

Semantic Frame Annotation on the French MEDIA corpus

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

Abstract

This paper introduces a knowledge representation formalism used for annotation of the French MEDIA dialogue corpus in terms of high level semantic structures. The semantic annotation, worked out according to the Berkeley FrameNet paradigm, is incremental and partially automated. We describe an automatic interpretation process for composing semantic structures from basic semantic constituents using patterns involving words and constituents. This process contains procedures which provide semantic compositions and generating frame hypotheses by inference. The MEDIA corpus is a French dialogue corpus recorded using a *Wizard of Oz* system simulating a telephone server for tourist information and hotel booking. It had been manually transcribed and annotated at the word and semantic constituent levels. These levels support the automatic interpretation process which provides a high level semantic frame annotation. The Frame based Knowledge Source we composed contains Frame definitions and composition rules. We finally provide some results obtained on the automatically-derived annotation.

1. Introduction

The complexity of Spoken Dialogue Systems (SDS) depends primarily on the complexity of the task concerned. Thus the systems dedicated to the routing of phone calls or the search of information, being able to be presented in forms (telephone directories, search for schedules for example), can rest on a relatively simple representation of semantic knowledge. At the opposite, since a system must be able to manage several joint requests or to integrate phases of negotiation, it requires a high level semantic representation. Such systems must have semantic knowledge being able to interact with complex information on the current state of the dialogue while supporting the errors introduced into the data processing sequence of the speech signal by the recognition and comprehension modules. The difficulty of obtaining such a functional and robust high level representation explains why the current dialogue systems are limited to rather restricted semantic spaces.

The construction of a new representation of knowledge would allow us to obtain a dialogue system able to be called into question, to propose alternatives and to adapt to the phases of negotiation included in the natural dialogue. The proposed representation of knowledge must allow the semantic composition within the speech turns during the course of the dialogue as well as the management of the references (phenomenon of which the appearance is strongly increased at the time of the negotiation phases). From this point of view, the choice of the FrameNet paradigm¹ as high level semantic representation appears suitable. In addition to the robustness of the paradigm, FrameNet has the advantage of producing standard annotations, easy to share and compare within the scientific community.

This work is supported by the 6th Framework Research Programme of the European Union (EU), Project LUNA, IST contract no 33549. For more information about the LUNA project, please visit www.ist-luna.eu.

¹ More informations on <http://framenet.icsi.berkeley.edu>

This paper introduces the semantic frame annotation realised on the French MEDIA dialogue corpus. The next section (2.) is dedicated to the description of the corpus. In section (3.) we describe in details the principles used to provide the Frame annotation. We then present in section (4.) the application of these principles to the MEDIA corpus. We finally conclude in section (5.).

2. The French MEDIA corpus

The MEDIA corpus (Bonneau-Maynard et al., 2005) is a French dialogue corpus simulating a telephone server for tourist information and hotel booking. It has been recorded using a *Wizard of Oz* system (a system in which speakers interact with a computer system that they believe to be autonomous, but which is actually being operated by an unseen human being). Eight *scenarii* categories were defined with different levels of complexity. The corpus accounts 1257 dialogs from 250 speakers and contains about 70 hours of dialogues. Each speaker recorded five different hotel reservation *scenarii*. The French MEDIA corpus is manually transcribed and conceptually rich with more than 80 basic concepts manually annotated. A semantic dictionary gathers these basic concepts.

The semantic dictionary utilized to annotate the French MEDIA corpus associates a *concept-value* pair to a word or a group of words then a *specifier* showing the relations between concepts and at last a *mode* (positive, negative, interrogative or optional) attached to the concept. By defining a set of 19 *specifiers* which are combined with the basic roles, the MEDIA annotation scheme preserve the relationships between concepts. It makes it possible to build a hierarchical representation of a utterance interpretation.

An example of the MEDIA annotation on a message translated from French (*I'd like to book a room for two nights in Marseille*) is given in table (1).

In this example, the *reservation* specifier is given to the *room-amount* and *night-amounts* concepts as a hierarchical structure that represents a reservation is triggered by the concept *command* and filled with the elements

W^c	concept c	mode	specifieur	valeur
<i>I'd like to book</i>	commande	+		reservation
<i>a room</i>	room-amount	+	reservation	1
<i>for two nights</i>	night-amount	+	reservation	2
<i>in Marseille</i>	location-town	+	hotel	Marseille

Table 1: Example of the MEDIA annotation on a message.

found in `room-amount` and `night-amount`. The specifier `hotel` associated to the `location-town` concept connects the town named in the segment “*in Marseille*” with the previous part of the utterance. The combination of the specifiers and the attribute names allows recomposing a hierarchical representation of a query from its flat annotation. This annotation provides labels comparable to semantic constituents hypothesized by a semantic shallow parser.

3. Frame annotation

Semantic structures can be derived from semantic knowledge obtained with a semantic theory. Examples are semantic networks to represent entities and their relations (Woods, 1975) or function/argument structures (Jackendoff, 1990). A convenient way for representing and reasoning about semantic knowledge is to represent it as a set of *logic formulae* from which computational structures such as frames can be derived. In this context, Frames can be seen as cognitive structuring devices used in the understanding process (Fillmore, 1982; Fillmore, 1985). A frame is a model for representing semantic entities and their properties (Petrucci, 1996).

The Berkeley FrameNet project (Lowe et al., 1997; Fillmore and Petrucci, 2003) provides on-line the FrameNet lexical database for English, currently contains more than 10,000 Lexical Units (LU), more than 6,100 of which are fully annotated, in nearly 825 hierarchically-related semantic frames, exemplified in more than 135,000 annotated sentences [February 2007]. A (LU) is a pairing of a word with a meaning, each sense of a polysemous word belonging to a different Frame. For a given Frame, the frame-evoking words are its LUs. A Frame describes a common or abstract situation involving roles called Frame Elements (FEs). For example, the `cooking-creation` Frame, describing food and meal preparation, contains 12 FEs among which Cook, Produced Food, Ingredients or Container and LUs as cook, prepare or bake. The figure (1) visualises the relations between the `cooking-creation` Frame and its related frames as a graph.

3.1. Frame definitions

The choice of a Frame annotation in this work is motivated by the ability of this framework to represent negotiation dialogues and to itself adapt to complex actions of the dialogue manager. We have manually defined a frame based Knowledge Source (KS) to describe the semantic composition knowledge of the MEDIA domain. Frames and FEs are described by a set of manually defined patterns. These patterns are made of LUs, conceptual units (CUs), words and can include features extracted from the compounds of

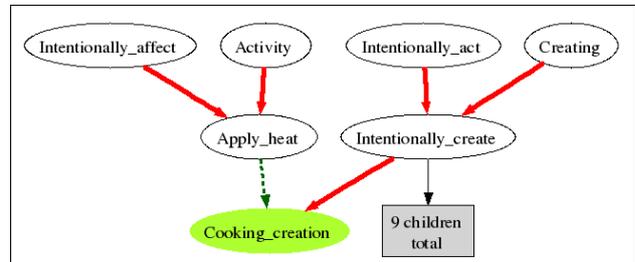


Figure 1: The FrameNet Frame `cooking_creation`

them. Most of the CUs matches some MEDIA basic concepts and some others are defined according to the KS Frames. The FrameNet dictionary is available for English but it does not exist such a database in French, covering the semantic domain of the French MEDIA corpus. To our knowledge, the only available resource in French is the French FrameNet project (Pado and Pitel, 2007) using cross-linguistical projection. Instead of using it, we decide to manually define our Frames for several reasons :

- a smaller but dedicated to the MEDIA domain KS limits annotation confusion
- the specific nature of the textual support (speech transcriptions) induces major differences in the LUs and CUs definitions
- the FrameNet Frames are perhaps too generic to satisfy the needs of a spoken dialogue system.

Thus we decide to define dedicated Frames to the MEDIA corpus domain. However, we try as much as possible to keep the Frame definition application independent. Consequently, some frames describing generic knowledge as spatial relations match FrameNet Frames and some others are application specific but always defined according to the *Berkeley FrameNet* (Lowe et al., 1997) paradigm. The figure (2) shows some of our Frames matching some FrameNet Frames.

The word triggering a Frame instantiation are the LUs associated to a Frames or its FEs. There are two kinds of LUs. The first one includes generic LUs, pointing out the generic focuses of a utterance. The second one involves specific LUs, giving values to predicates. Values of LUs can be words and their synonyms, n-grams or nuples of words appearing in a part of speech, without any constraint on their position in the utterances. An example of LU category and CU associated to a FE is given in table (2).

These patterns allow Frame instantiation then inferences based on instantiated Frames and FEs are performed. Thus,

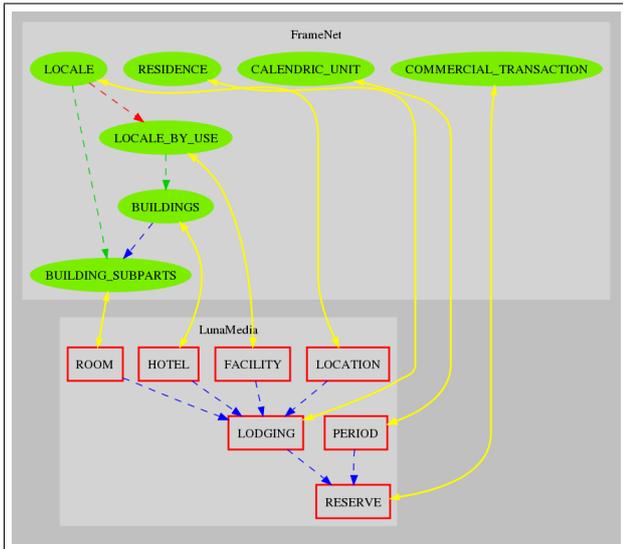


Figure 2: Example of matching between Frames of the KS dedicated to the French MEDIA corpus and FrameNet Frames

```
<framelement fename="location_town">
<concept value="town"/>
<generic_lexical_units value="city,town,village"/>
<specific_lexical_units value="paris,lyon,marseille..."/>
</framelement>
```

Table 2: LU category and CU associated to the FE location_town

the instantiation strategy is a two step strategy : a first step based on Frame and FE definitions previously described and a second step performed by inference rules. This second step is explained in the following subsection.

3.2. Inference rules

The inference process is able to perform inferences about instantiated Frames and FEs using first order logic formulae. Frames and FEs obtained by the instantiation process previously described determine the truth values of the logical rules. According to these truth values, Frames or FEs are instantiated, deleted, modified or connected. For example, if the FE `reserve_theme` associated to the `lodging` concept and the Frame `LODGING` have been instantiated by the first pattern matching, the first order logic rule given in the table (3) creates a link between the FE `reserve_theme` and the Frame `LODGING`. This link indicates that the reservation relates to a lodging.

```
do_link(RESL,L) :- is_fe(reserve_theme,RESL),
is_concept_of(lodging,RESL),
is_fr(lodging,L).
```

Table 3: First order logic Prolog rule linking the FE `reserve_theme` and the Frame `LODGING`

Approximately thirty rules are currently used. They mainly

consist in creating links between Frames and FEs, instantiating Frames and FEs non discovered by pattern matching and deleting redundancies. These rules do not depend on the words of the utterance and on the sequentiality of the proposal.

4. Application

In order to build the KS associated to the MEDIA corpus, the Frame and FE patterns and the inference rules, an initial set of 463 turns from 15 dialogues has been manually annotated. The KS (Figure (3)) contains 21 Frames and 86 FEs. The size of our KS is very small compared with the French FrameNet (Pado and Pitel, 2007) or the English Berkeley FrameNet. Table (4) presents the sizes of the three databases, MEDIA KS, French FrameNet and Berkeley FrameNet. The Frame `REQUEST` appearing in these three databases is mentioned with its FEs according to each database.

Annotations are described in XML documents containing additional information such as time references of words supporting the hypotheses. The manually annotated dialogues were provided thanks to a dedicated tool developed to make manual annotation easier. The average manual annotation time per turn is around 4 minutes.

The automatic extraction of the LU and CU components included in the corpus ensures the patterns completely cover the domain. A XML pattern file gather these components. The pattern file associated to the KS we proposed contains 106 CUs and about 1100 LUs. The French MEDIA corpus initially contained 83 concepts. Our patterns use 60 of them as CUs and 46 new CUs essential to annotating semantic entities not initially considered in the French MEDIA corpus. For example, the CUs associated to the agent Frame `PERSON` (Table (5)) are not initially included in the French MEDIA corpus.

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

Table 5: CUs associated to the agent Frame `PERSON`

A frame visualization tool dedicated to process speech dialogues, called `FriZ`, supports visualization, verification and correction of automatic annotations. This tool, using the `GraphViz`² libraries provides, for each speech turn, a general view of instantiated Frames, FEs and links. `FriZ` allows a direct access to a chosen speech turn, plays the associated audio file and is able to interact with the annotation tool to make the manual corrections easier. Figure (4) shows an example of visualization of the instantiated Frames and FEs when the speech turn contains the speaker message : “Well, I would like to book, hum, two simple rooms, hum, in a hotel in Orange, hum well, for

² <http://www.graphviz.org/>

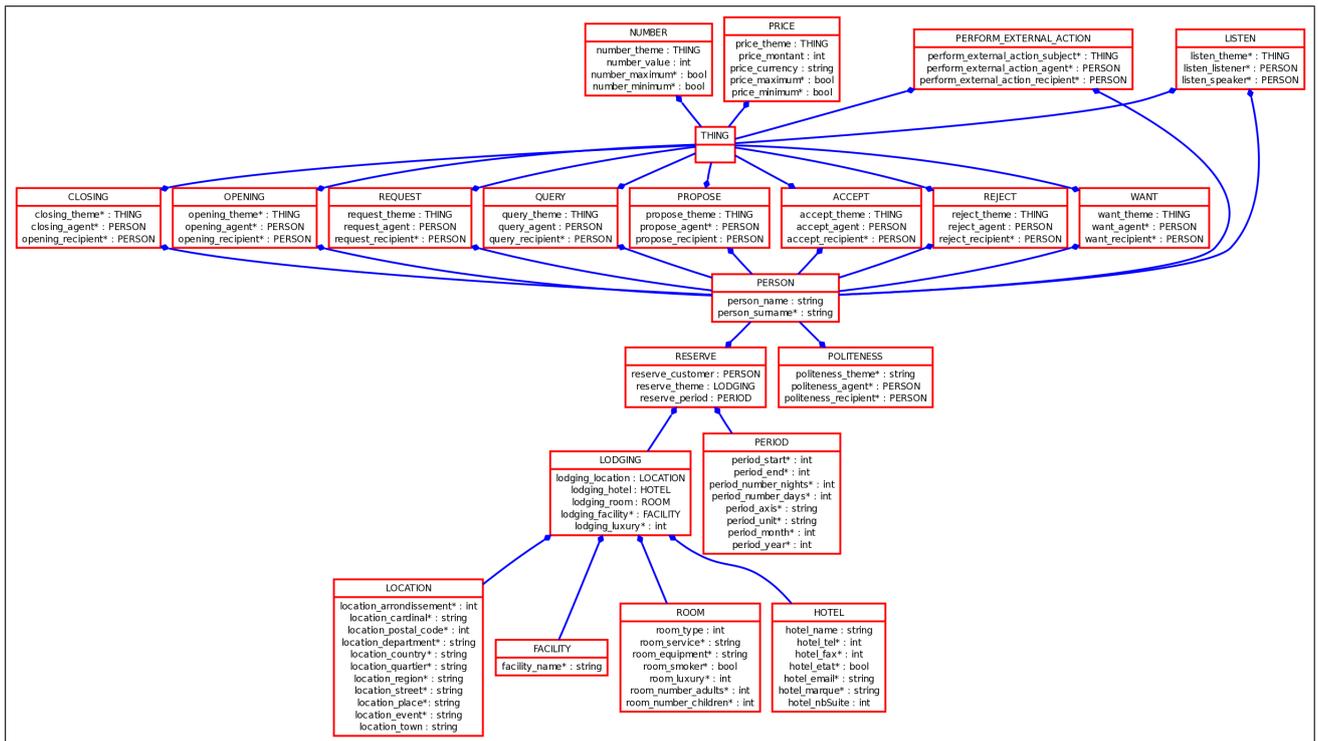


Figure 3: The KS associated to the MEDIA corpus

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

Table 4: Size comparison : MEDIA KS - French FrameNet - Berkeley FrameNet

twenty nights, since July the 12th to July th 31th, hum, I would expected a price less than 100 euros, and, hum, with a swimming-pool.”

Strategies used to improve the annotation quality are close to the approaches suggested in (Scheffczyk and Ellsworth, 2006). It is not possible to manually annotate the whole corpus at the Frame level, hence a random sampling on the test user turns was performed by a human expert to manually assessing the accuracy of the automatic structure annotation. An F-measure of 0.90 (0.96 precision and 0.85 recall) was computed on 100 turns when comparing manual annotations and automatic frame annotations of exact transcriptions. This high accuracy allows to use the automatically-derived annotations as reference annotations. The incremental annotation process used to provide the Frame annotated MEDIA corpus is described in the figure (5).

Lacking a standard evaluation process on such a task, the total relevance of the Frame annotation is difficult to estimate and strongly depends on the application field. However, using a task-independent formalism is original and would improve the quality of annotation and the ability

to catch various domains in SDS. Two approaches are in progress. Taking into account the dialogues acts included in a utterance seems to improve the inference quality and using a stochastic inference model is about to be evaluated.

5. Conclusion

A knowledge representation formalism is proposed. This FrameNet based formalism is used for incremental and partially automated annotation of the MEDIA corpus in terms of semantic structures. An automatic interpretation process is introduced for composing semantic structures from basic semantic constituents using patterns involving conceptual units and lexical units. The process includes procedures for obtaining semantic compositions and for generating Frame hypotheses by inference. Patterns are generalized by progressively annotating data with available knowledge, evaluating confidence of the results, manually annotating samples with low confidence and so on. The F-measure obtained from comparing manual annotations and automatic frame annotations allows to use the automatically-derived annotations as reference annotations. The frame annotated

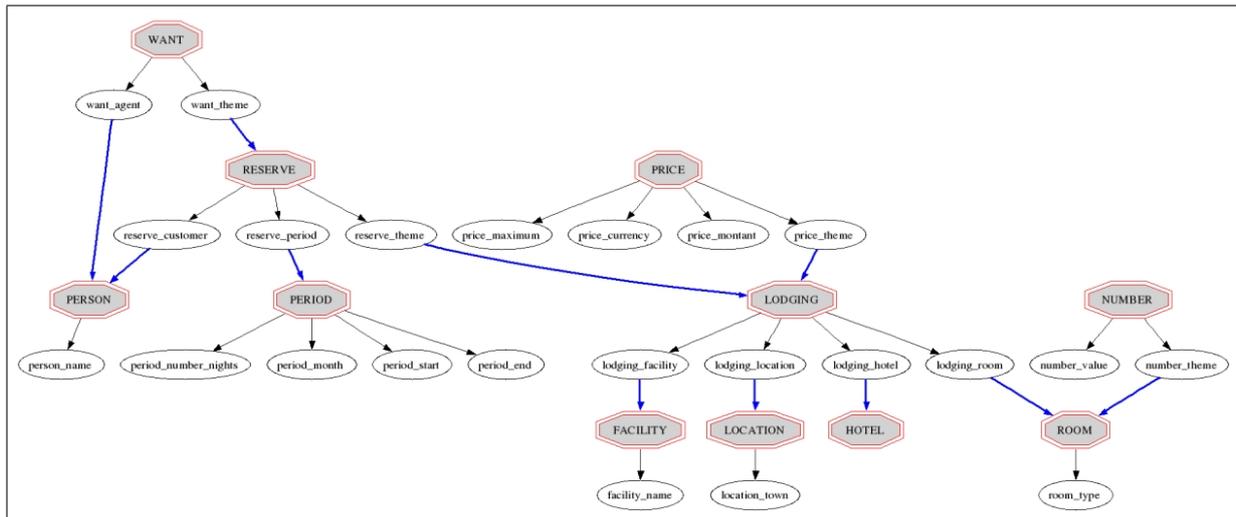


Figure 4: Frame visualization

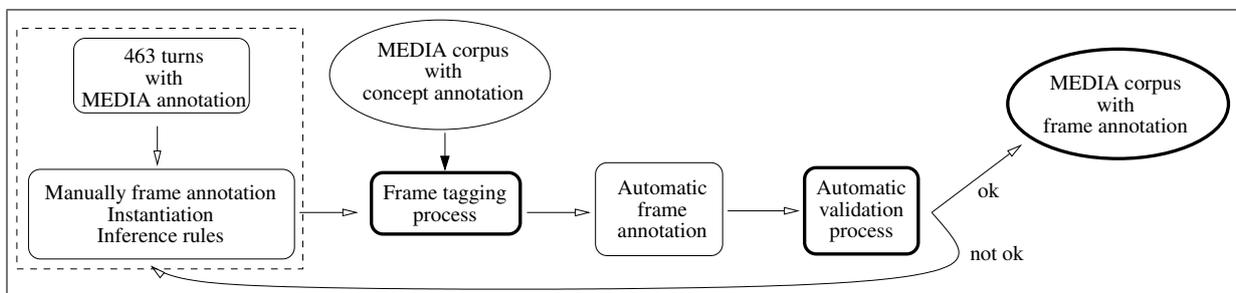


Figure 5: Incremental annotation process

MEDIA corpus and the frame annotation tools will be soon available. In a future work, we plan to produce automatic annotations in the dialogue system context (including the word and concept errors) and provide stochastic Frames taking into account all the available levels (acoustic, language model, understanding model, Frame definition and inference).

6. References

Hélène Bonneau-Maynard, Sophie Rosset, Christelle Ayache, Anne Kuhn, and Djamel Mostefa. 2005. Semantic annotation of the french media dialog corpus. In *Eurospeech*, Lisboa, Portugal.

Christopher R. Johnson Fillmore, Charles J. and Miriam R.L. Petruck. 2003. Background to framenet. *International Journal of Lexicography*, 16.3:235–250.

Charles J. Fillmore. 1982. *Frame Semantics*. Linguistics in the Morning Calm, Seoul.

Charles J. Fillmore. 1985. Frames and the semantics of understanding. *Quaderni di Semantica*, VI(2):222–254.

R. Jackendoff. 1990. *Semantic structures*. The MIT Press, Cambridge Mass.

J.B. Lowe, C.F. Baker, and C.J. Fillmore. 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, April.

Sebastian Pado and Guillaume Pitel. 2007. Annotation précise du français en sémantique de rôles par projection cross-linguistique. In *Proceedings of TALN-07*, Toulouse, France.

Miriam R.L. Petruck. 1996. *Frame semantics*. Handbook of Pragmatics. John Benjamins, Philadelphia.

J. Scheffczyk and M. Ellsworth. 2006. Improving the quality of framenet. In *Workshop on quality assurance and quality measurement for language and speech resources*, pages 8–13, Genoa, Italy.

W.A. Woods. 1975. *What's in a Link: Foundations for Semantic Networks*. Bolt, Beranek and Newman.