

Using MMIL for the High Level Semantic Annotation of the French MEDIA Dialogue Corpus.*

Lina Maria Rojas-Barahona

LORIA/INRIA, France

lina.rojas@loria.fr

Thierry Bazillon

Univ. Avignon, France

thierry.bazillon@univ-avignon.fr

Matthieu Quignard

LORIA/INRIA, France

matthieu.quignard@loria.fr

Fabrice Lefevre

Univ. Avignon, France

fabrice.lefevre@univ-avignon.fr

Abstract

The MultiModal Interface Language formalism (MMIL) has been selected as the High Level Semantic (HLS) formalism for annotating the French MEDIA dialogue corpus. This corpus is composed of human-machine dialogues in the domain of hotel reservation and tourist information. Utterances in dialogues have been previously annotated with a concept-value flat semantics for studying and evaluating spoken language understanding modules in dialogue systems. We are now interested in investigating the use of more complex representations to improve the understanding capability. The MMIL intermediate language is a high level semantic formalism that bears relevant linguistic information, from syntax up to discourse. This representation should increase the expressivity of the current annotation though at the expense of the annotation process complexity. In this paper we present our first attempt in defining the annotation guidelines for the HLS annotation of the MEDIA corpus and its effect on the annotation process itself, revealed by annotators' disagreements due to the different levels of hierarchy and the granularity of the features defined in MMIL.

1 Introduction

MMIL is an ontology-oriented representation language that has been used in several natural language processing (NLP) applications, Denis et al. (2010). It permits the integration of divergent resources in distributed systems as well as the representation of various levels of linguistic analysis. In this work we are particularly interested in exploring the representation of these linguistic levels for analyzing utterances in the context of human-machine interactions. To be able to evaluate the representation on a large set of data the French MEDIA dialogue corpus is used, Bonneau-Maynard et al. (2005). The MEDIA corpus collects about 70 hours of spontaneous speech in the task of hotel room reservation and tourist information. It has been created using a Wizard-of-Oz technique, as a consequence, the utterances are made of many disfluencies, hesitations, false starts, truncations or fillers words (e.g., euh or ben). Thus, the syntactic analysis is relevant for keeping valuable information for further processing (e.g., reference resolution). The semantics describe fine grained predicates, arguments and features based on the domain knowledge. Similarly, the possibility of link references for pragmatic analysis and the representation of the illocutionary force of utterances are relevant to improve the understanding in NLP applications. We selected MMIL for the semantic annotation because it supports the representation of all these features.

Although these features enrich the semantic annotation of utterances in the corpus, they also increase the complexity of the annotation and compromise the agreement between annotators. The possibility of representing different instantiations in MMIL has been the main cause of disagreement between annotators. On the one hand, linguists tend to annotate the surface form of the utterance. On the other

*This work is supported by the French *Agence Nationale de la Recherche* (ANR) and is part of the Project PORT-MEDIA (www.port-media.org).

hand, application designers are more biased towards its canonical representation by keeping relevant task oriented actions and features. The trade-off between these two lines of representation is significant for building appropriately the annotation guidelines for the semantic annotation. The annotation would keep the most valuable information in a multilevel representation for enhancing the understanding capability of NLP applications. In this paper we introduce briefly MMIL and we describe the annotation methodology and the inter-annotation agreement.

2 The High Level Representation

MMIL permits the representation of communicative actions that are represented as *components*. A component is a structure that gathers the communicative event and its propositional content. Components are made up of two main types of **entities**: *events*, which are entities anchored in the time dimension, and *participants*, which are entities not bounded by time. Entities are linked together by *relations* and are described by sets of *features* (i.e. pairs of attribute-value), Denis et al. (2010). Every component has a unique communicative event with the illocutionary force represented by means of the *dialogueAct* feature. The propositional content is represented as a *main event* with its arguments, which can be either events or participants, linked to the communicative event by a relation **propContent**. In this representation, predicates are usually represented as events and predicate arguments are usually represented as participants. Relations between participants and events usually describe the thematic roles.

French: "/1euh vous venez de dire que précédemment qu' il n' a y avait plus de chambres disponibles à ces dates et maintenant vous en avez/2 donc je voulais juste m' assurer qu' au Novotel vous avez bien une chambre double euh pour un couple avec un enfant avec une baignoire dans la chambre euh il me il me faut un Parc Ã proximit  et euh cent dix euros maximum la nuit est-ce-que vous pouvez v rifier"
English: "/1um you just said earlier that there are not more rooms available on these dates and now there are/2 so I just wanted to be sure that you have at the Novotel a double room for uh a couple with one child with a bath in the room uh I need a park nearby and uh hundred and ten euros up at night is that you can check"

Figure 1: Example of a complex utterance of the MEDIA Corpus.

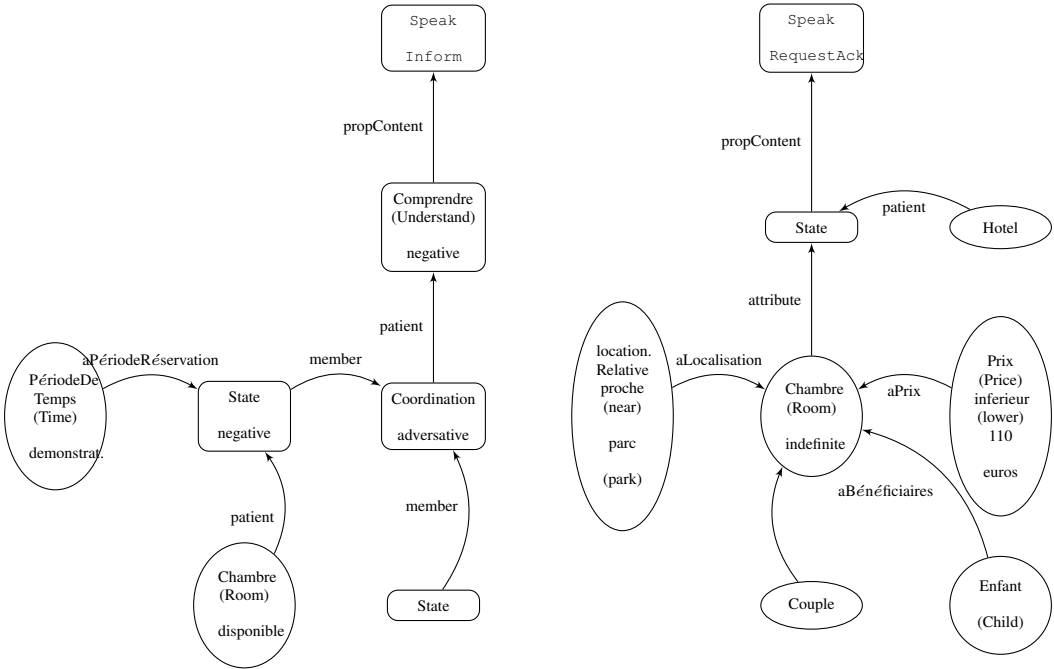


Figure 2: HLS as an abstraction of the meaning of the French utterance shown in Figure 1. Left: this component expresses the inform of a misunderstanding of the first segment ("1" in Figure 1). Right: this component is a request acknowledgment, representing the second segment ("2" in Figure 1). Note that *events* are exemplify by square boxes while *participants* are exemplify by ellipses.

Let us focus on the MMIL representation for a typical utterance of the MEDIA corpus, given in Figure 1. In this utterance the user first announces an inconsistency, then asks for clarification. Thus, two MMIL components with different communicative actions, *inform* and *request acknowledgment*, have been used, as shown in Figure 2. The component on the left has a main event that describes the misunderstanding expressed in the first segment¹ of the utterance. It is represented by the ontological concept "Understand" and by the syntactic feature *polarity* with the *negative* value. It also contains a coordinated entity mirroring an *adversative* coordination between two events, *state*. The event *state* represents the status of something, therefore the *negated state event* can be understood as "there are not more rooms available on these dates" while the *positive state* represents "now there are". The participants symbolize the arguments "rooms" and "dates" respectively. The component on the right expresses the clarification request of the second segment. It verifies the status of the hotel with the specific constraints.

3 The Annotation Methodology

In the process of defining the annotation guidelines, we elaborated a specification document that describes the representation of dialogue acts, events and exemplifies the high-level semantics. Moreover, it delves into the methodology that might be applied for the automatic and manual annotation. Afterwards, a linguist expert and a project designer were in charge of defining the annotation guidelines. For this purpose, they annotated manually a subset of utterances which were supposed to be representative of the most complex aspects of the HLS annotation, in terms of their semantic constituents. 330 utterances were selected. They are all directly related to the reservation task (first two rows in Figure 4) and mostly occurred in the first 3 turns of the dialogues when the user is describing his goal, defined as an overall objective along with a set of constraints. Hereafter, we present the preliminary evaluation of the experts' agreement on these utterances.

The annotation process has been supported by an annotation tool: ATool. It accesses two knowledge-bases, one for the MMIL formalism and the other for the MEDIA domain. The latter is adapted from the MEDIA evaluation campaign, Bonneau-Maynard et al. (2006). ATool permits annotators to navigate through utterances, while displaying the MMIL representation. Annotators can design the MMIL components graphs, define the MMIL entities by associating features, values and segment. ATool will suggest the possible features and values for the MMIL formalism and for the domain according to the knowledge-bases ensuring the integrity of the constructed MMIL components in the annotation.

The MEDIA corpus is rich in expressions that evoke several communicative actions. Figure 4 shows a few examples. For the purpose of the task, we are interested in the underlying meaning of sentences, thus politeness and indirectness are discarded from the HLS representation. For this reason, in *requests* the speaker is the patient, while the hearer is the agent (see Figure 4). Because when translating the utterance into its deep instantiation, the speaker will benefit from the execution of the action, while the hearer has the obligation to perform the action. All the expressions in the corpus that bear the semantics of "command for a reservation" (e.g., *je veux réserver*, *je souhaite réserver*, *je voudrais faire une réservation*, *j'aimerais faire une réservation*, all equivalent to *I would like to reserve*), have been normalized with the deep component shown in Figure 3, exemplifying unequivocally the user's desire to request for a reservation. The possible arguments and roles have been detailed in the domain knowledge-base. As a consequence the knowledge-base defines relations between hotels, rooms, customers, prices, equipments, services, locations and dates. Besides, the grammatical relations and features, such as coordination, have been defined in the MMIL knowledge-base. Coordination is indicated with the "coordtype" feature and it is used in cases of *conjunction* (*je veux une chambre simple et deux chambres double*, *I want a single room and two double bedrooms*), *disjunction* (*Paris ou en proche banlieue*, *in Paris or suburbs*) or *adversation* (*en ville mais pas trop loin de la mer*, *in the city but not too far from the sea*).

For annotating events we can find the main verb in the utterance and represent it as the main event in MMIL by following a domain-specific classification of verbs, from which Figure 4 shows some equivalences among dialogue acts and verbs. For each participant or event, several features can be

¹Segments are sequence of words that are depicted as "/*i*", where *i* is the number of the segment.

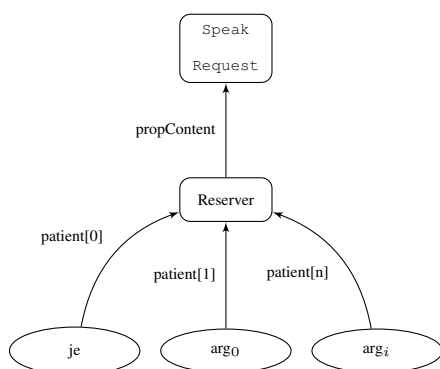


Figure 3: Canonical representation of a booking request in MEDIA.

D. Act	EvType	Examples	Semantic Roles
Request	Reserver	réserver [la chambre]	aObjetRéservé
		réserver [pour le troisième week-end de novembre une nuit] [à Clermont-Ferrand] [pour quatre chambres doubles]	aPériodeRéreservation aLocalisation aObjetRéservé
Inform	Inform	[j'] ai des informations supplémentaires	agent
Request	Inform	[j'] aurais aimÃ avoir exactement [les dates]	patient[0], patient[1]
Request	State	[Il] est [Ã combien]	patient, aPrix
Request	Repeter	pouvez-[vous] répéter	agent
Inform	Repeter	[je] vais me répéter	agent
Accept		oui	
Reject		non	

Figure 4: Some of the observed dialogue acts and main events with their arguments in the corpus.

added. The most important of them are “object type” (for participants) or “event type” (for events), which specify their ontological concepts. They may be réserver (reserve), hôtel (hotel), chambre (room), périodedetemps (time), ville (city), person, adulte (adult), enfant (child), localisationnommée (places), among others. There are more specific features, for instance, the journey dates, hotel features (e.g., name, standing, services, etc). Some of these features have predefined values, such as the *gender* of an object (either masculine or feminine). On the other side, features such as *cardinality*, have not predefined values, in that case, the annotator has to manually indicate the correct value.

Obviously, the annotation task difficulty increases with the utterance’s complexity. The representation is rather tedious to define in elliptical utterances, such as multiple reservations, in which implicit and explicit information must be taken under consideration. Furthermore, the MMIL formalism does not support the association of discontinuous segments to entities, generating some imprecisions in the HLS annotation. For instance, in *je voudrais une chambre pour deux personnes euh simple* (I would like a room for two people uh simple), “une chambre” (a room) and “simple” should be linked to an unique participant, having as *object type* (“Room”) and as *type of room* “simple”. However, given that the speaker has not mentioned “simple” right after “chambre”, there is a new element imbricated between them: “pour deux personnes”. As a result, the annotator must integrate the subsegment ‘pour deux personnes’ in the “Room” participant. Even though this subsegment is also associated to the “Personne” participant.

4 Results

When analyzing the sample of 330 utterances that were annotated, we found a perfect agreement between annotators in the detection of dialogue-acts, main events, as well as main arguments. In contrast, when measuring fine-grained features inside components we found eight types of disagreement, namely *conjunctions*, *disjunctions*, *creation of participant for simple features*, *groups of features inside entities*, *features of entities*, *values of features*, *relation names* and *relation among entities*. The most frequent cases concern the first two, which refer to coordination: conjunctions (20%) and disjunctions (5%). The inter-annotator agreement for the coordinate entities was computed, obtaining the kappa measure, Carletta (1996), of 0.25 for conjunctions and 0.15 for disjunctions, meaning a fair and slight agreement respectively. Although the other cases were less frequent, the inter-annotator agreement was even lower, indicating no agreement.

In spite of the disagreement, when measuring the global similarity between the MMIL components created by both annotators we found a high score of 98%. This metric measures the graph similarity

by computing the similarity between entities and relations, including the fine-grained features inside entities. The speech-act, main-event and main arguments are in compliance with the specifications in both annotations.

Case	Annotator 1	Annotator 2
Conjunctions	68	56
Disjunctions	18	10
Part. for simple feats.	11	0
Grouping feats.	0	2

Case	Discrepancy
Features	4
Features' values	5
Name of relations.	5
Relation among entities.	2

Figure 5: Left: the Table displays the number of utterances by annotator for the listed cases. Annotator 1, is the linguist expert, Annotator 2 is the project designer. Right: the Table shows the number of utterances with a completely discrepant annotation: different features for same entities, different values for same features, different relation between same entities and entities related differently in a component.

These issues show that the disagreement cases were less frequent. So far, annotators have not being so rigorous when segmenting the text inside features. Therefore, segmentation needs to be checked in both annotations. After this experiment, we are defining the final certified annotation and deriving the annotation guidelines formally.

5 Discussion

Defining the annotation guidelines for high level semantic representation is controversial. The multiple features that can be represented in the selected MMIL formalism, as well as the multiple instantiations offer different possibilities for representing the same utterance. In general representing spoken utterances is cumbersome, because of the linguist phenomena present in spontaneous speech. As a consequence, annotators have to deal not only with the explicit, but also with the implicit information, and in some cases the representations might be subjective. For these reasons, we defined the standard for the annotation, and based on it, we carried out an annotation experiment on a sample of 330 complex utterances, directly related to the reservation task; involving two annotator profiles i.e., a linguist and a project designer. Afterwards, we measured the similarity between the annotated MMIL components and the inter-annotation agreement obtaining a 98% of similarity and only eight major cases of disagreement, coordination discrepancy being the most frequent. Right now, we are refining the final annotation guidelines based on these results. This first experiment analyzes the most complex and numerous utterances in the corpus covering reservation requests and affirmations. Subsequently, misunderstanding, questions and clarifications will be analyzed following the same methodology. As a result, we will be able to reduce the disagreement between annotators in order to produce the annotation of the whole MEDIA corpus, which will be made freely available to the research community.

References

- Bonneau-Maynard, H., C. Ayache, F. Bechet, A. Denis, A. Kuhn, F. Lefèvre, D. Mostefa, M. Quignard, S. Rosset, C. Servan, , and J. Villaneau (2006). Results of the french evalda-media evaluation campaign for literal understanding. In 5th International Conference on Language Resources and Evaluation (LREC2006).
- Bonneau-Maynard, H., S. Rosset, C. Ayache, A. Kuhn, and D. Mostefa (2005). Semantic annotation of the french media dialog corpus. In *INTERSPEECH-2005*, 3457-3460.
- Carletta, J. (1996). Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* 22(2), 249–254.
- Denis, A., L. M. Rojas-Barahona, and M. Quignard. (2010). Extending MMIL semantic representation: Experiments in dialogue systems and semantic annotation of corpora. In *proceedings of the Fifth Joint ISO-ACL/SIGSEM Workshop on Interoperable Semantic Annotation (ISA-5), Hong Kong, January 2010*.