

STRATEGIC SELECTION FOR AN ATN SYNTACTIC PARSER

Giacomo Ferrari and Oliviero Stock

Istituto di Linguistica Computazionale - CNR, Pisa

Performance evaluation in the field of natural language processing is generally recognised as being extremely complex. There are, so far, no pre-established criteria to deal with this problem.

1. It is impossible to measure the merits of a grammar, seen as the component of an analyser, in absolute terms. An "ad hoc" grammar, constructed for a limited set of sentences is, without doubt, more efficient in dealing with those particular sentences than a grammar constructed for a larger set. Therefore, the first rudimentary criterion, when evaluating the relationship between a grammar and a set of sentences, should be to establish whether this grammar is capable of analysing these sentences. This is the determination of linguistic coverage, and necessitates the definition of the linguistic phenomena, independently of the linguistic theory which has been adopted to recognise these phenomena.

2. In addition to its ability to recognise and coherently describe linguistic phenomena, a grammar should be judged by its capacity to resolve ambiguity, to bypass irrelevant errors in the text being analysed, and so on. This aspect of a grammar could be regarded as its "robustness" [P.Hayes, R.Reddy 1979].

3. Examining other aspects of the problem, in the analysis that we propose we will assume a grammar which is capable of dealing with the texts which we will submit to it.

Let an ATN grammar N , with n nodes, be of this type. N will be maintained constant for the following discussion.

By text we intend a series of sentences, or of utterances by one of the speakers in a dialogue. When analysing such a text, once a constant N has been assumed, it is likely that, in addition to the content (the argument of the discourse) indications will appear on the grammatical choices made by the author of the text (or the speaker) when expressing himself on that argument (how the argument is expressed).

When these indications have been adequately quantified, they can be used to correctly select the perceptive strategies (as defined in [Kaplan 72]) to be adopted in order to achieve greater efficiency in the analysis of the following part of the text.

4. For our experiments we have used ATNSYS [Stock 76], and an Italian grammar with $n = 50$ (127 arcs) [Cappelli et al.77]. In this system, search is depth-first and the parser interacts with a heuristic mechanism which orders the arcs according to a probability evaluation. This probability evaluation is dependent on the path which led to the current node and is also a function of the statistical data accumulated during previous analyses of a "coherent" text.

The mechanism can be divided into two stages. The first stage consists of the acquisition of statistical data; i.e. the frequency, for each arc exiting from a node, of the passages across that arc, in relation to the arc of arrival: for each arriving arc there are as many counters as there are exiting arcs.

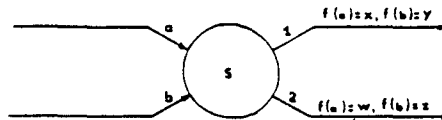


Fig. 1

In this way, in Fig. 1 arc 1 has been crossed x times coming from a and y times coming from b . In the second stage, during parsing, in state S , if coming from a and $w > x$, arc 2 is tried first.

4.1 Thus, a first evaluation of the linguistic choices made is provided by the set of probability values associated to each arc. These figures can to some extent describe the "style" of any "coherent" text analysed. (For this one should also take into account the different linguistic significance of each arc. In fact a CAT or PUSH arc directly corresponds to a certain linguistic component, while a JUMP or VIRT arc occurs in relation to the technique by which the network has been built, the linguistic theory adopted, and other variables.)

4.2 The second part of the mechanism, the dynamic reordering of the arcs, coincides with a reordering of the comprehension strategies. In this way, a matrix can be associated to each node, giving the order of the strategies for each arc in arrival.

For each text T, there is a set of strategies S_T , ordered as described above. While the analysis of the probability values for two distinct texts T and T' can give global indications of their linguistic characteristics, if we focus on the comprehension of the sentence, it is more meaningful to give evaluations in relation to the sets of strategies, S_T and $S_{T'}$, which are selected.

Fig. 2 shows, for some nodes, a comparison between the orders of the arcs for the first 11 sentences from two texts, a science fiction novel (SPN, upper boxes) and a handbook of food chemistry (FC, lower boxes). The arc numbers are referred to the order in the original network. The figures which appear after the = in the heading indicate the number of parses for each sentence. An empty box indicates the same order as that shown in the previous box.

STATE	COMING FROM	SPN=10 FC1 =14	SPN=4 FC2 =1	SPN=2 FC3 =10	SPN=1 FC4 =4	SPN=5 FC5 =1	SPN=2 FC6 =5	SPN=4 FC7 =2	SPN=2 FC8 =5	SPN=1 FC9 =4	SPN=4 FC10 =2	SPN=4 FC11 =5	=39 =51
S	S/S2	123				132							
		312											
	T/DART	312											
		312											
S/NOS	GP/GN1	312											
				312									
	S1/RIFL1	51324 52134	51342							51342		52143	
		51234		52134	52143	52134							
S/SOOG1				51234									
	GV/NOG1	41235 41235		45123 43125									
GV/T		4123 4213						4123		4123			
				2413						4213			
GP	GN/PREP1	213											
				123									
	S1	213	213										
				123									
	GV/SLOT1 1	123											
GV/FRAM2 1				123									
		123											

5.1 It is to be expected that this mechanism, in so far as it introduces a heuristics, will increase the efficiency of the system used for the linguistic analysis. The results of our experiments so far confirm this. This improved efficiency can be measured in three ways:

- locally, in terms of the computational load, due to non-determinism, which is saved in each node. In fact, by some experiments, it is possible to quantify the computational load of each type of arc. The computational load of a node is then a linear combination of these values and one can compare it with the actual load determined by the sequence of arcs attempted in that point after the reordering.
- in terms of an overall reduction in computing time;
- in terms of penetrance, i.e. the ratio between the number of choices which actually lead to a solution and the total number of choices made.

5.2 If T is a text containing r sentences, the average penetrance will be:

$$P_m(S_T, T) = \frac{\sum_{i=1}^r P(S_T, s_i)}{r}$$

where s_i stands for each of the sentences in T.

If T is analysed using the set of strategies chosen for a different text, T', then the penetrance is, on average, no greater than with S_T .

$$P_m(S_{T'}, T) \leq P_m(S_T, T)$$

In our experiments, for instance, the average penetrance for the first text (SFN) parsed with its own strategies (S_{SFN}) is $P_m(S_{SFN}, SFN) = 0.52$, while parsed with the strategies of the second text (S_{FC}) is $P_m(S_{FC}, SFN) = 0.39$. We have attempted to evaluate experimentally the relationship between the difference of the average penetrances, which we call discrepancy

$$D(S_T, S_{T'}, T) = P_m(S_T, T) - P_m(S_{T'}, T)$$

and the distance between two sets of strategies. However we think we need more experimentation before formalizing this relationship.

Returning to our science fiction novel, the discrepancy using its set of strategies and the one inferred by the food chemistry text is

$$D(S_{SFN}, S_{FC}, SFN) = 0.13$$

6. In addition to the definition of a heuristic mechanism which is capable of improving the efficiency of natural language processing, and which can be evaluated as described above, our research aims at providing a means to characterise a text by evaluating the grammatical choices made by the author while expressing his argument.

We are also attempting to take into account the expectations of the listener. In our opinion, the listener's expectations are not limited to the argument of the discourse but are also related to the way in which the argument is expressed; this is the equivalent of the choice of a sub-grammar [Kittredge 78]. We intend to verify the existence of such expectations not only in literature or when listening to long speeches, but also in dialogue.

References

1. Cappelli A., Ferrari G., Moretti L., Prodanof I., Stock O.- 'An Experimental ATN Parser for Italian Texts' Technical Report, LI.C-CNR, Pisa 1977.
2. Kaplan R.- 'Augmented Transition Networks as Psychological Models of Sentence Comprehension' Artificial Intelligence 3 1972, Amsterdam - New York - Oxford.
3. Hayes P., Reddy R. - 'An anatomy of Graceful Interaction in Spoken and written Man-Machine Communication', CMU-CS-79-144, Pittsburgh PA, 1979.
4. Kittredge R.- 'Textual Cohesion Within Sublanguages: Implications for Automatic Analysis and Synthesis', COLING 78, Bergen, 1978.
5. Stevens A., Rumelhart D.- 'Errors in Reading: An Analysis Using an Augmented Transition Network Model of Grammar' in Norman D., Rumelhart D. eds., Explorations in Cognition, Freeman, S.Francisco, 1975, pp. 136-155.
6. Stock o. - 'ATNSYS: Un sistema per l'analisi grammaticale automatica delle lingue naturali', NI-B76-29, IEI, Pisa, 1976.

