

FLEXIBLE PARSING OF DISCRETELY
 UTTERED SENTENCES

Luigi Borghesi Chiara Favareto
 Elettronica San Giorgio - ELSAG S.p.A.
 Via Hermada, 6
 16154 Genova
 I T A L Y

In this paper we describe a syntactic semantic parser of spoken sentences pertaining to a subset of natural Italian language. Error-free and fast analysis, partial interpretation ability, man-machine dialogue trend, different semantic environment adaptability and natural language usage are its main characteristics. All of these features are supported by a technique of input reliability evaluation. Particular attention is devoted to the description of the knowledge internal representation and of the mechanism that manages, at different points of the analysis, the whole process.

1. PARSER QUALIFICATIONS

The parser to be described performs its analysis starting from an intrinsically unreliable input that is the result of an isolated word speech recognizer. The lack of certainty on the single items of the input sentence is one of the main problems in such a vocal parser. The representation of each uttered word, following the recognition stage, is, in fact, an ordered list of possible interpretations with associated dissimilarity measures. As a consequence, it is possible to have doubts not only about every single word of the sentence, but also on complete sentence parts. Moreover, irrecoverable recognition errors may require the capability of parsing incomplete sentences.

Spoken Sentences: TOGLI TUTTO DALLA STANZA

Parsing Input :

word	dis	word	dis	word	dis	word	dis
TOGLI	32.1	UN	46.4	DALLA	16.7	RAGGIO	42.4
OTTO	34.7	QUANTO	46.8	DAL	19.4	NOVANTA	42.8
DAMMI	34.8	UNA	46.9	AL	20.9	STANZA	42.9
TOGLIERE	35.1	VENTI	47.0	TAVOLO	23.7	QUADRATO	44.1
COSTRUISCI	35.9	OTTO	48.1				
		UND	48.5				
		TUTTO	49.5				

Figure 1
 EXAMPLE OF A TYPICAL INPUT OF THE PARSER

Fig. 1 shows an example of a typical parsing input where each input word is replaced by a complete list of possible alternatives with associated distance score. It

is interesting to notice that not only the sentence that was actually spoken "T0-GLI TUTTO DALLA STANZA" (= remove everything from the room), is reported but, in this case, also some other correct sentences can be found by the parser (for example "COSTRUISCI UN TAVOLO QUADRATO" (build a square table).

An efficient parser must also be able to solve other problems not strictly connected to a particular kind of input. In fact it should, of course, achieve fast operations; that requires the ability to minimize the number of alternative parses[1].

Furthermore the parser should be designed in such a way as to satisfy the "generality" expectations; that is it should be easily adaptable to any semantic domain at least in the limited semantic domain cases. Since the parser results should be followed by the execution of some operation in any practical application it is required that it produced trustworthy results and, in particular, that it always included the right sentence interpretation within all the output ones.

Finally, to allow a graceful dialogue with its users the parser must be able to analyse also partial sentences (for example elliptical or fragmentary ones), thus making it possible to use naturally expressed sentences [2].

2. MAIN PARSER'S CHARACTERISTICS

The main features of our parser, that permit to satisfy the above mentioned requirements, are the following:

- 1) representation of the language in terms of a network whose elements are syntactic groups and syntactic features;
- 2) definition of a confidence measure of the recognition results and its extension also to groups of words (syntactic groups);
- 3) adoption of a recursive working strategy which anchors the parsing on the most reliable words in a first step and on the most reliable groups in a second one.

We selected the furnishing of a living room as the discourse domain and we defined a vocabulary of a 116 words.

This vocabulary, although limited, leads to a total number of over 10^5 possible sentences that include commands for constructing or moving pieces of furniture, assignment of labels, definition of unit lengths, inquiries about mutual distances, etc.

2.1. Language representation

We describe each sentence of the language by a sequence of syntactic groups. A syntactic group is defined as a sentence part with a well precise semantic meaning. Often, but not necessary, a syntactic group corresponds to a classical grammatical object. For example, we defined the verb, the direct object, the location object, etc.

In this way (see Fig. 2) each sentence of the language can be described by a sequence of some of these groups.

On the whole we introduced only 9 groups; in our opinion this set of syntactic groups is enough to describe, at a syntactic level, all the possible sentences pertaining to this semantic environment.

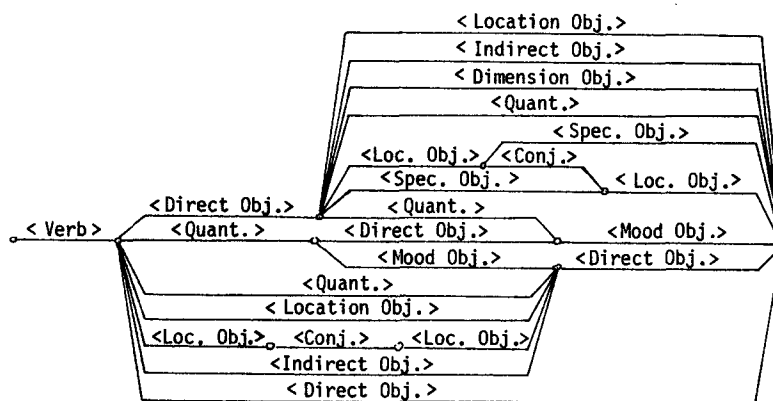


Figure 2

LANGUAGE REPRESENTATION

Each syntactic group is, in turn, represented by a number of possible word sequences, or, more precisely, of sequences of associated syntactic features. Figure 3 shows, for example, how the direct object is represented.

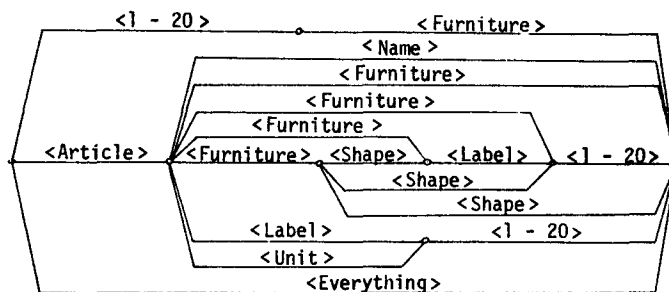


Figure 3

DIRECT OBJECT REPRESENTATION

It is important to notice that one feature can represent more than one word and every new word does not always need a new feature definition. So the present vocabulary can be easily increased to a certain degree within the semantic domain, with out any change in the grammar.

2.2. Reliability evaluation

The doubts connected with the vocal input suggested the need for a tool that measured the goodness of each word recognition. To this purpose a method to evaluate the reliability of recognition results was defined [3].

By this method every word, in the ordered list, is associated with a confidence score indicating its probability of being the correct one (see Fig. 4).

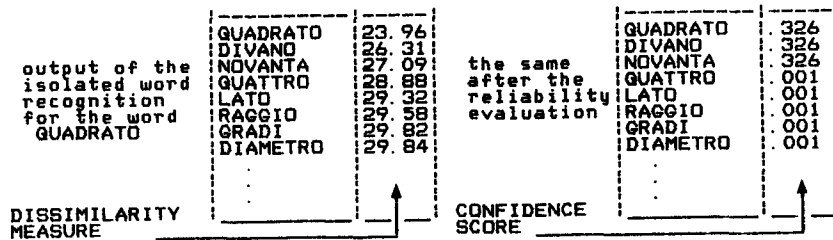


Figure 4

RELIABILITY EVALUATION

In this way, as described below, the most reliable words of the sentence can be selected and the parser anchored to them. The same reliability score is also used to evaluate the syntactic groups found and to decide which, among alternative groups, is the most probable one.

2.3 Island driven working strategy

All the operations of the parser are centered around the concept of reliability score. In fact, in a first step, the parser anchors its analysis to the most reliable word of the sentence (that we named "guide word") and searches, both to the right and to the left of it for all the syntactic groups that include the features associated to the guide word. Each of these syntactic groups is named "island".

Not only the first word in the ordered list can be used for this aim, but sometimes also the second and the third ones are taken into consideration. For each island a cumulative reliability score, function of the single word scores, is computed.

The same procedure is then applied to the remaining words until the whole sentence has been examined and there are no more guide words; at this point a lattice of island, possibly overlapping, is obtained.

In a second step the parser, in an almost identical fashion as before, searches for the most reliable island (that we named "guide island"), anchoring to it the exploration of the language network to get a match with one of the possible sentences. When this is not possible, because of very unreliable recognition of a whole syntactic group, the partial sentence recovered is proposed in output together with a hypothesis about the missing constituent.

At this point a module for graceful man-machine interaction could be activated, in order to obtain the needed information by means of an appropriate dialogue.

In addition there are some parameters, specifying the number of retained alternatives at various points of the parsing, that allow to control parser's performance both in terms of speed and confidence. These parameters allow the parser to work with different degrees of flexibility and so, they must be carefully selected, according to the application, i.e. according to the risk that can be tolerated when accepting an acoustically unclear sentence.

3. RUNNING EXAMPLES

In Fig. 5 the main steps of the analysis of a particular sentence are summarized. To make the comprehension easier we report a simulated english example that corresponds to a real italian sentence processed by the parser.

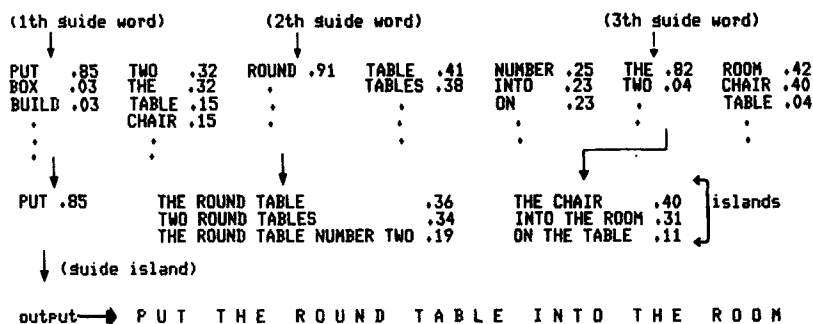


Figure 5
RUNNING SIMULATED EXAMPLE

The input sentence is: PUT THE ROUND TABLE INTO THE ROOM. In the first step, starting respectively from the 1st, 2nd and 3rd guide word (PUT, ROUND, THE), the parser finds some possible islands with associated reliability score. In a second step, starting from the guide island the parser searches a match between a path in the language network and the islands. The final result is the correct interpretation of the sentence even if there were three recognition errors.

Sometimes the parser outputs are not univocal as in the previous example. In fact, if the reliability score of a whole island is too low, the parser provides an output in which, instead of a detailed word-by-word interpretation, an hypothesis about the type of the missing syntactic group appears as shown in the example below:

PUT < direct object > IN THE MIDDLE OF THE ROOM

If, on the contrary, there are two or more words with approximately the same reliability score and the same syntactic role, then the parser supplies in the output those alternatives with their associated reliability scores and the whole decision will be deferred to a following pragmatic module or dialogue component. For example we can have an output like this one:

PUT THE { TABLE .32 } NUMBER TWO IN FRONT OF THE DOOR
 { CHAIR .29 }

4. RESULTS

The present parser has been tested on a set of 50 sentences spoken by three different speakers (two males and one female). A poor word recognizer was adopted in order to stress parser's capabilities. We compare in table 1 the parser performances and those of the recognizer alone. For each speaker the first column reports the percent of success of the recognizer alone, i.e. how many times all the words of a sentence were in the first position. The second column reports the percent of success of the parser (i.e. how many times the parser was able to interpret the sentence).

Each row corresponds to a case in which there are, respectively, none, 1, 2, and 3 lost islands whose reliability was not sufficient to take any decision.

We want to notice that for the 1st speaker the parser locates the correct sentence in the 92% of the cases and achieves a 96% correct interpretation if it assumes that there is one lost island. For the 2nd speaker these values increase more slowly because of a very unreliable input (10% of success for the recognizer).

However the main result is that the parser never took a decision that did not contain the correct interpretation.

# of lost islands	SPEAKER N. 1		SPEAKER N. 2		SPEAKER N. 3	
	recognizer	parser	recognizer	parser	recognizer	parser
0	30	92	10	10	58	82
1		96		54		98
2				88		100
3				94		

Table 1

PERCENT OF SENTENCE RECOGNITION (see text)

5. CONCLUSIONS

We have described a parser of spoken sentences in which the design decisions taken are a necessary step to satisfy the requirements of a voice interactive system.

In fact the choice of using features instead of vocabulary entries and of describing the sentences in terms of syntactic groups agrees with the needs of improving the analysis speed, of making the semantic environment representation oriented to a partial interpretation analysis and of allowing the adaptability to different semantic environments.

Finally we want to notice the characteristics that make this parser oriented to a graceful man-machine interaction. They are its ability to provide in the output more than one choice with its associated reliability score and its ability to interpret also partial input.

The first characteristic, in fact, allows a pragmatic module to make a choice among various alternatives, looking at their reliability score or, if in doubt, to activate a dialogue module which requests the needed information. The second characteristic permits the management of a dialogue in which also elliptical sentences are allowed or, anyhow, the maximum freedom of expression is permitted.

REFERENCES:

- [1] Woods W.A., Syntax, semantics and speech, in: Reddy D.R. (ed.), Speech Recognition (Academic Press, 1975).
- [2] Hayes P., Reddy R., An anatomy of graceful interaction in spoken and written man-machine communication, Report CMU-CS-79-144, Carnegie Mellon Univ. (sept. 1979).
- [3] Musso G., Borghesi L., Favareto C., Confidence evaluation for an Isolated Word Recognizer, 4th F.A.S.E. Proceedings (april 1981) 293-296.