

# SPONTANEOUS SPEECH AND NATURAL LANGUAGE PROCESSING

## ALPES: A ROBUST SEMANTIC-LED PARSER

Jean-Yves Antoine

CLIPS-IMAG

Domaine universitaire, BP 53, F-38041 Grenoble Cedex 9 - FRANCE

Email : Jean-Yves.Antoine@imag.fr

### 1. INTRODUCTION

The need of robust parsers is more and more essential as spoken human-machine communication meets an impressive development. Because of its uncontrolled nature, spontaneous speech presents indeed a high rate of extragrammatical constructions (hesitations, repetitions, self-corrections, etc.). As a result, spontaneous speech rapidly catches out most syntactic parsers, in spite of the frequent addition of some corrective methods [7]. Therefore, most dialog systems restrict the linguistic analysis of the spoken utterances to a simple extraction of keywords [2]. This selective approach led to significant results in some restricted applications (ATIS), but it does not seem appropriate for higher level tasks, for which the utterances cannot be reduced to a simple set of keywords.

As a result, neither the syntactic methods nor the selective approaches can fully satisfy the constraints of robustness and exhaustivity required by the human-machine communication. This paper precisely presents a detailed semantic parser (ALPES) which masters most spoken utterances.

### 2. STRUCTURAL SEMANTICS

Most syntactic formalisms (HPSG [5], TAG [3], etc.) give a major importance to subcategorization, which accounts for the syntactic dependencies inside the sentence. Since extragrammatical spoken utterances remain understandable, we consider on the contrary that subcategorization directly issues from semantics.

Semantic case		Example
DET	determiner	Remove <i>the</i> stairway
AGT	agent	<i>I</i> remove the stairway
ATT	attribute	Remove the <i>left</i> balcony
OBJ	object	Remove the <i>balcony</i>
LOC	location	Place a window on the <i>frontage</i>
OWN	ownership	The <i>floor</i> of the <i>kitchen</i> is tiled
MOD	modality	<i>I want to erase</i> the balcony
INS	instrument	Remove it with <i>the erase</i>
DAT	dative	Give the <i>room</i> a balcony
TPS	time	Save before <i>leaving</i>
SCE	source	Move it from the <i>back</i> to the front
QTE	quantity	There are <i>two</i> stairs
BUT	goal	Move it to <i>place</i> the stairway
COO	coordination	Draw a <i>door and</i> a <i>window</i>
REF	anaphora	Remove the <i>door</i> which is there

Table 1 — Semantic cases (linked words are italicized).

Thus, ALPES aims at building alone a semantic parse tree which fully describes the meaning dependencies inside the sentence. The corresponding relations are labeled by several semantic cases (Table 1) which only intend to cover the system's application field (computer-aided drawing).

ALPES achieves a fully lexicalized analysis. Indeed it relies on a semantic lexicon in which every lexeme — lexical unit of meaning — is described by a features structure (figure 1):

PRED     lexeme identifier  
MORPH    morphological realizations  
SEM       semantic domain  
SUBCAT   semantic subcategorization frame

```

Pred = 'to draw'
Morph = {'draw', 'draws', 'drew', 'drawn'}
Subcat = [
  AGT = / element / + / animate /
  OBJ = / element / + / concrete /
  (LOC) = / property / + / on_element / + / place /
]
Sem = / task - domain /

```

Figure 1 — Semantic features structure of "to draw"

The subcategorization frames describe the meaning dependencies governed by the lexeme. The corresponding arguments are not ordered. Optional arguments are in brackets, by opposition to the compulsory ones. Adverbial phrases are not subcategorized.

### 3. SEMANTIC PRIMING

Any speech recognition system involves a high perplexity which requires the definition of top-down parsing constraints. This is why we based the semantic parsing on a priming process.

#### 3.1. Priming Process

Semantic priming is a predictive process where already uttered words (*priming words*) are calling other ones (*primed words*) through various meaning associations. It aims to a double goal:

- it constrains the speech recognition system.
- it characterizes the whole meaning dependencies inside the sentence.

Each priming step involves two successive processes. At first, the *contextual adaptation* favors the priming words which match with the semantic context. This context is modeled by several semantic

fields. Then, the *relational priming* identifies the lexemes which share a semantic relation with one of the already uttered words. Those relations directly issue from the subcategorization frames of those priming words.

### 3.2. Priming Network

The priming process is carried out by an associative multi-layered network (figure 2) which results from the compilation of the lexicon. Each cell of the network corresponds to a specific lexeme. The inputs represent the priming words. Their activities are propagated up to the output layer which corresponds to the primed words. An additional layer handles the coordinations and the prepositions.

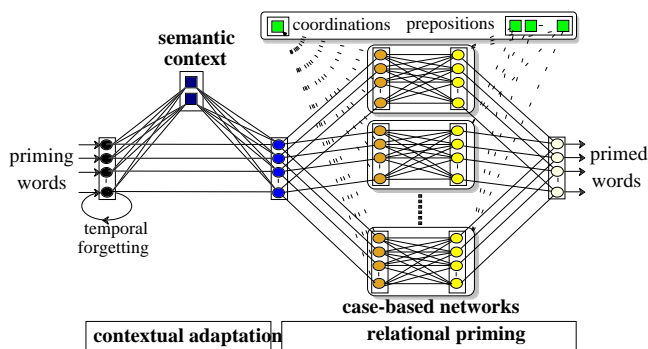


Figure 2 — Structure of the associative priming network

**Temporal forgetting** — At the beginning, the input activities are slightly modulated by a process of temporal forgetting. Although it favors the most recent lexemes, this process does not prevent long distance primings.

**Contextual adaptation** — The contextual adaptation is achieved by the first three layers. Each cell of the second layer represents a precise semantic field. Its activity depends on the semantic affiliations of the priming words. As a result, the current semantic context corresponds to the higher activated semantic field. These contextual cells finally modulate the initial priming activities by means of activatory or inhibitory links. Therefore, the priming words which match with the current context are favored.

**Relational Priming** — Then, the priming activities are dispatched among several subnetworks which perform parallel primings on distinct semantic cases (figure 2). The inputs of those subnetworks represent the priming powers on each semantic case. The dispatching weights are dynamically adapted during the parsing (see § 4), while the inner synaptic weights of the subnetworks issue from the subcategorisation frames of the priming lexemes. The outputs are finally calculated through a maximum heuristic. Then, the primed words — highest outputs — correspond to the possible arguments of the priming ones.

The primed words constrain the speech recognition, thereby warranting the semantic coherence of the analysis. Recognized words are finally handled by the parsing process (see section 4).

### 3.3. Prepositions

On the whole, prepositions restrict the semantic assignment of the objects they introduce. As a result, the prepositional cells should modulate the case-based dispatching weights to prevent any incoherent priming. At last, the preposition is assigned the TAG argument of the prepositional object.

### 3.4. Coordinations

ALPES only deals with logical coordinations (and, or, but...). In such cases, the coordinated words always share the same case:

(a) *Draw a circle and a square*

This constraint is worked out by the recall of the already fulfilled semantic relations: every time a coordination occurs, the dispatching step is restricted to the recalled relations. The coordinated words are finally considered the COO arguments of the conjunction, which is assigned to the shared semantic case:

$$\left[ \begin{array}{l} \text{Pred} = \text{'to draw'} \\ \text{COO} = \left[ \begin{array}{l} \text{Pred} = \text{'circle'} \\ \text{DET} = [\text{Pred} = \text{'a'}] \end{array} \right] \\ \text{OBJ} = \text{Pred} = \text{'and'} \\ \text{COO} = \left[ \begin{array}{l} \text{Pred} = \text{'square'} \\ \text{DET} = [\text{Pred} = \text{'a'}] \end{array} \right] \end{array} \right]$$

Figure 3 — Coordination: semantic structure of the utterance (a).

### 3.5. Back Priming

In some cases, a primed lexeme might occur before its priming word. In (b), *small* is for instance back primed by *window*.

(b) *I want to enlarge the small window*

Such situations are handled through the following algorithm :

- For every new current word:
  1. If it was not primed, it is pushed into a back priming stack.
  2. Otherwise, a specific step of priming checks whether this word back primes some stacked ones. Back primed words are then popped out.

## 4. SEMANTIC PARSING

The parsing process relies on the unification of the subcategorization frames of the recognized lexemes. Four unification principles have been defined:

1. *Unicity* — Any subcategorized argument must at the most be fulfilled by a unique lexeme or by a coordination.
2. *Coherence* — Any lexeme must fulfil at the most a unique subcategorized argument.
3. *Coordination* — Coordinate lexemes must fulfil the same subcategorized argument.
4. *Relative completeness* — Any subcategorized case might remain unfulfilled although the parser must always favor the most complete analysis.

The principle of relative completeness is motivated by the high number of incomplete utterances involved in spontaneous speech. The parser only aims at extracting an unfinished semantic structure pragmatics should then complete. As noticed previously, these principles preventively govern the contextual adaptation of the priming network, so that any incoherent priming is excluded.

## 5. LINGUISTIC ABILITIES

As shown in the following section, ALPES achieves a linguistic coverage which noticeably meets the needs of the spoken human-machine communication. Its suitability is however still more patent when considering spontaneous speech (section 5.2.).

### 5.1. Linguistic Coverage

ALPES is mainly dedicated to French applications. We will study several French linguistic phenomena easily mastered by the parser.

**Compound tenses and passive mode** — According to the semantic point of view, the auxiliaries appear as a mark of modality of the verb. As a result, the parser considers any auxiliary as an ordinary MOD argument of the verb, with the VP corresponding to a compound tense (c) or a passive (d).

- |     |  |  |
|-----|--|--|
| (c) | <i>J'ai mangé</i><br>*I has eaten.<br>I ate.       | $\left[ \begin{array}{l} \text{Pred} = \text{'manger' } \\ \text{MOD} = [\text{Pred} = \text{'avoir' } ] \\ \text{AGT} = [\text{Pred} = \text{'je' } ] \end{array} \right]$  |
| (d) | <i>Le carré est effacé</i><br>The square is erased | $\left[ \begin{array}{l} \text{OBJ} = \left[ \begin{array}{l} \text{Pred} = \text{'carré' } \\ \text{DET} = [\text{Pred} = \text{'le' } ] \end{array} \right] \\ \text{Pred} = \text{'effacer' } \\ \text{MOD} = [\text{Pred} = \text{'ê trê' } ] \end{array} \right]$ |

**Interrogations** — Three interrogative forms are found in French:

- |     |                   |   |
|-----|-------------------|---|
| (e) | subject inversion | <i>Déplace-t-on le carré ?</i><br>* <i>Move we the square?</i>                    |
| (f) | est-ce que        | <i>Est-ce qu'on déplace le carré ?</i><br>* <i>Is-it-that we move the square?</i> |
| (g) | intonation        | <i>On déplace le carré ?</i><br>* <i>We move the square?</i>                      |

Since ALPES ignores word-order considerations, these questions are ordinarily processed like any declarative utterance. At last, closed questions are characterized either by the prosodic analysis (e,g) or by the adverbial phrase *est-ce que* (f) which represents a mark of interrogative modality (MOD argument of the verb).

- (h) *On déplace le carré où ?*  
 \* *We move the square where?*

On the contrary, open questions (h) are explicitly marked by an interrogative pronoun which stands for the missing argument:

$$\left[ \begin{array}{l} \text{LOC} = [\text{Pr ed} = \text{'où' } ] \\ \text{AGT} = [\text{Pr ed} = \text{'nous' } ] \\ \text{PRED} = \text{'déplacer' } \\ \text{OBJ} = \left[ \begin{array}{l} \text{PRED} = \text{'carré' } \\ \text{DET} = [\text{Pr ed} = \text{'le' } ] \end{array} \right] \end{array} \right]$$

**Relative clauses** — Every relative clause is considered as an argument of the lexeme the relative pronoun refers to. However, the parse trees of the main and the relative clauses are kept distinct to respect the principle of coherence. The corresponding structures are indirectly related by an anaphoric relation (REF):

- |     |                                  |   |
|-----|----------------------------------|---|
| (i) | <i>It encumbers the door</i>     | $\left[ \begin{array}{l} \text{AGT} = [\text{Pr ed} = \text{'it' } ] \\ \text{Pr ed} = \text{'to encumber' } \\ \text{OBJ} = \left[ \begin{array}{l} \text{Pr ed} = \text{'door' } = [1] \\ \text{DET} = [\text{Pr ed} = \text{'the' } ] \end{array} \right] \end{array} \right]$ |
|     | <i>...which is on the right.</i> | $\left[ \begin{array}{l} \text{OBJ} = \left[ \begin{array}{l} \text{Pr ed} = \text{'that' } \\ \text{REF} = [1] \end{array} \right] \\ \text{Pr ed} = [\text{Pr ed} = \text{'is' } ] \\ \text{LOC} = [\text{Pr ed} = \text{'on the right' } ] \end{array} \right]$                |

**Subordinate clauses** — Provided the dependent clause is not a relative one, the subordinate verb is subcategorized by the main one. Therefore, every subordinate clause is considered as an ordinary argument of the verb.

### 5.2. Spontaneous Speech

Since it does not take into account syntactic considerations, our semantic parser masters most of the spontaneous ungrammatical constructions.

**Repetitions and self-corrections** — Repetitions (j) and self-corrections (k) seem to violate the principle of unicity. Indeed, they involve indeed several lexemes which share the same case:

- (j) \* *Select the device ... the right device.*  
 (k) \* *Close the display ... the window.*

These structures are actually considered as particular coordinations in which the conjunction is missing [4 : 484-495]. They are consequently parsed like any coordination.

**Ellipses and interruptions** — Thanks to the principle of relative completeness, ALPES is able to extract the incomplete parse tree of any interrupted utterance. On the contrary, this principle is deficient for most of the ellipses:

- (l) \* *[Move] The left door here.*

For instance, the upper predicate *to move* is missing in (l). As a result, the parser lacks information to unify *door* and *here*. Such ellipses should nevertheless be recovered at a pragmatic level.

**Comments** — Generally speaking, comments do not share any semantic relation with the sentence they are inserted in:

- (m) \* *Draw a vertical line ... that's it ... on the right.*

In (m), the adverbial phrase *that's it* is thus a pragmatic mark of confirmation which shares no meaning relation with the main clause. As a result, ALPES can not unify the inserted comment and the main clause. However, we expect further studies on pragmatic marks to enhance the parsing of these constructions.

## 6. EXPERIMENTAL RESULTS

This section presents several experiments which were carried out on ALPES as well as on *Ln2\_3*, a well-known French LFG parser [8]. These experiments were achieved on the literal written transcription of three spoken corpora each including around 200 utterances. These corpora correspond all to a task of computer-aided drawing and were obtained through the Wizard of Oz paradigm. The dialogues were totally unconstrained, so that the utterances correspond to natural spontaneous speech.

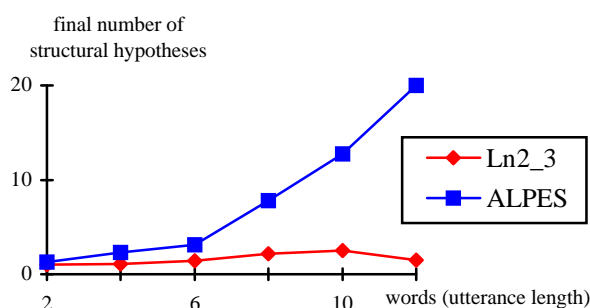
### 6.1. Robustness

Table 2 provides the accuracy rates of the two parsers. These results show the benefits of our approach. Around four utterances out of five ( $\bar{x} = 83.5\%$ ) are correctly processed by ALPES whereas *Ln2\_3* accuracy rates are limited to 40% on the two first corpora. Its robustness is noticeably higher on the third corpus, which presents a moderate ratio of ungrammatical utterances.

Parser	corpus 1	corpus 2	corpus 3	$\bar{x}$	$\sigma_n$
LFG	0.408	0.401	0.767	<b>0.525</b>	0.170
ALPES	0.853	0.785	0.866	<b>0.835</b>	0.036

**Table 2** — Average robustness of the studied parsers. Accuracy rate = number of correct analyses / number of utterances.

Obviously, *Ln2\_3* has not been designed for parsing spontaneous speech. Nevertheless, this comparison suggests that our semantic approach is suitable for spontaneous speech. Besides, the independence of semantics from the grammatical shape of the utterances warrants that our parser's robustness remains relatively unaltered in all circumstances (standard deviation  $\sigma_n = 0.036$ ).



**Figure 4** — Structural perplexity of the studied parsers

### 6.2. Structural Perplexity

The structural perplexity values the bottom-up degree of semantic ambiguity. As mentioned above, ALPES ignores most of the word-order considerations. This tolerant approach is motivated by the frequent order violations involved in spontaneous speech, especially in French. It unfortunately leads to a noticeable increase of perplexity (figure 4).

At first, we offered to reduce this perplexity through a cooperation between the semantic analyzer and a LFG parser [1]. Although this cooperation achieves a noticeable reduction of the perplexity, it is ineffective when the LFG parser fails. This is why we now intend to add some order constraints which are never violated in spontaneous speech. [6] established that any order rule should be expressed lexically. We consequently suggest to order partially the arguments of every lexical subcategorization. Thus, each frame will be assigned a few equations which will characterize some ordering priorities among its arguments. We expect this solution to restrict the perplexity of the semantic parser without altering its robustness.

## 7. CONCLUSION

In this paper, we argued the structural variability of spontaneous speech prevents its parsing by standard syntactic analyzers. We described ALPES, a semantic-led parser which aims at parsing spontaneous speech without considering syntax. The linguistic coverage of ALPES, as well as its robustness, have clearly shown the benefits of this approach.

## REFERENCES

1. Antoine J.Y., Caillaud B., Caelen J., "Automatic adaptive understanding of spoken language by cooperation of syntactic parsing and semantic priming", ICSLP'94, Yokohama, Japan, 799:802, 1994.
2. Appelt D., Jakson E., "SRI International February 1992 ATIS Benchmark Test Results", 5th DARPA Workshop on Speech and Natural Language, Harriman, NY, Morgan Kaufman Inc., 1992.
3. Joshi A., "The relevance of Tree Adjoining Grammar to generation", in G. Kempen (ed.), "Natural Language Generation", Reidel, Dordrecht, NL, 1987.
4. Levelt W., "Speaking: from intention to articulation", MIT Press, Cambridge, Ma, 1989.
5. Pollard C., Sag I., "Information based syntax and semantics", CSLI Lectures notes, 13, University of Chicago Press, IL, 1987.
6. Rambow O., Joshi A., "A Formal Look at Dependency Grammars and Phrase-Structure Grammars, with Special Consideration of Word-Order Phenomena", in L. Wanner (ed.), "Current Issues in Meaning-Text Theory", Pinter, London, 1994.
7. Seneff S., "Robust Parsing for Spoken Language Systems", ICASSP'92, vol. I, 189-192, San Francisco, CA, 1992.
8. Zweigenbaum P., Ca vazza M., "Deep-text understanding in a restricted domain", COLING'90, Helsinki, Fin., 1990.