A Language Model for Human-Machine Dialog: The Reversible Semantic Grammar

Jérôme Lehuen and Thierry Lemeunier

LIUM, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans, France

Abstract. In this paper we present algorithms for analysis and generation in a human-machine dialog context. The originality of our approach is to base these two algorithms on the same knowledge. The latter combines both semantic and syntactic aspects. The algorithms are based on a double principle: the correspondence between offers and expectations, and the calculation of a heuristic score. We present also some results obtained by performing an evaluation based on the MEDIA French corpus.

1 Introduction

The human-machine spoken dialog systems generally dissociate analysis processes from generation processes. Yet, it is useful to pool the knowledge of the analysis and of the generation processes so that the system can say what it understands and can understand what it says. There are at least three consequences. Firstly, the user's utterances can be quoted or reformulated. Secondly, the machine can control itself during the generation processes. Thirdly, a joint improvement of these two skills is possible as soon as the knowledge increases.

We propose a specific model of language called the Reversible Semantic Grammar (RSG) which combines both semantic and syntactic aspects. The originality of our approach consists in the reversibility: this language model is used at the same time to analyze and to generate utterances. Another important point of this dual representation is that it allows a robust analysis of the spoken utterances and a rather subtle generation.

The RSG takes place in an existing dialog engine which is structured into four levels (see Fig. 1). The three upper ones (language, dialog and interactive task) communicate through a shared working memory, in the same way as the "blackboard" architectures. This article focuses on the language level. The RSG is presented in section 2. The analysis and the generation algorithms are detailed and exemplified in section 3. Some results and a comparison with other systems are given in section 4.

2 The RSG Model of Language

The knowledge model is based on the notions of concept and of relations between concepts. This last aspect is inspired by the model of dependency structures proposed by the French linguist Lucien Tesnière [1]. A concept represents an object, an action, an

Lehuen J. and Lemeunier T.
A Language Model for Human-Machine Dialog: The Reversible Semantic Grammar.
DOI: 10.5220/0003017200170026
In *Proceedings of the 7th International Workshop on Natural Language Processing and Cognitive Science (ICEIS 2010)*, page
ISBN: 978-989-8425-13-3

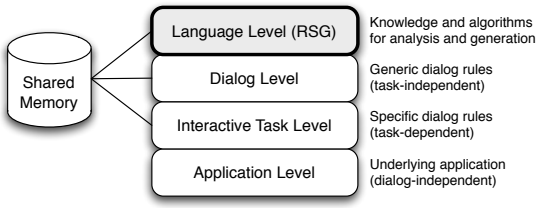

Fig. 1. RSG in the dialog engine.

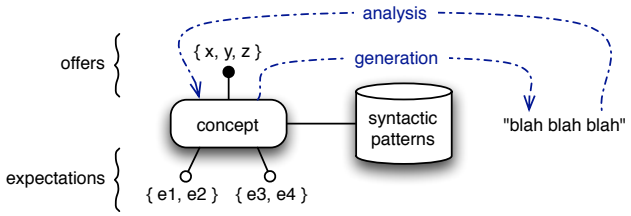

Fig. 2. Model of a concept of valence 2.

attribute, etc. The features of a concept are its offers, its expectations, and its syntactic patterns (see Fig. 2).

The offers are semantic categories and define the concept [2]. The expectations are also semantic categories but describe the potential links with other concepts. An expectation is not necessarily required. Each concept has a library of syntactic patterns. The latter contains words and references to the expectations of the concept. Each pattern can be characterized by distinctive features (pragmatic and/or morphological ones) which allow a precise description of the pattern's use (level of language, politeness, illocutory force, gender, number, etc.). This full-form encoding simplifies the generation process.

The representation of an utterance consists in an oriented graph. The nodes in the graph, called "comprehension granules", are concepts instanciated in the context of utterance. For example the following French utterance: *"bonjour, je voudrais une baguette bien cuite s'il vous plat"* (hello, I'd like a nice crisp baguette please) is represented in Fig. 3.

3 The Analysis and the Generation of Utterances

3.1 The Analysis Algorithm

According to the knowledge model, the analysis consists in building a structure of granules from a string of characters. The process composed of five stages creates and links granules according to syntactic criteria (the pattern matching) and/or semantic criteria (the correspondence between offers and expectations). These two kinds of criterion are either combined, or used independently from one another in order to generate hypotheses. The five stages of the analysis are:

1. **Nucleus Identification:** the instantiation of certain types of "leaf granules" (such as dates) using regular expressions or local grammars;

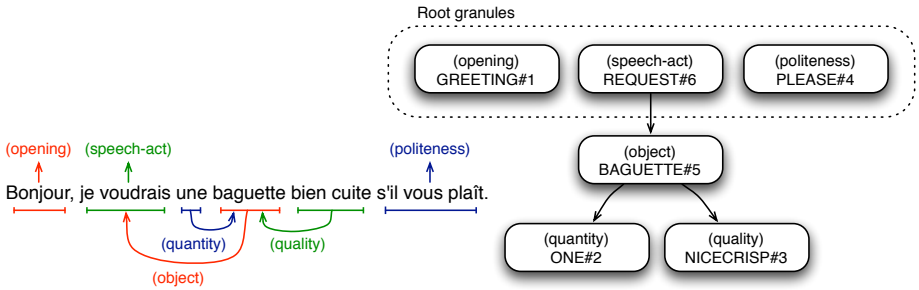

Fig. 3. Dependency structure produced by the RSG.

2. **Utterance Segmentation:** the identification of parts of the utterance likely to contain a granule using the concepts' expectations and the syntactic patterns. The aim is to generate some lexical hypotheses (word or expressions) as hypothetical granules;
3. **Structure Construction:** the instantiation of granules as well as the connections between them using the concepts' expectations and the syntactic patterns. The algorithm is based on the opportunistic filling of a chart and on a heuristic function;
4. **Conflict Resolution:** the removal of the remaining conflicts of position between granules. The "weak granules" are suppressed for the benefit of the "strong granules" using an evaluation function;
5. **Granule Rescuing:** the linking of "orphan granules" to others thanks to hypothetical links using the correspondence between offers and expectations, as well as a proximity criteria.

The method we use to analyze without backtracking is "chart parsing" [3] [4]. Basically, a chart parser emulates parallel processing by concurrently pursuing all alternative analyses at each step. The chart parser can identify the multiple understandings. It is up to the dialog level to select one through interaction. At the end, there only remain the "strong granules" or conflictual granules of the same score (Fig. 4 presents an analysis result).

A score of a granule G is computed by taking into account its coverage (number of words), its dispersion (number of words that are not taken into account), the scores of its links Ai and the scores of its children Gi. The score of a link is equal to the number of common features between the offers and the expectations. An example of score computing is given in Fig. 4.

$$score(G) = coverage(G) - dispersion(G) + \sum_{i=1}^{val} (10 \times score(A_i) \times score(G_i))$$

3.2 Some Particular Cases of Analysis

This section presents examples illustrating particular performances, such as the identification of tonic or periphrastic questions and as the resolution of certain ambiguities and as the generation of hypotheses.

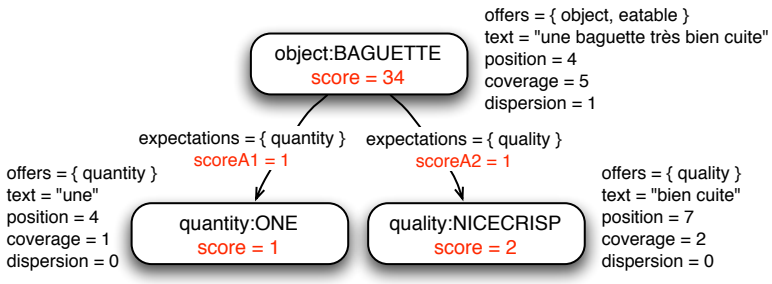

Fig. 4. Calculation of the score of the granule BAGUETTE.

– The first example illustrates the difference between the analysis of a tonic question and the analysis of a periphrastic one. In the representations below, the tonic character is indicated by the underlined categories and the periphrastic character is indicated by the speech-act granule:

a) *Le second semestre commence quand ?* (When does the 2nd semester start?)

```
⇒ (information:DATE question tonic
    A1:(event:S2))
```

b) *Je voudrais savoir à quelle date commence le second semestre* (I'd like to know when the 2nd semester starts)

```
⇒ (speech-act:REQUEST
    A1:(information:DATE
        A1:(event:S2)))
```

– The second example shows how the analyzer naturally solves the rare (and artificial) cases of homonymous ambiguity, if the context so permits. Here, the problem with the French word "avocat" (which can be both a lawyer or an avocado) is solved thanks to its adjective:

c) *Je voudrais un avocat bien mûr* (I'd like a well ripe avocado)

```
⇒ (speech-act:REQUEST
    A1:(object:AVOCADO
        A1:(quantity:ONE)
        A2:(aspect:RIPE)))
```

d) *Je voudrais un avocat compétent* (I'd like a competent lawyer)

```
⇒ (speech-act:REQUEST
    A1:(person:LAWYER
        A1:(quantity:ONE)
        A2:(skill:COMPETENT)))
```

– The coordinating-conjunction *"et"* (and) is built according to the categories or the features of the granules situated on the left and on the right of the conjunction. The category of the concept AND is given by its two child-granules:

e) *Puis-je avoir une baguette et avez-vous deux croissants ?* (Can I have one baguette and have you got two croissants?)

```
⇒ (speech-act:AND
    A1:(speech-act:REQUEST
        A1:(object:BAGUETTE
            A1:(quantity:ONE)))
    A2:(speech-act:REQUEST
        A1:(object:CROISSANT
            A1:(quantity:TWO))))
```

f) *Je voudrais une baguette et deux croissants* (I'd like one baguette and two croissants?)

```
⇒ (speech-act:REQUEST
    A1:(object:AND
        A1:(object:BAGUETTE
            A1:(quantity:ONE))
        A2:(object:CROISSANT
            A1:(quantity:TWO))))
```

– The rescuing stage enables the analyzer to deal with dislocated utterances which produce "orphan granules". The analyzer will try to connect them to nearby granules if possible. The codes of the hypothetical connections are underlined in the examples below:

g) *Le second semestre je voudrais savoir la date* (the 2nd semester I would like to know the date)

```
⇒ (speech-act:REQUEST
    A1:(information:DATE
        ?A1:(event:S2)))
```

h) *Le second semestre la date je voudrais savoir* (the 2nd semester the date I would like to know)

```
⇒ (speech-act:REQUEST
    ?A1:(information:DATE
        ?A1:(event:S2)))
```

– Finally, thanks to the step of segmentation, the analyzer is able to make lexical hypotheses concerning unknown words. The unknown segments *"le truc"* and *"le bidule"* are identified with the syntactic pattern *"ranger [objet] dans [rangement]"* (to tidy [object] in [place]).

i) *Ranger le truc dans le bidule*

```
⇒ (action:TIDY
    A1:(?object:[le truc])
    A2:(?place:[le bidule]))
```

3.3 The Generation Algorithm

The objective is to build a sentence from a granule structure by using the most adequate syntactic patterns. The principle is based on the propagation of a "global generation goal" *Tg* from the root to its leaves. A score is calculated for each node and for each verbalization according to *Tg* and to a local criterion that takes into account its distinctive features, the scores of its children and links. The final sentence is obtained by combining the best verbalizations. Let *V* be a verbalization and *Tp(V)* its distinctive

features. Let *Vi* be a candidate verbalization for the i^{th} child granule and *Tp(Vi)* its distinctive features. The score of *V* is computed by the following formula:

$$score(V) = card(Tp(V), Tg) + \sum_{i=1}^{val} (10 \times score(Vi) + card(Tp(V), Tp(Vi)))$$

where $card(s1, s2) = cardinality(s1 \cap s2)$ for two sets s1 and s2. In Fig. 5, the score of the verbalization *"une baguette bien cuite"* is 25, whereas the score of the wrong verbalization *"un baguette bien cuit"* is 23 (there are two gender agreement errors: *"un"* instead of *"une"* and *"cuit"* instead of *"cuite"*). Let the generation goal be *Tg* = {colloquial}, then the score of the verbalization *"je voudrais une baguette bien cuite"* will be 251 whereas the score of the best verbalization *"file-moi une baguette bien cuite"* will be 252.

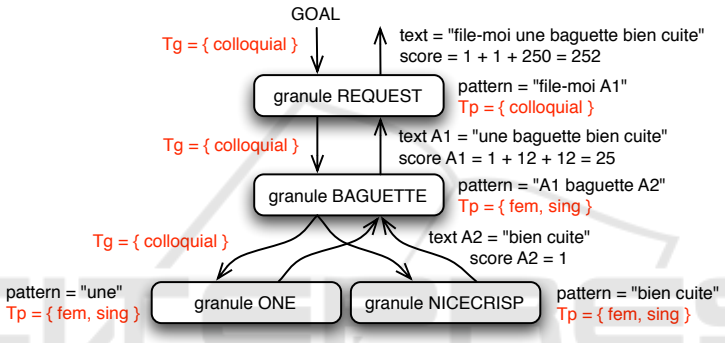

Fig. 5. Example of a generation process.

3.4 An Example of Use of the Generation Capabilities

The following example shows how RSG can be used in order to reformulate a dislocated utterance like: *"Je voudrais baguette... une"* (Can I have baguette... one). The result of the analysis of this utterance is:

```
⇒ (speech-act:REQUEST
   A1:(object:BAGUETTE
       ?A1:(quantity:ONE)))
```

There is an uncertain link between the two granules BAGUETTE and ONE (it's the result of the "granule rescuing" stage (see 3.1)), so the dialog engine has to check it by engaging a validation tactic. The generated response is: *"Pardon, j'ai un doute sur la quantité exprimée, vous voulez bien une baguette ?"* (Sorry, I have doubts about the quantity you have expressed, do you want one baguette?). This response consists of two parts. The first one (Sorry [...] expressed) uses a description pattern of the expectation A1 of the granule BAGUETTE. The second one (do you want one baguette) is based on a verification pattern of the granule REQUEST.

4 The Evaluation of the Analyzer

4.1 The MEDIA Corpus

At this time, we have evaluated only our analyzer using the French MEDIA dialog corpus. This corpus has already been used in the MEDIA evaluation campaign [5] [6] that aimed to define and test an evaluation methodology to compare and diagnose the understanding capability of spoken dialog systems. Both context-independent and context-dependent understanding evaluations had been planned but only the first possibility was performed.

The MEDIA task is the reservation of hotel rooms with tourist information, using information obtained from a web-based database. The dialog corpus was recorded using a WOZ system simulating a vocal tourist information phone server. Each caller believed he or she was talking to a machine whereas in fact he was talking to a human being (a wizard) who simulated the behavior of a tourist information server. In this way, 1257 have been recorded from 250 different speakers where each caller carried out 5 different hotel reservation scenarios. The final corpus is about 70 hours of dialogs transcribed then conceptually annotated according to a semantic representation defined within the campaign. It is divided into a training part of 11000 utterances and an unseen test part of 3000 utterances.

Four kinds of scores have been proposed according to the characteristics of the semantic representation in which there are four modes (affirmative, negative, interrogative or optional) and a set of attributes (representing the meaning of a sequence of words) more or less specified (by adding some specifiers taken from a limited list). The full scoring uses the whole set of attributes whereas in the relax scoring, the specifiers are no longer considered. Another distinction is done by taking into account either the 4 modes or only 2.

4.2 The MEDIA Semantic Representation

The MEDIA evaluation is based on a flat semantic representation [5] which is based on a list of attribute-value structures in which conceptual relationships are implicitly represented by the names of the attributes. More precisely, the meaning of a segment of words is represented by a triplet which contains the name of the attribute, the value of the attribute, and a mode which can be affirmative (+), negative (-) or interrogative (?). For example, the utterances *"Je voudrais réserver un hôtel à Paris près de la gare Saint-Lazare du quatorze au seize"* (I'd like to book a hotel in Paris near the Gare Saint-Lazare from the fourteenth to the sixteenth) is represented in Table 1.

This flat representation is less expressive than ours, which is a hierarchical one, but we must project it to this flat one so as to be able to perform an evaluation using the common evaluation tools. We have identified 3 major steps to project the granules structures to triplets sequences:

1) The creating of incomplete triplets (attribute, value, +) from the granules: concepts become values and categories become attributes. For example, the concept "near"

which offers the category ”localization-relativeDistance” produces the attribute ”localization-relativeDistance” with the value ”near”. The mode is set to + by default, unless it is noted in the syntactic pattern.

2) The appending of missing specifiers to attributes on the basis of propagation rules. For example, the specifier ”-hotel” is appended to the attribute ”localization-relativeDistance” because of the preceding triplet (BDobject, hotel, +).

3) The adding or the changing of some modes on the basis of inference rules. For example, the mode of all the triplets of the second utterance are changed into an interrogative one because of the interrogative marker *”Quels sont”*. The latter is recognized by our analyzer but does not produce any triplet because it does not correspond to any attribute of the MEDIA semantic dictionary.

Table 1. An example of a MEDIA semantic representation.

Word sequence	Attribute name	Normalized value	Mode
je voudrais réserver	command-task	reservation	+
un hôtel	BDobject	hotel	+
à Paris	localization-town-hotel	Paris	+
près de	localization-relativeDistance-hotel	near	+
la gare Saint-Lazare	localization-relativeNamedPlace-hotel	gare Saint-Lazare	+
du quatorze	time-day-month-from-reservation	14	+
au seize	time-day-month-to-reservation	16	+

4.3 The Results

We present in Table 2 error rates calculated from comparisons of our results and expected ones. This rate is: (SUB + DEL + INS) / TOT where SUB is the number of changed triplets, DEL is the number of deleted triplets, INS is the number of added triplets, and TOT the number of triplets we have to identify. We obtain a pretty good score in the most difficult mode (full 4 modes). The deviations of 4.7% and 5% between the ”2 modes” rates and the ”4 modes” rates correspond to the identification of questions which are difficult to identify without taking into account the context of the application. Our system is more efficient than others on this point. There is an average deviation of 2.9% between the ”relax modes” rates and the ”full modes” rates. This can be explained by the relative proximity between our semantic representation and the MEDIA one, which avoids errors of projection of granules to triplets.

We obtain 5986 correct concepts, 1363 SUB, 1038 DEL and 499 INS. The major problem is the very important number of substitutions. If we have a look on the alignment charts, we can see that this is a side effect of the Levenshtein algorithm. In fact, a lot of substitutions are consequences of alignment errors, which are due to the deletions and to the insertions. So to reduce the substitution errors rate, clearly we have to work in order to reduce the deletions rate and the insertions rate. The important number of substitutions comes also from a source of errors already identified in [7] related to the projection of a hierarchical representation to a linear one. Indeed, the triplets follow the order of words, while structures of granules are not dependent on that order. Whatever

the projection method, we observe inversions compared to the expected order, which together increase the SUB rate and the INS rate.

If we sort the errors by category (cf. Table 3), we can notice that an important percentage of errors concern the recognition of the concept *refLink-coRef* (26.18% of the deletions, and 16.35% of the insertions). In the annotated corpus, this concept corresponds to some articles (*le, la, les*), to some pronouns (*il, elle*), and to some demonstrative adjectives (*ce, cet, cette*), in anaphoric reference contexts. We tried to enhance the recognition rate by working on syntactic patterns. We observed a communicating vessels effect between the DEL rate and the INS rate, but no global significant improvement. Therefore we think that we can't resolve this problem without introduce anaphora resolution capabilities.

Another problematic issue is the recognition of the concept *object*. A lot of terms are both a concept name, and a possible value of the concept *object* (example: hotel, room, equipment, localization, number, time, etc.). This ambiguity causes 15.74% of the deletions, 6.21% of the substitution and 6% of the insertions. If we try to work on the syntactic patterns, we observe a communicating vessels effect, like for the *refLink-coRef* concept. We think that these problems denote one limit of the "knowledge" approaches that probabilistic approaches don't have. However, we also show that our approach gives good results in full mode, i.e. when we have to complete a triplet (concept, value, mode).

Table 2. Results in terms of error rate.

Full		**Relax**	
4 modes	**2 modes**	**4 modes**	**2 modes**
LIMSI-1 (29.0)	LIMSI-2 (23.2)	LIMSI-1 (27.0)	LIMSI-2 (19.6)
LIMSI-2 (30.3)	LIMSI-1 (23.8)	LIMSI-2 (27.2)	LIMSI-1 (21.6)
LIUM (34.7)	LORIA (28.9)	LIA (29.8)	LIA (24.1)
LORIA (36.3)	LIUM (30.0)	LIUM (31.9)	LORIA (24.6)
VALORIA (37.8)	VALORIA (30.6)	LORIA (32.3)	LIUM (26.9)
LIA (41.3)	LIA (36.4)	VALORIA (35.1)	VALORIA (27.6)

5 Conclusions

This article presents a new dialog-oriented model of language: the Reversible Semantic Grammar which uses the same knowledge for analysis as well as generation. Its originality comes from a high integration of syntactical and semantical aspects, which is present in the lexical representation, but also in the algorithms. That's why there is not a real grammatical parser in our system, in opposition to other systems. This allows us to focus on semantics: the lexicon is a collection of concepts which includes syntactic patterns. Our approach can be compared with Gardent's [8]. The common point is that a non-deterministic algorithm is used for the generation. The differences come from the linguistic model, and the way to enforce determinism. We use semantic/pragmatic constraints, whereas Gardent uses grammatical ones. The reason is that we work on spoken dialog, with sentences that are mainly ungrammatical and not well-structured.

Table 3. The nine most important errors sorted by category.

Number	Type	Concept	Percentage
266	DEL	refLink-coRef	26.18 of 1016 DEL
160	DEL	object	15.74 of 1016 DEL
123	INS	connectProp	20.74 of 593 INS
97	INS	refLink-coRef	16.35 of 593 INS
91	SUB	hotel-name	6.98 of 1326 SUB
81	SUB	object	6.21 of 1326 SUB
70	INS	response	11.80 of 593 INS
64	SUB	refLink-coRef	4.94 of 1326 SUB
58	DEL	connectProp	5.70 of 1016 DEL

The results exposed in this article only concern the analyzer. The evaluation is based on the MEDIA corpus which is the French reference dialog corpus with evaluation tools. The linguistic model that we implement, which is a kind of context-free grammar, is pretty simple. Since the scores we obtain are equivalent to those of other systems, this proves that this approach is good enough in our dialog context.

References

1. Tesnière, L.: Éléments de syntaxe structurale. Klincksiek, Paris (1959)
2. Katz, J., Fodor, J.: The structure of a semantic theory. Language 2 (1963) 170–210
3. Kay, M.: Algorithm shemata and data structures in syntactic processing. Technical Report CSL-80-12, Xerox Corporation (1980)
4. Grosz, B., Jones, K., B., W., eds.: Readings in Natural Language Processing. Morgan Kaufmann Publishers Inc (1986)
5. Bonneau-Maynard, H., Rosset, S., Ayache, C., Kuhn, A., Mostefa, D., the MEDIA consortium: Semantic annotation of the french media dialog corpus. In: Proceeding of INTERSPEECH 2005, Lisbon, Portugal (2005) 3457–3460
6. Bonneau-Maynard, H., Ayache, C., Bechet, F., Denis, A., Kuhn, A., Lefevre, F., Mostefa, D., Quignard, M., Rosset, S., Sevran, C., Villaneau, J.: Results of the french evalda-media evaluation campaign for literal understanding. In: Proceedings of LREC 2006, Genoa, Italy (2006) 2054–2059
7. Villaneau, J., Lamprier, S.: Corpus de dialogue homme/machine: annotation sémantique et compréhension. In: Actes des Journées Internationales de Linguistique de Corpus, Lorient, France (2005) 221–229
8. Gardent, C., Kow, E.: A symbolic approach to near-deterministic surface realisation using tree adjoining grammar. In: Proceeding of ACL 2007, Prague, Czech Republic (2007) 328–335