

Annotation en Frames Sémantiques du corpus de dialogue MEDIA

Marie-Jean Meurs Frédéric Duvert Frédéric Béchet Fabrice Lefèvre
Renato de Mori

Université d'Avignon et des Pays de Vaucluse

Laboratoire Informatique d'Avignon (EA 931), F-84911 Avignon, France.

{marie-jean.meurs, frederic.duvert, frederic.bechet, fabrice.lefevre,
renato.demori}@univ-avignon.fr

Résumé. Cet article présente un formalisme de représentation des connaissances qui a été utilisé pour fournir des annotations sémantiques de haut niveau pour le corpus de dialogue oral MEDIA. Ces annotations en structures sémantiques, basées sur le paradigme FrameNet, sont obtenues de manière incrémentale et partiellement automatisée. Nous décrivons le processus d'interprétation automatique qui permet d'obtenir des compositions sémantiques et de générer des hypothèses de frames par inférence. Le corpus MEDIA est un corpus de dialogues en langue française dont les tours de parole de l'utilisateur ont été manuellement transcrits et annotés (niveaux mots et constituants sémantiques de base). Le processus proposé utilise ces niveaux pour produire une annotation de haut niveau en frames sémantiques. La base de connaissances développée (définitions des frames et règles de composition) est présentée, ainsi que les résultats de l'annotation automatique.

Abstract. This paper introduces a knowledge representation formalism, used for incremental and partially automated annotation of the French MEDIA dialogue corpus in terms of semantic structures. We describe an automatic interpretation process for composing semantic structures from basic semantic constituents using patterns involving constituents and words. The process has procedures for obtaining semantic compositions and generating frame hypotheses by inference. This process is applied to MEDIA, a dialogue corpus manually annotated at the word and semantic constituent levels, and thus produces a higher level semantic frame annotation. The Knowledge Source defined and the results obtained on the automatically-derived annotation are reported.

Mots-clés : compréhension automatique de la parole, système de dialogue oral, frames sémantiques, décodage conceptuel, annotation sémantique, inférence sémantique.

Keywords: spoken language understanding, spoken dialogue system, semantic structures, semantic frames, conceptual decoding, semantic annotation, semantic inference.

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, recherche d'horaires par exemple) 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 il requiert une représentation sémantique de haut niveau. De tels systèmes doivent disposer de connaissances sémantiques pouvant interagir avec les 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. La difficulté à obtenir une telle représentation de haut niveau fonctionnelle et robuste explique que les systèmes de dialogue actuels soient limités à des espaces sémantiques assez restreints.

Obtenir 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" passe donc par 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 ainsi que la gestion des références (phénomène dont l'apparition est fortement augmentée lors des phases de négociation). Dans cette perspective, le choix du paradigme FrameNet¹ comme représentation sémantique de haut niveau paraît approprié. Outre la robustesse du paradigme, FrameNet présente l'avantage de produire des annotations standards qu'il est facile de partager et de comparer au sein de la communauté scientifique.

Les travaux présentés dans cet article portent sur l'annotation en frames² du corpus de dialogue en français MEDIA. Le corpus est décrit dans la partie 2. Puis, la partie 3 détaille les principes utilisés pour l'annotation en frames. Enfin, l'application de ces principes au corpus MEDIA est présentée dans la partie 4.

2 Description des données MEDIA

Le corpus MEDIA (Bonneau-Maynard *et al.*, 2005) est un corpus de dialogue en français issu de la simulation d'un serveur téléphonique d'informations touristiques et de réservation d'hôtels. Il a été enregistré selon le protocole du *Magicien d'Oz* (système de dialogue simulé par un opérateur humain). Huit catégories de scénarii sont définies avec différents niveaux de complexité. Le corpus contient 1250 dialogues produits par 250 locuteurs, pour une durée totale d'environ 70 heures d'enregistrement audio. Chaque locuteur a enregistré cinq scénarii de réservation d'hôtel.

Le corpus est transcrit manuellement et enrichi par une annotation également manuelle utilisant 83 concepts de base rassemblés dans un dictionnaire sémantique de concepts. Cette annotation manuelle a été réalisée par deux annotateurs de la société ELDA. L'accord inter-annotateur (IAg) obtenu est toujours supérieur à 80% et atteint jusqu'à 90% dans le meilleur cas (Bonneau-Maynard *et al.*, 2005).

¹ Plus d'informations sur <http://framenet.icsi.berkeley.edu>

² Le terme *frame* a été conservé dans l'article, plutôt que sa traduction en *cadre sémantique*, afin d'éviter toute confusion.

Le dictionnaire sémantique utilisé pour l’annotation du corpus MEDIA associe à un mot ou un groupe de mots une paire *concept-valeur* puis un spécifieur définissant des relations entre concepts et enfin un *mode* affirmatif, négatif, interrogatif ou éventuel, attaché au concept. En définissant un ensemble de 19 spécifieurs pouvant être associés aux concepts de base, le schéma d’annotation MEDIA représente les relations entre les concepts et permet ainsi de construire une représentation hiérarchique de l’interprétation d’un énoncé.

Un exemple de message annoté du corpus MEDIA est donné dans le tableau (1). La première colonne contient les groupes de mots W^c supports de chaque concept, présenté 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 groupes de mots W^c .

W^c	<i>concept c</i>	<i>mode</i>	<i>specifieur</i>	<i>valeur</i>
<i>Je voudrais réserver</i>	commande	+ (affirmatif)		reservation
<i>une chambre</i>	chambre-quantite	+	reservation	1
<i>pour deux nuits</i>	sejour-nbNuit	+	reservation	2
<i>à Marseille</i>	localisation-ville	+	hotel	Marseille

TAB. 1 – Exemple d’annotation sémantique MEDIA

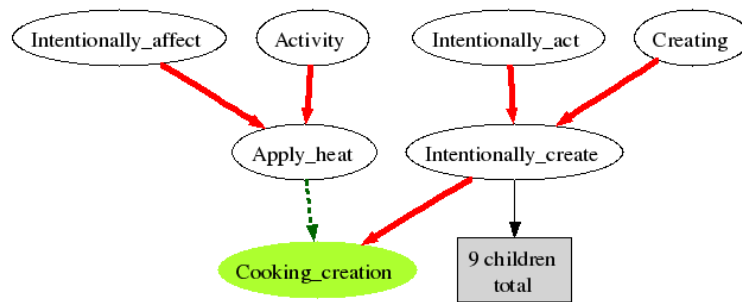
Dans cet exemple, le spécifieur *reservation* est lié aux concepts *chambre-quantite* et *sejour-nbNuit*. On obtient ainsi une structure hiérarchique représentant une réservation associée au concept *commande* et développée grâce aux valeurs des concepts *chambre-quantite* et *sejour-nbNuit*. Le spécifieur *hotel* 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 donc de recomposer une représentation hiérarchique de la requête de l’utilisateur à partir de l’annotation à plat.

3 Annotation en frames

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/reliations ou (Jackendoff, 1990) avec les structures de type fonction/argument. Une façon efficace de modéliser et de manipuler les connaissances sémantiques est de les représenter comme un ensemble de formules logiques sur lequel repose le processus de compréhension. Dans ce contexte, Fillmore définit les frames comme des structures cognitives associées au processus de compréhension (Fillmore, 1982; Fillmore, 1985). Chaque frame comporte des Frame Elements (FEs), rôles sémantiques qui lui sont propres. Une frame est donc un modèle représentant des entités sémantiques et leurs propriétés (Petruck, 1996).

Le projet FrameNet (Lowe *et al.*, 1997; Fillmore & Petruck, 2003) fournit pour la langue anglaise une base de données de frames détaillant leurs relations, leurs FEs et les unités lexicales (*Lexical Units*, LU) qui leurs correspondent ainsi que de nombreux exemples de textes annotés. L’exemple de la frame *cooking_creation* définie dans le dictionnaire FrameNet est donné dans la figure (1) qui illustre les relations de cette frame aux autres frames du dictionnaire.

FIG. 1 – La frame `cooking_creation` de FrameNet

3.1 Définition des frames

Le choix d’une annotation en frames dans ce travail est motivé par sa capacité à représenter des dialogues de négociation et à s’adapter aux actions complexes du gestionnaire de dialogue. Nous avons manuellement définie notre propre source de connaissance (KS) de frames pour décrire nos connaissances en termes de composition sémantique adaptée au domaine du corpus MEDIA. Les frames et les FEs sont décrits par un ensemble de modèles définis manuellement. Ces modèles sont composés d’unités lexicales (LUs), d’unités conceptuelles (CUs) et peuvent inclure des associations caractéristiques de ces composants.

Le dictionnaire FrameNet est disponible pour la langue anglaise mais il n’existe pas actuellement de FrameNet français comparable, capable de couvrir le domaine sémantique du corpus MEDIA. La seule ressource disponible pour le français, obtenue par projection automatique du FrameNet anglais (Pado & Pitel, 2007), a été écartée au profit d’une définition manuelle des frames pour plusieurs motifs :

- l’usage d’une KS de taille plus réduite mais couvrant parfaitement le domaine du corpus MEDIA limite la confusion lors de l’annotation ;
- la nature particulière du support textuel (transcriptions de parole) induit des différences importantes dans la définition des LU et CU ;
- les frames de FrameNet risquent d’être trop génériques pour satisfaire les besoins d’un système de dialogue automatique.

La définition de frames dédiées au domaine du corpus MEDIA a donc été privilégiée dans un premier temps. Toutefois, une attention particulière a été apportée afin de maintenir la définition des frames la plus indépendante possible de l’application. Ainsi certaines frames décrivant des connaissances générales comme les relations spatiales sont conformes à leur version FrameNet, quand d’autres plus spécifiques à l’application ont été créés ou largement modifiées. Toutefois la plupart des frames définies restent en correspondance avec des frames présentes dans le projet FrameNet. La figure (2) montre un extrait de ces correspondances.

Les mots déclenchant l’instanciation d’une frame sont les LUs associées à la frame ou à ses FEs. On distingue deux types de LUs : générique ou spécifique. Une LU est générique lorsqu’elle indique le sujet global d’une proposition. Une LU spécifique est un argument possible d’un prédicat. Un exemple de contenu des catégories de LUs est donné dans le tableau (2). Les valeurs prises par les LUs peuvent être des mots et leurs synonymes, des n-grammes ou des n-uplets de mots apparaissant dans des portions de dialogues, sans aucune contrainte sur leur position dans les phrases. Dans la continuité des annotations MEDIA, les CUs sont associés aux LUs. Ces associations sont utilisées dans le processus d’inférence décrit dans la partie suivante de cet article.

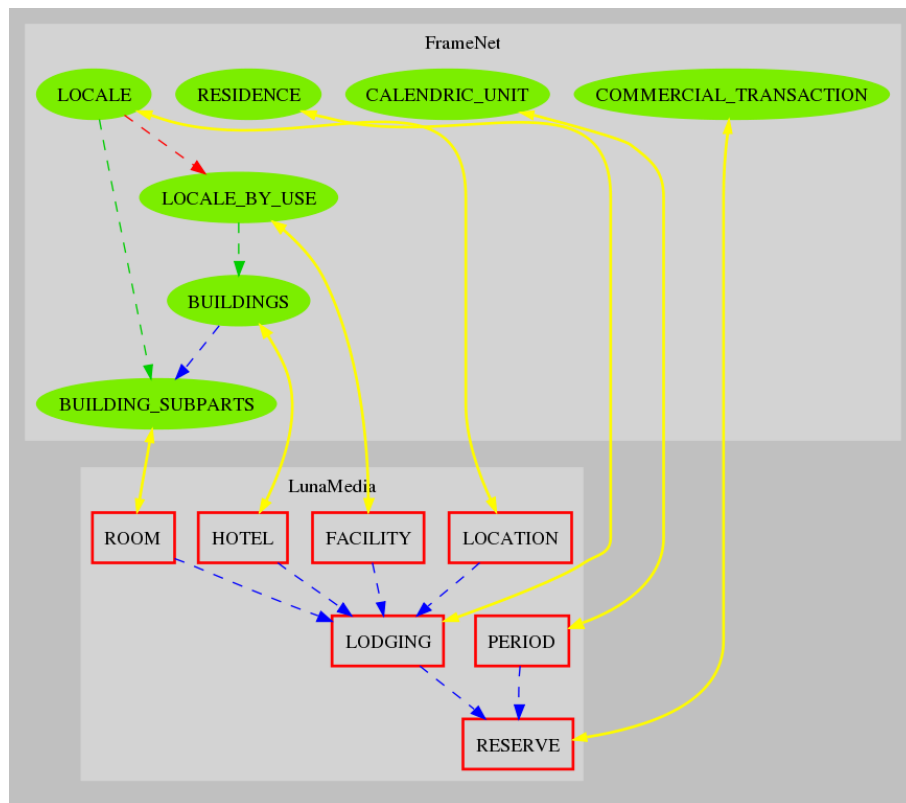


FIG. 2 – Exemples de correspondances entre les frames de la KS associée au corpus MEDIA et les frames FrameNet

<framelement fename="location_town">	
<generic_lexical_units	valeur="ville,cité,village"/>
<specific_lexical_units	valeur="paris,lyon,marseille..."/>
</framelement>	

TAB. 2 – Catégories de LUs pour le Frame Élément **location_town**

3.2 Règles d'inférence

La stratégie d'annotation se compose en deux étapes : une première étape reposant sur l'instanciation de frames à partir de la définition des frames et une seconde étape de composition des frames exécutée par des règles d'inférence. Le processus d'inférence appliqué aux instances de frames est décrit par des formules de la logique du premier ordre.

Les frames et FEs obtenus par le processus d'instanciation initial conditionnent les valeurs de vérité des règles logiques. Selon ces valeurs de vérité, frames et FEs sont instanciés, supprimés, modifiés ou reliés. Par exemple, si les frames RESERVE, HOTEL et LOCATION satisfont la règle présentée ci-après, une frame LODGING est créée, liée directement aux frames HOTEL et LOCATION ainsi qu'à la frame RESERVE par l'intermédiaire de la frame LODGING.

```
[ Frame RESERVE (FE_reservation_theme, concept = reservation_hotel)
^ Frame HOTEL (FE_hotel_name, concept = hotel)
^ Frame LOCATION (FE_location_town, concept = location_hotel) ]
=>
[ Frame LODGING (FE_lodging_hotel = Frame HOTEL,
                  FE_lodging_location = Frame LOCATION)
^ Frame RESERVER (FE_reservation_theme = Frame LODGING) ]
```

Environ trente règles d'inférence sont utilisées actuellement. Elles consistent principalement à proposer des relations entre les objets (frames, FEs) précédemment instanciés, à éliminer les redondances et à instancier des frames abstraites non découvertes pendant la première étape d'annotation reposant sur les définitions de frames. Ces règles utilisent les annotations fournies par l'étape de définition ainsi que les CUs et leurs valeurs présentes dans le corpus. Elles ne sont pas directement dépendantes des mots employés et s'affranchissent ainsi de la séquentialité de la proposition.

4 Application

La KS (Figure 3) associée au corpus MEDIA est composée de 21 frames de base pour un total de 86 FEs. Cette KS est de taille très modeste en comparaison des FrameNet français (Pado & Pitel, 2007) ou anglais. Le tableau (3) indique les dimensions des trois ressources, MEDIA KS, FrameNet français et FrameNet. A titre de comparaison, la frame REQUEST, commune à ces trois ressources, est présentée avec ses FEs selon les différentes versions.

	MEDIA KS	FrameNet français		FrameNet	
Frames	21	138		> 825	
FEs	86	1371		6800	
Frame	REQUEST	REQUEST		REQUEST	
FEs associés	<i>FE_agent</i>	<i>FE_speaker</i>	<i>FE_medium</i>	<i>FE_speaker</i>	<i>FE_medium</i>
	<i>FE_recipient</i>	<i>FE_addresse</i>	<i>FE_manner</i>	<i>FE_addresse</i>	<i>FE_manner</i>
	<i>FE_theme</i>	<i>FE_message</i>	<i>FE_means</i>	<i>FE_message</i>	<i>FE_means</i>
		<i>FE_topic</i>		<i>FE_topic</i>	<i>FE_beneficiary</i>
				<i>FE_time</i>	

TAB. 3 – Comparaison MEDIA KS - FrameNet français - FrameNet

Les annotations sont rassemblées dans des documents au format XML contenant des informations additionnelles telles que les références temporelles des segments de dialogues associés aux frames et aux FEs .

Les annotations manuelles des dialogues ont été produites grâce à un outil développé spécifiquement et qui permet d'obtenir un temps moyen d'annotation par tour de parole d'environ 4 à 5 minutes. Un annotateur a annoté manuellement quinze dialogues, soit 463 tours de parole.

L'extraction automatique des composants (LU, CU) présents dans le corpus permet de s'assurer de la couverture complète du domaine par les modèles. Ces composants sont rassemblés dans un fichier XML de modèles. Le corpus MEDIA est conceptuellement annoté initialement à l'aide de 83 concepts (dont 60 sont repris comme CUs dans la KS). Des CU additionnelles viennent les compléter pour permettre l'annotation d'entités sémantiques non initialement annotées dans

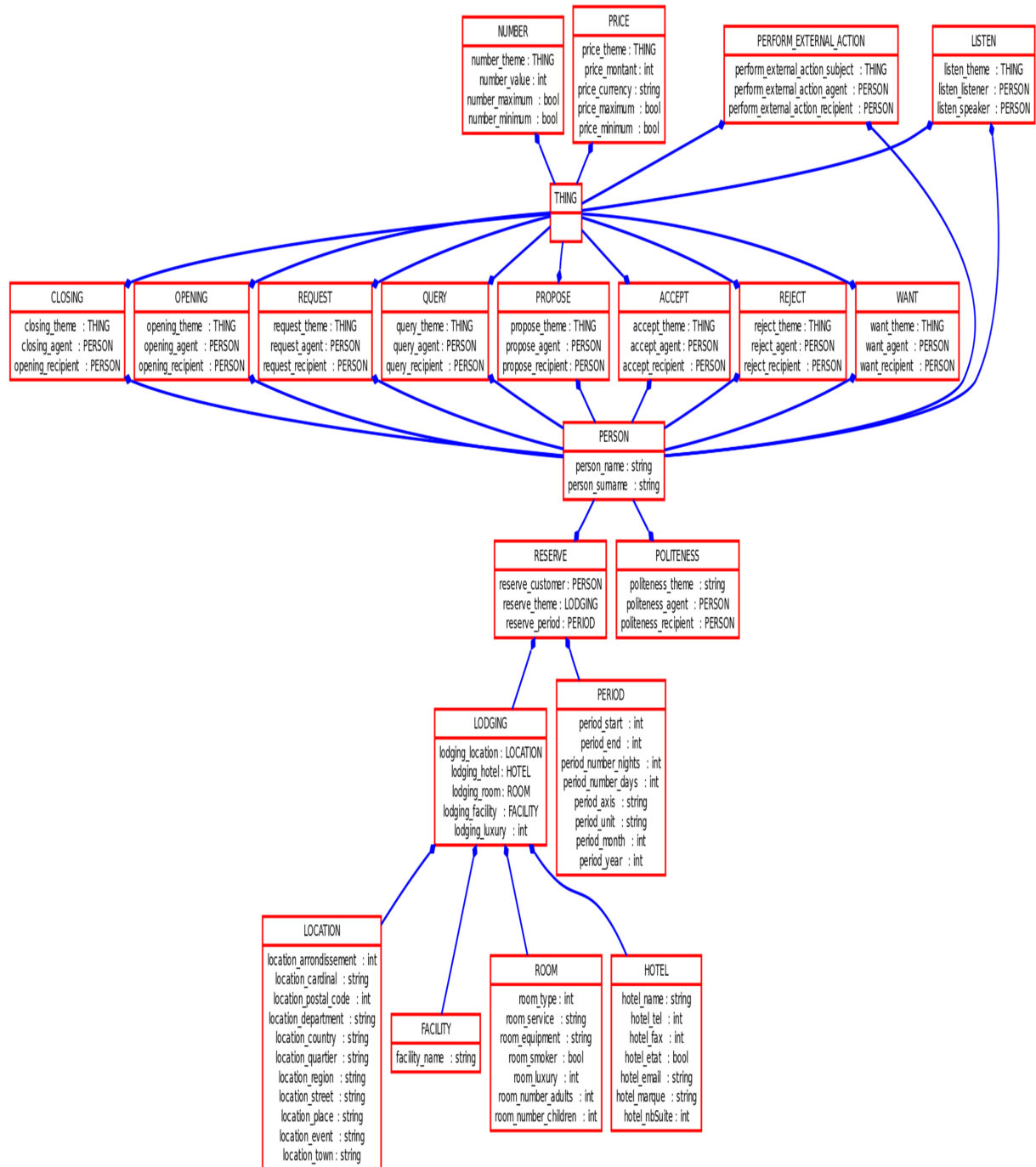


FIG. 3 – KS associée au corpus MEDIA

le corpus. Par exemple, les CUs associées à la frame PERSON, représentant un agent (Tableau 4), ne sont pas initialement annotées dans le corpus MEDIA. La KS contient actuellement 106 CUs et environ 1100 LUs.

Un outil de visualisation des frames, dédié à la manipulation de dialogues, apporte une aide précieuse à l'annotation manuelle, la vérification et la correction des annotations. Cet outil fournit pour chaque tour de parole une vue d'ensemble des frames et FEs instanciés ainsi que des relations entre ces objets. Il permet un accès direct à un tour de parole sélectionné, diffuse le fichier audio correspondant et est capable d'interagir avec l'outil d'annotation pour faciliter les corrections manuelles.

```

<frame      fname="PERSON">
  <concept valeur = "person"/>
  <frlmt fename="person_name">
    <concept valeur = "identifier"/>
  </frlmt>
  <frlmt fename="person_surname">
    <concept valeur = "surname"/>
  </frlmt>
</frame>

```

TAB. 4 – CUs associées à la frame PERSON

La figure (4) donne un exemple de visualisation des frames associées à un tour de parole dans lequel le client dit : *“alors j’aurais souhaité réserver euh deux chambres individuelles euh dans un hôtel à Orange pour vingt nuits du douze juillet au trente et un juillet euh je souhaiterais que le prix soit inférieur à cent euros ou alors un peu plus s’il y a une piscine euh je souhaiterais que dans un cadre très calme et avec une piscine donc si possible”*

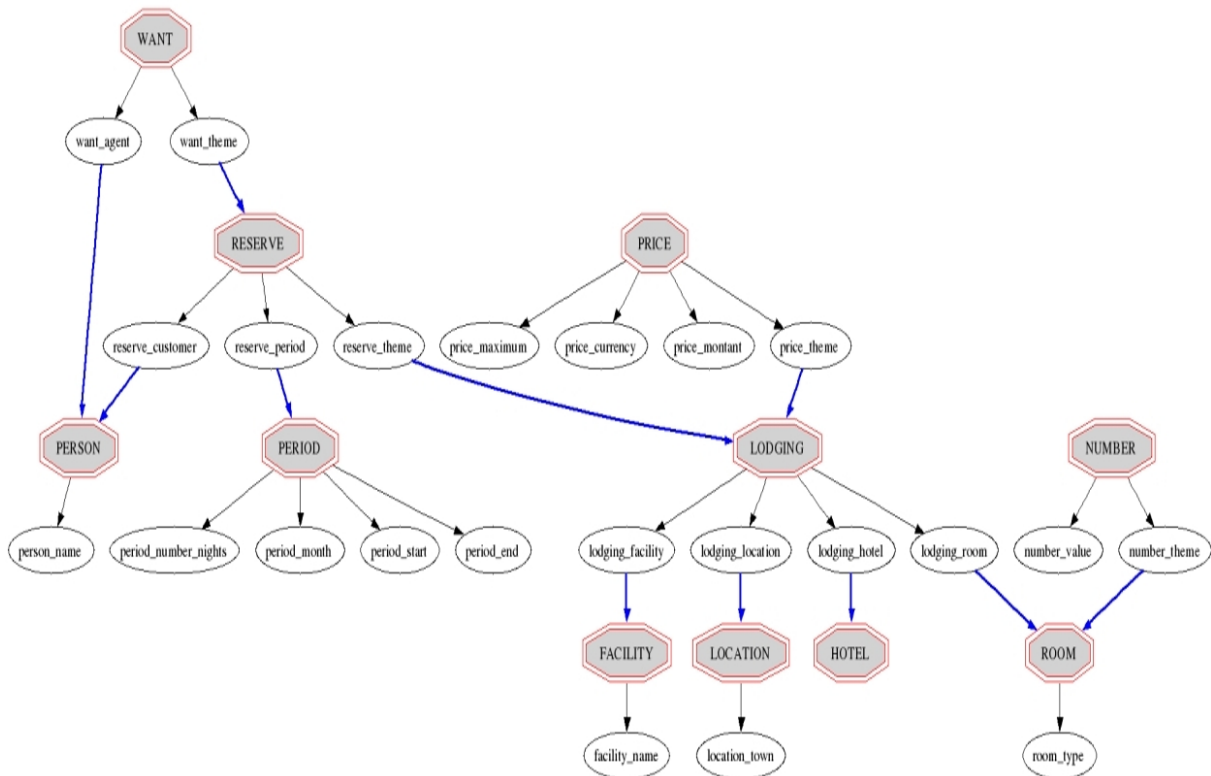


FIG. 4 – Visualisation des frames et FEs d’un tour de parole

Les stratégies utilisées pour améliorer la qualité des annotations sont largement inspirées des approches proposées par (Scheffczyk & Ellsworth, 2006). L’annotation manuelle en frames n’étant pas disponible sur l’ensemble du corpus, l’annotation automatique a été évaluée sur 100 tours de parole sélectionnés par tirage aléatoire. Les annotations présentes ont été contrôlées manuellement par deux experts humains. La comparaison des annotations manuelles et des annotations automatiques fournit une F-mesure de 0,90, pour une précision de 0,96 et un rappel de 0,85.

Le processus incrémental d’annotation utilisé pour annoter en frames le corpus MEDIA est

décrit dans la figure (5).

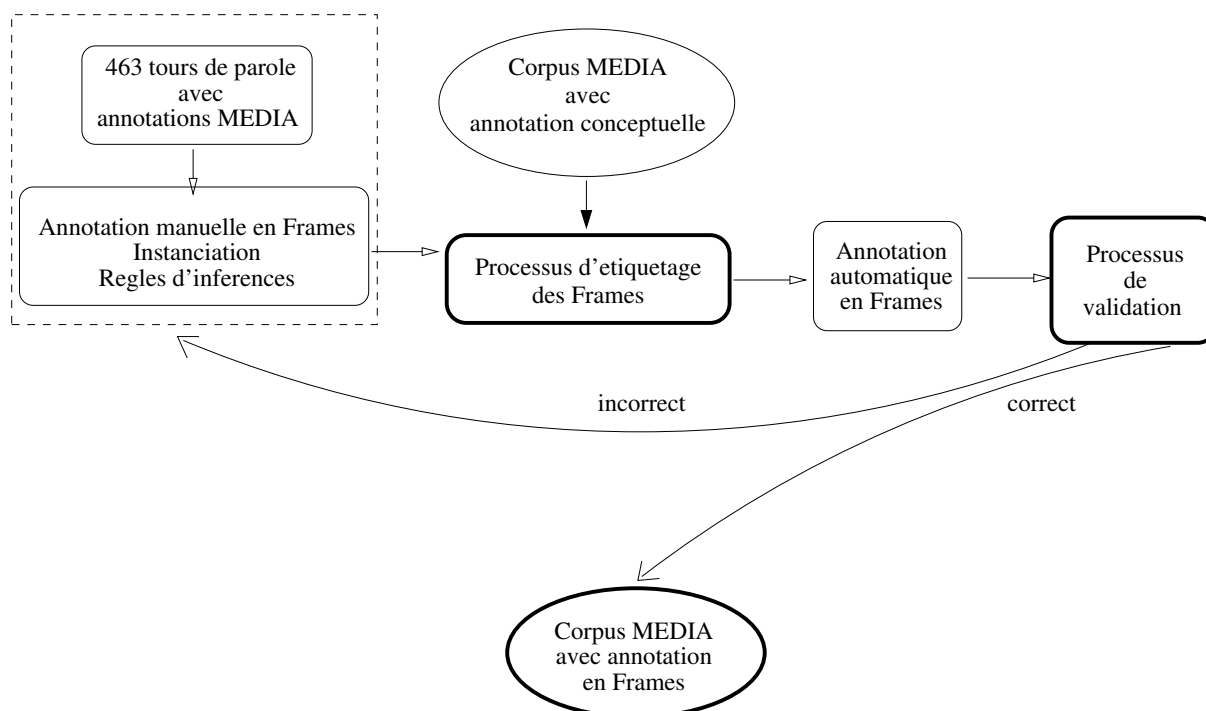


FIG. 5 – Processus incrémental d'annotation

Ne disposant pas de procédure standard d'évaluation sur une telle tâche, il n'est pas possible de se prononcer sur la pertinence globale de l'annotation. Sa valeur effective est évidemment très liée au contexte d'application. Toutefois le recours à un formalisme indépendant de la tâche est une nouveauté qui devrait améliorer grandement la mise au point de l'annotation et les possibilités de l'élargir à d'autres domaines de dialogue homme-machine. Deux approches complémentaires sont en cours d'exploration. Du point de vue qualitatif, la prise en compte des actes de dialogues associés à chaque proposition semble permettre d'améliorer la qualité des inférences. Du point de vue quantitatif, l'utilisation de probabilités lors de l'étape d'inférence, en vue d'améliorer sa robustesse, est en cours d'évaluation.

5 Conclusion

Un formalisme de représentation des connaissances basé sur le paradigme FrameNet a été présenté. Ce formalisme est mis en oeuvre pour annoter en frames sémantiques le corpus de dialogue français MEDIA, de façon incrémentale et partiellement automatique. Pour cela, un processus d'interprétation automatique a été proposé pour composer les structures sémantiques à partir des composants sémantiques de base, utilisant des modèles contenant des unités lexicales et conceptuelles. La F-mesure évaluant la qualité des annotations automatiques produites est suffisamment élevée pour permettre de les considérer comme annotations de référence du corpus. La suite de ce travail consiste à produire des annotations automatiques dans le contexte du système de dialogue (introduction des erreurs en mots et en concepts de base) pour lesquelles nous serons capables de fournir des probabilités associées aux hypothèses de frames et FE's, en tenant compte de tous les niveaux traités (acoustique, modèle de langage, modèle de compréhension, définition des frames et inférence).

Références

- BONNEAU-MAYNARD H., ROSSET S., AYACHE C., KUHN A. & MOSTEFA D. (2005). Semantic annotation of the french media dialog corpus. In *Eurospeech*, Lisboa, Portugal.
- FILLMORE C. J. (1982). *Frame Semantics*. Linguistics in the Morning Calm, Seoul.
- FILLMORE C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, VI(2), 222–254.
- FILLMORE, CHARLES J. C. R. J. & PETRUCK M. R. (2003). Background to framenet. *International Journal of Lexicography*, 16.3, 235–250.
- JACKENDOFF R. (1990). Semantic structures. *The MIT Press, Cambridge Mass.*
- LOWE J., BAKER C. & FILLMORE C. (1997). A frame-semantic approach to semantic annotation. In *Proceedings of the SIGLEX Workshop on Tagging Text with Lexical Semantics : Why, What, and How ?*, Washington D.C., USA.
- PADO S. & PITEL G. (2007). Annotation précise du français en sémantique de rôles par projection cross-linguistique. In *Proceedings of TALN-07*, Toulouse, France.
- PETRUCK M. R. (1996). *Frame semantics*. Handbook of Pragmatics. John Benjamins, Philadelphia.
- SCHEFFCZYK J. & ELLSWORTH M. (2006). Improving the quality of framenet. In *Workshop on quality assurance and quality measurement for language and speech resources*, p. 8–13, Genoa, Italy.
- WOODS W. (1975). *What's in a Link : Foundations for Semantic Networks*. Bolt, Beranek and Newman.