="F_02" semid="263" />
    </frame>
    <frame fname="RESERVE" id="F_03" semid="263">
      <frlmt fename="res_customer" id="FE_02" frid="F_03" semid="263 subfrid="F_02"/>
    </frame>
    <frame fname="WANT" id="F_04" semid="263">
      <frlmt fename="want_theme" id="FE_03" frid="F_04" semid="263 " subfrid="F_03"/>
      <frlmt fename="want_agent" id="FE_04" frid="F_04" semid="263 " subfrid="F_02"/>
    </frame>
  </frameAnnotation>
</turn>

```

Annexe 3

Algorithm 1

Algorithme de projection d'arbres

Entrée : $\{c_i\}$ séquence de concepts associés à la phrase, \mathcal{T} arbre de CS et CSE représentant la phrase

Sortie : \mathcal{B} ensemble des branches

```

1:  $\mathcal{B} \leftarrow \emptyset$ 
2: pour tout  $c \in \{c_i\}$  faire
3:    $\text{branche } b_c \leftarrow \emptyset$ 
   Génération des branches principales
4:   pour tout  $l \in \text{feuilles}(\mathcal{T}, c)$  faire
5:      $b_c.\text{ajouter}(\text{extraire\_branche}(l))$ 
6:   fin pour
   Contrôle des branches internes
7:   pour tout  $n \in \text{noeuds}(\mathcal{T}, c)$  faire
8:     si  $n \notin b_c$  alors
9:        $b_c.\text{ajouter}(\text{extraire\_branche}(n))$ 
10:    fin si
11:   fin pour
12:    $\mathcal{B} \leftarrow b_c$ 
13: fin pour
14: retourner  $\mathcal{B}$ 

```

Fonction extraire_branche

Entrée : c_i concept, n noeud

Sortie : branche $b \leftarrow \emptyset$

```

15: répéter
16:    $b. = n$ 
17:    $n \leftarrow n.\text{père}()$ 
18: jusqu'à  $!(n \text{ ET } (n.\text{concept} \in \{c_i, \text{null}\}))$ 
19: retourner  $b$ 

```

end

ANNEXE POUR LE SERVICE FABRICATION
A FOURNIR PAR LES AUTEURS AVEC UN EXEMPLAIRE PAPIER
DE LEUR ARTICLE ET LE COPYRIGHT SIGNE PAR COURRIER
LE FICHER PDF CORRESPONDANT SERA ENVOYE PAR E-MAIL

1. ARTICLE POUR LA REVUE :
Author version - 2010
2. AUTEURS :
Marie-Jean Meurs, Fabrice Lefèvre et Renato de Mori
3. TITRE DE L'ARTICLE :
Approche bayésienne de la composition sémantique dans les systèmes de dialogue oral
4. TITRE ABRÉGÉ POUR LE HAUT DE PAGE MOINS DE 40 SIGNES :
Composition sémantique bayésienne
5. DATE DE CETTE VERSION :
13 mai 2013
6. COORDONNÉES DES AUTEURS :
 - adresse postale :
Université d'Avignon et des Pays de Vaucluse
Laboratoire Informatique d'Avignon (EA 931), F-84911 Avignon, France.
{marie-jean.meurs,fabrice.lefevre,renato.demori}@univ-avignon.fr
 - téléphone :
 - télécopie :
 - e-mail :
7. LOGICIEL UTILISÉ POUR LA PRÉPARATION DE CET ARTICLE :
L^AT_EX, avec le fichier de style `article-hermes.cls`,
version 1.2 du 03/03/2005.
8. FORMULAIRE DE COPYRIGHT :
Retourner le formulaire de copyright signé par les auteurs, téléchargé sur :
<http://www.revuesonline.com>

SERVICE ÉDITORIAL – HERMES-LAVOISIER
14 rue de Provigny, F-94236 Cachan cedex
Tél : 01-47-40-67-67
E-mail : revues@lavoisier.fr
Serveur web : <http://www.revuesonline.com>