Parsing spoken language without syntax

Jean-Yves Antoine
CLIPS-IMAG
BP 53 — F-38040 GRENOBLE Cedex 9, FRANCE
Jean-Yves Antoine@imag.fr

Abstract

Parsing spontaneous speech is a difficult task because of the ungrammatical nature of most spoken utterances. To overpass this problem, we propose in this paper to handle the spoken language without considering syntax. We describe thus a microsemantic parser which is uniquely based on an associative network of semantic priming. Experimental results on spontaneous speech show that this parser stands for a robust alternative to standard ones.

1. Introduction

The need of a robust parsing of spontaneous speech is a more and more essential as spoken human - machine communication meets a really impressive development. Now, the extreme structural variability of the spoken language balks seriously the attainment of such an objective. Because of its dynamic and uncontrolled nature, spontaneous speech presents indeed a high rate of ungrammatical constructions (hesitations, repetitions, a.s.o...).
As a result, spontaneous speech catch rapidly out most syntactic parsers, in spite of the frequent addition of some *ad hoc* corrective methods [Seneff 92]. Most speech systems exclude therefore a complete syntactic parsing of the sentence. They on the contrary restrict the analysis to a simple keywords extraction [Appelt 92]. This selective approach led to significant results in some restricted applications (ATIS...). It seems however unlikely that it is appropriate for higher level tasks, which involve a more complex communication between the user and the computer.
Thus, neither the syntactic methods nor the selective approaches can fully satisfy the constraints of robustness and of exhaustivity spoken human-machine communication needs. This paper presents a detailed semantic parser which masters most spoken utterances. In a first part, we describe the semantic knowledge our parser relies on. We then detail its implementation. Experimental results, which suggest the suitability of this model, are finally provided.

2. Microsemantics

Most syntactic formalisms (LFG [Bresnan 82], HPSG [Pollard 87], TAG [Joshi 87]) give a major importance to subcategorization, which accounts for the grammatical dependencies inside the sentence. We consider on the contrary that subcategorization issue from a lexical semantic knowledge we will further name *microsemantics* [Rastier 94].

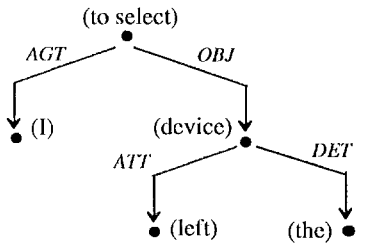

Figure 1: Microsemantic structure of the sentence I select the left device

Our parser aims thus at building a *microsemantic structure* (figure 1) which fully describes the meaning dependencies inside the sentence. The corresponding relations are labeled by several microsemantic cases (Table 1) which only intend to cover the system's application field (computer-helped drawing).
The microsemantic parser achieves a fully lexicalized analysis. It relies indeed on a microsemantic lexicon in which every input represents a peculiar lexeme[1]. Each lexeme is described by the following features structure :

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

[1] Lexeme = lexical unit of meaning.

Example : *to draw*

$$\begin{bmatrix} Pred = \text{'to draw'} \\ Morph = \{\text{'draw','draws','drew','drawn'}\} \\ Subcat = \begin{bmatrix} AGT = / element / + / animate / \\ OBJ = / element / + / concrete / \\ (LOC) = / property / + / place / \end{bmatrix} \\ Sem = / task-domain / \end{bmatrix}$$

The microsemantic subcategorization frames describe the meaning dependencies the lexeme dominate. Their arguments are not ordered. The optional arguments are in brackets, by opposition with the compulsory ones. At last, the adverbial phrases are not subcategorized.

Table 1 : Some examples of microsemantic cases.

Label	Semantic case
DET	determiner
AGT	agent
ATT	attribute
OBJ	object / theme
LOC	location / destination
OWN	meronomy / ownership
MOD	modality
INS	instrument
COO	coordination
TAG	case marker (préposition)
REF	anaphoric reference

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 microsemantic parsing on a priming process.

3.1. Priming process

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

- It constrains the speech recognition.
- It characterizes the meaning dependencies inside the sentence.

Each priming step involves two successive processes. At first, the *contextual adaptation* favors the priming words which are consistent with the semantic context. The latter is roughly modeled by two semantic fields: the task domain and the computing domain. On the other hand, the *relational priming* identifies the lexemes which share a microsemantic relation with one of the already uttered words. These relations issue directly from the subcategorization frames of these 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 (*Structural layer S*) handles furthermore the coordinations and the prepositions.

We will now describe the propagation of the priming activities. Let us consider :

- t current step of analysis
- $a_i^j(t)$ activity of the cell j of the layer i at step t ($i \in \{1, 2, 3, 4, 5, 6, S\}$)
- $\omega_{i,j}^{k,l}(t)$ synaptic weight between the cell k of the layer i and the cell l of the layer j.

Temporal forgetting — At first, the input activities are slightly modulated by a process of temporal forgetting :

$a_1^i(t) = a_{max}$ if i is to the current word.

$a_1^i(t) = a_{max}$ if i is to the primer of thisword.

$a_1^i(t) = \text{Max}\ (0,\ a_1^i(t-1) - \Delta_{forget})$ otherwise.

Although it favors the most recent lexemes, this process does not prevent long distance primings.

Contextual adaptation — Each cell of the second layer represents a peculiar semantic field. Its activity depends on the semantic affiliations of the priming words :

$$a_2^j(t) = \sum_i \omega_{1,2}^{i,j}(t) \cdot a_1^i(t) \qquad (1)$$

with : $\omega_{1,2}^{i,j}(t) = \omega_{max}$ if i belongs to j.

$\omega_{1,2}^{i,j}(t) = -\omega_{max}$ otherwise.

Then, these contextual cells modulate the initial priming activities :

$$a_3^j(t) = a_1^j(t) + \sum_i \omega_{2,3}^{i,j}(t) \cdot a_2^i(t)$$

with : $\omega_{2,3}^{i,j}(t) = \Delta_{context}$ if j belongs to i.

$\omega_{2,3}^{i,j}(t) = -\Delta_{context}$ otherwise.

The priming words which are consistent with the current semantic context are therefore favored.

Relational Priming — The priming activities are then dispatched among several sub-networks which perform parallel analyses on distinct cases (fig. 3). The dispatched activities represents therefore the priming power of the priming lexemes on each microsemantic case :

$$a^i_{4\alpha}(t) = \sum_j \omega^{i,j}_{3,4}(t) \cdot a^j_3(t) = \omega^{i,i}_{3,4}(t) \cdot a^i_3(t)$$

The dispatching weights are dynamically adapted during the parsing (see section 4). Their initial values issue from the compilation of the lexical subcategorization frames :

$\omega^{i,i}_{3,4}(0) = \delta_{max}$ if the case α corresponds to a compulsory argument of the lexeme i or if the latter should fulfill α alone.

$\omega^{i,i}_{3,4}(0) = \delta_{min}$ if the case α corresponds to an optional argument of i or if the latter should fulfill α thanks to a preposition.

$\omega^{i,i}_{3,4\alpha}(0) = 0$ otherwise.

$\omega^{i,j}_{3,4\alpha}(0) = 0$ if $i \neq j$

The inner synaptic weights of the case-based sub-networks represent the relations between the priming and the primed words :

$\omega^{i,j}_{4\alpha,5\alpha}(t) = \omega_{max}$ if i and j share a microsemantic relation which corresponds to the case α.

$\omega^{i,j}_{4\alpha,5\alpha}(t) = \omega_{min}$ otherwise.

The outputs of the case-based sub-networks, as well as the final priming excitations, are then calculated through a maximum heuristic :

$$a^j_{5\alpha}(t) = \text{Max}_{(3)^i} \left(\omega^{i,j}_{4,5\alpha}(t) \cdot a^j_{4\alpha}(t) \right) \quad (4)$$

$$a^i_6(t) = \text{Max}_i \left(a^i_{5\alpha}(t) \right) \quad (5)$$

The lexical units are finally sorted in three coarse sets :

$a^i_6(t) > T_{high}$ primed words

$T_{high} > a^i_6(t) > T_{low}$ primable words

$a^i_6(t) < T_{low}$ rejected words

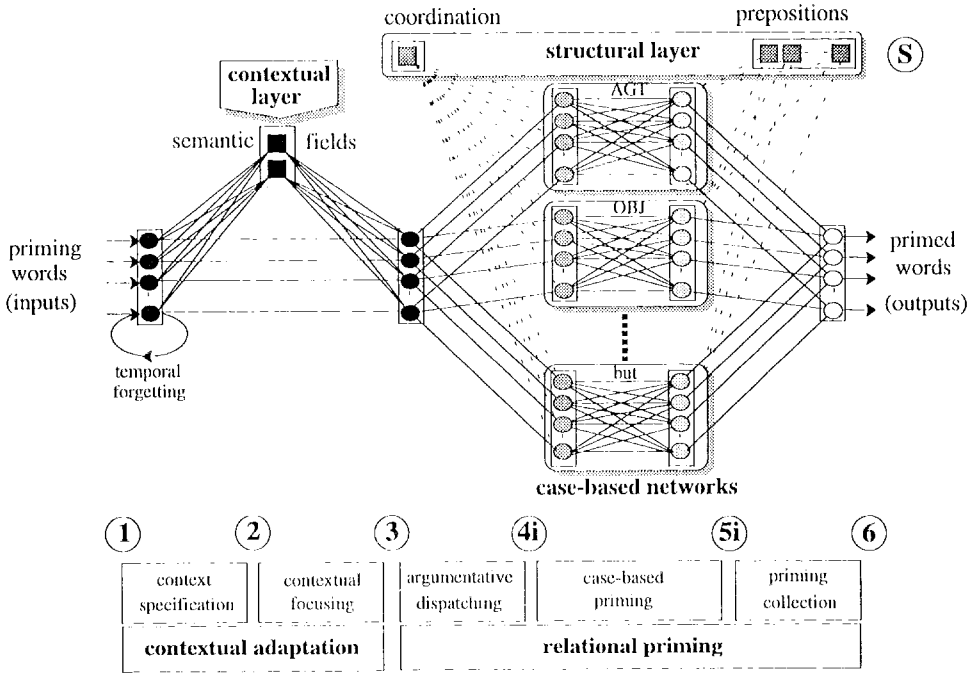

Figure 2 – Structure of the priming network

The primed words aims at constraining the speech recognition, thereby warranting the semantic coherence of the analysis. These constraints can be relaxed by considering the primable words. Every recognized word is finally handled by the parsing process with its priming relation (see section 4).

3.3. Prepositions

Prepositions restrict the microsemantic assignment of the objects they introduce. As a result, the prepositional cells of the *structural layer* modulate dynamically the case-based

dispatching weights to prohibit any inconsistent priming. The rule (3') stands therefore for (3) :

$$a^i_{4\alpha}(t) = \omega^{i,i}_{3,4\alpha}(t) \cdot \left(\sum_k \omega^k_\alpha(t) \cdot a^k_S(t) \right) \cdot a^i_3(t) \quad (3')$$

with : $\omega^k_\alpha(t) = \omega_{max}$ if α is consistent with the preposition k.

$\omega^k_\alpha(t) = 0$ otherwise.

and : $a^k_S(t) = a_{max}$ while the object of k is not assigned a case.

$a^k_S(t) = 0$ otherwise.

At last, the preposition is assigned the TAG argument of the introduced object.

3.4. Coordinations

The parser deals only for the moment being with logical coordinations (and, or, but...). In such cases, the coordinated elements must share the same microsemantic case. This constraint is worked out by the recall of the already fulfilled microsemantic relations, which were all previously stacked. The dispatching is thus restricted to the recalled relations every time a coordination occurs :

$\omega^{i,j}_{3,4\alpha}(t) = \omega^{i,j}_{3,4\alpha}(0)$ for a stacked relation

$\omega^{i,j}_{3,4\alpha}(t) = 0$ otherwise.

The coordinate words are finally considered the COO arguments of the conjunction, which is assigned to the shared microsemantic case.

3.5. Back priming

Generally speaking, the priming process provides a set of words that should follow the already uttered lexemes. In some cases, a lexeme might however occur before its priming word :

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

Back priming situations are handled through the following algorithm :

> Every time a new word occurs :
> 1. If this word was not primed, it is pushed it in a back priming stack.
> 2. Otherwise, one checks whether this word back primes some stacked ones. Back primed words are then popped out.

4. Microsemantic parsing

4.1. Unification

The microsemantic parsing relies on the unification of the subcategorization frames of the lexemes that are progressively recognized. This unification must respect four principles :

Unicity — Any argument must be at the most fulfilled by a unique lexeme or a coordination.

Coherence — Any lexeme must fulfil at the most a unique argument.

Coordination — Coordinate lexemes must fulfil the same subcategorized argument.

Relative completeness — Any argument might remain unfulfilled although the parser must always favor the more complete analysis.

The principle of relative completeness is motivated by the high frequency of incomplete utterances (ellipses, interruptions...) spontaneous speech involves. The parser aims only at extracting an unfinished microsemantic structure pragmatics should then complete. As noticed previously with the coordinations, these principles govern preventively the contextual adaptation of the network weights, so that any incoherent priming is excluded.

5. LINGUISTIC ABILITIES

As illustrated by the previous example, the microsemantic parser masters rather complex sentences. The study of its linguistic abilities offers a persuasive view of its structural power.

5.1. Linguistic coverage

Although our parser is dedicated to French applications, we expect our semantic approach to be easily extended to other languages. We will now study several linguistic phenomena the parser masters easily.

Compound tenses and passive — According to the microsemantic point of view, the auxiliaries appear as a mark of modality of the verb. As a result, the parser considers ordinarily any auxiliary an ordinary MOD argument of the verb.

(d) *J'ai mangé*
**I has eaten.*
I ate.

$$\begin{bmatrix} \text{Pred} = \text{'manger'} \\ \text{MOD} = [\text{Pred} = \text{'avoir'}] \\ \text{AGT} = [\text{Pred} = \text{'je'}] \end{bmatrix}$$

(e) *Le carré est effacé*
The square is erased

$$\begin{bmatrix} \text{OBJ} = \begin{bmatrix} \text{Pred} = \text{'carré'} \\ \text{DET} = [\text{Pred} = \text{'le'}] \end{bmatrix} \\ \text{Pred} = \text{'effacer'} \\ \text{MOD} = [\text{Pred} = \text{'ê tre}] \\ \text{AGT} = \begin{bmatrix} \text{Pred} = \text{'logiciel'} \\ \text{DET} = [\text{Pred} = \text{'le'}] \\ \text{TAG} = \text{'par'} \end{bmatrix} \end{bmatrix}$$

Interrogations — Three interrogative forms are met in French : subject inversion (f1), *est-ce-que* questions (f2) and intonative questions (f3).

(f1) *déplaçons nous le carré ?*
(f2) *est-ce-que nous déplaçons le carré ?*
(f3) *nous déplacçons le carré ?*

Since the parser ignores most word-order considerations, the interrogative utterances are processed like any declarative ones. This approach suits perfectly to spontaneous speech, which rarely involves a subject inversion. Closed questions are consequently characterized either by a prosodic analysis or by the adverbial phrase *est-ce-que*.

(g) *où déplaçons nous le carré ?*

Open questions (g) are on the contrary introduced explicitly by an interrogative pronoun which stands for the missing argument.

Relative clauses — Every relative clause is considered an argument of the lexeme the relative pronoun refers to.

(h) *It encumbers the window which is here*

The microsemantic structures of the main and the relative clauses are however kept distinct to respect the principle of coherence. The two parse trees are indirectly related by an anaphoric relation (REF) .

Subordinate clauses — Provided the dependent clause is not a relative one, the subordinate verb is subcategorized by the main one.

(i) *Draw a circle as soon as the square is erased*

As a result, subordinate clauses are parsed like any ordinary object.

5.2. Spontaneous constructions

The suitability of the semantic parser is really patent when considering spontaneous speech. The parser masters indeed most of the spontaneous ungrammatical constructions without any specific mechanism :

Repetitions and self-corrections — Repetitions and self-corrections seem to violate the principle of unicity. They involve indeed several lexemes which share the same microsemantic case :

(l_1) **Select the device ... the right device.*
(l_2) **Close the display ... the window.*

These constructions are actually considered a peculiar coordination where the conjunction is missing [De Smedt 87]. Then, they are parsed like any coordination.

Ellipses and interruptions — The principle of relative completeness is mainly designed for the ellipses and the interruptions. Our parser is thus able to extract alone the incomplete structure of any interrupted utterance. On the contrary, the criterion of relative completeness is deficient for most of the ellipses like (t), where the upper predicate *to move* is omitted :

(n) * [Move] *The left door on the right too.*

Such wide ellipses should nevertheless be recovered at a upper pragmatic level.

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

(o) * *Draw a line ... that's it ... on the right..*

For instance, the idiomatic phrase *that's it* is related to (o) at the pragmatic level and not at the semantic one. As a result, the microsemantic parser can not unify the main clause and the comment. We expect however further studies on pragmatic marks to enhance the parsing of these constructions. Despite this weakness, the robustness of the microsemantic parser is already substantial. The following experimental results will thus suggest the suitability of our model for spontaneous speech parsing.

6. Results

This section presents several experiments that were c arried out on our microsemantic analyzer as well as on a LFG parser [Zweigenbaum 91]. These experiments were achieved on the literal written transcription of three corpora of spontaneous speech (table 2) which all correspond to a collaborative task of drawing between two human subjects (wizard of Oz experiment).

Table 2 : Description of the experimental corpora.

Corpus	Number of utterances	Average length of utterances
corpus 1	260	11.8
corpus 2	157	11.3
corpus 3	179	5.7

The dialogues were totally unconstrained, so that the corpora are corresponding to natural

spontaneous speech. We compared the two parser according on their robustness and their perplexity.

6.1. Robustness

The table 3 provides the accuracy rates of the two parsers. These results show the benefits of our approach. Around four utterances over five ($\overline{x}$=83.5%) are indeed processed correctly by the microsemantic parser whereas the LFG's accuracy is 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. The overall performances of the LFG suggest nevertheless that a syntactic approach is not suitable for spontaneous speech, by opposition with the microsemantic one.

Table 3 : Average robustness of the LFG and the microsemantic. Accuracy rate = number of correct analyses / number of tested utterances.

Parser	corpus 1	corpus 2	corpus 3	$\overline{x}$	σ_n
LFG	0.408	0.401	0.767	**0.525**	0.170
Semantics	0.853	0.785	0.866	**0.835**	0.036

Besides, the independence of microsemantics from the grammatical shape of the utterances warrants its robustness remains relatively unaltered (standard deviation σ_n = 0.036).

6.2. Perplexity

As mentioned above, the microsemantic parser ignores in a large extent most of the constraints of linear precedence. This tolerant approach is motivated by the frequent ordering violations spontaneous speech involves. It however leads to a noticeable increase of perplexity. This deterioration is particularly patent for sentences which include at least eight lexemes (Table 4).

Table 4 : Number of parallel hypothetic structuresl according to utterances' length

Length	LFG parser	Microsemantic
4 words	1,5	2,5
6 words	1,5	3,5
8 words	2	8
10 words	2	12,5
12 words	1,25	19,75

At first, we proposed to reduce this perplexity through a cooperation between the microsemantic analyzer and a LFG parser [Antoine 94]. Although this cooperation achieves a noticeable reduction of the perplexity, it is however ineffective when the LFG parser collapses. This is why we intend at present to insert directly some ordering constraints spontaneous speech never violates. [Rambow 94] established that any ordering rule should be expressed lexically. We suggest consequently to order partially the arguments of every lexical subcategorization. Thus, each frame will be assigned few equations which will characterize some ordering priorities among its arguments.

7. Conclusion

In this paper, we argued the structural variability of spontaneous speech prevents its parsing by standard syntactic analyzers. We have then described a semantic analyzer, based on an associative priming network, which aims at parsing spontaneous speech without considering syntax. The linguistic coverage of this parser, as well as several its robustness, have clearly shown the benefits of this approach. We expect furthermore the insertion of word-order constraints to noticeably decrease the perplexity of the microsemantic analyzer.

References

J.Y. Antoine, J. Caelen, B. Caillaud (1994). "Automatic adaptive understanding of spoken language", ICSLP'94, Yokoham, Japan, 799:802.

D. Appelt, E. Jakson (1992), "SRI International ATIS Benchmark Test Results", 5th DARPA Workshop on Speech and Natural Language, Harriman, NY.

J. Bresnan, J. Kanerva (1989). "Locative inversion in Chichewa", *Linguistic Inquiry*, 20, 1-50.

A. Joshi (1987) "The relevance of TAG to generation", in G. Kempen (ed.), "Natural Language Generation", Reidel, Dordrecht, NL.

W. Levelt (1989). "Speaking : from intention to articulation", MIT Press, Cambridge, Ma.

C. Pollard, I. Sag (1987), "Information based syntax and semantics", CSLI Lectures notes, 13, University of Chicago Press, IL.

O. Rambow, A. Joshi (1994). "A Formal Look at Dependancy Grammars and Phrase-Structure Grammars ", in L. Wanner (ed.), "Current Issues in Meaning-Text Theory", Pinter, London, 1994.

F. Rastier et al (1994). "Sémantique pour l'analyse", Masson, Paris.

S. Seneff (1992). "Robust Parsing for Spoken Language Systems", ICASSP'92, vol. I, 189-192, San Francisco, CA.