REVISING AN ATN PARSER

Giacomo Ferrari, Irina Prodanof

Istituto di Linguistica Computazionale, CNR, Pisa, Italy

1. An ATN parser for Italian has been developed and
tested in a series of experiments on complex sentences taken
from narrative texts (1,2,3,4). The construction of a complex
grammar as well as the results of our experiments showed some
limits and inadequacies of Woods´ ATN as it stands (11),
especially when used to parse Italian, a relatively free-
order language with a well developed morphological system.

1.1. ´Parameter-passing´ actions such as SENDR and
LIFTR have been often judged ´dirty´ operations under a form-
al view-point. Moreover it is almost impossible to keep the
control of the cross-level passed information, unless the
number of registers is increased. Such an increase is other-
wise unmotivated. This difficulty is more evident in Italian
relative clauses such as:

L´ uomo   la   cui   crudelta´ tutti   conosciamo
The man (the) whose cruelty   everyone knows
L´ uomo   la crudelta´ del quale tutti   conosciamo
The man   the cruelty of whom   everyone knows
L´ uomo   del desiderio   di vendetta del quale ti ho parlato
The man   about the desire for revenge of whom   I spoke to you

in which a very heavy use of such actions, as well as of
flags, is required. These phenomena, together with the rich-
ness of morphological inflexion, increase the necessity for
morphological and morpho-syntactic short, long-distance and
cross-level tests.

1.2. Another problem is proposed by the structural ambiguity of sequences of PP's, for which both the 'embedding' and the 'same-level' hypotheses are presented; a sentence with 3 PP's has 5 parses according to the schema

(PP1 (PP2 (PP3))), (PP1 (PP2)(PP3)), (PP1 (PP2) PP3),(PP1 PP2 (PP3)), (PP1 PP2 PP3)

An attempt to limit this explosion was made by adding certain functional labels depending upon the verb frame. This experiment showed that attention should be focussed on developing an adequate system of functional tests.

1.3. A statistically based, dynamic heuristic mechanism has been put to work in the traversal of the network. Those paths which have been more frequently traversed in previous parsed fragments of a given text are attempted first, during the analysis of the current sentence (6).

In this way, the concept of performance based syntactic expectation has been introduced and tested. It appeared that ATN is very suited to the introduction of heuristic techniques and also that these techniques should include a scheduling procedure based on arc types (10).

2. A revision of the linguistic formalism as well as of the parsing algorithm is now in progress, and should take all these considerations into account. Furthermore, those linguistic phenomena related to conversation such as anaphora and the analysis of partial and partially correct input must be taken into account in the designing of the formalism and the ATN actions and forms, not only for applicative purposes, but above all in order to enlarge the range of phenomena to be included in a grammar.

2.1. The parser relies both on structural and functional information (7), i.e. a limited collection of surface structures and functional labels. A set of fixed (control) patterns might also be included in this static syntactic knowledge

(8, 9). The output representation, however, is in terms of functional labels. The utilization of these three data types will be fully discussed.

2.2. It is not our intention to substantially modify the general structure of the algorithm. Storing and retrieving will be modified, in accordance with our observations in 1.1. A single register external to the levels of computation will be progressively set to the results of the analysis during parsing. Access to precise location of this structure is guaranteed by the functional labels. Traditional ATN registers can be locally used for special purposes. A detailed description of our whole mechanism including examples will be given.

Already assigned functional labels cannot be changed; left concatenation of them has been preferred (8). A mechanism has also been provided to move pointers from constituents to gaps or anaphorical pronouns (inside the sentence borders) (5). This is an attempt to treat by a uniform action or sequence of actions equi-NP deletion, gapping, coordination.

2.3. Some local procedures connected with the lexicon have been added to these mechanisms. These procedures include the transmission of features to higher level nodes in the structure, the proceduralisation of idiosyncrasies of single lexical entries and ´paraphrase´. The latter is expected to reduce nominalization to the corresponding verb with its frame components and to account for linguistic units which have identical distribution and syntactic behaviour.

2.4. The present list of ´linguistically´ motivated processes is open-ended, as we are now trying to identify others, which can be considered to be linguistic (procedural) ´universals´.

(1) A.Cappelli, G. Ferrari, L. Moretti, I. Prodanof, O. Stock: 'Parsing an Italian text with an ATN parser', report ILC, Pisa, 1978.

(2) A.Cappelli, G. Ferrari, L. Moretti, I. Prodanof, O. Stock: 'An ATN parser for Italian: an experiment', Atti del COLING 78, Stoccolma (microfiches).

(3) A.Cappelli, G. Ferrari, L. Moretti, I. Prodanof, O.Stock: 'Analisi di un testo italiano con un analizzatore ATN', Rivista di Informatica, X 3 (1980), pp. 5-13.

(4) A.Cappelli, G.Ferrari, L. Moretti, I. Prodanof, O. Stock: 'Analisi automatica dell' italiano: alcune osservazioni', in La ricerca di base in psicologia. Atti del XVIII Congresso degli Psicologi Italiani, Acireale 29/X - 2/XI 1979, 1980, pp. 246-267.

(5) A. Cappelli,G.Ferrari, L. Moretti, I.Prodanof, O.Stock 'Il trattamento di alcuni fenomeni anaforici mediante un ATN', Atti del seminario 'L'anafora'tenuto presso l'Accademia della Crusca, Firenze 14-16/12/78. In corso di Stampa.

(6) G.Ferrari, O. Stock: 'Strategie selection for an ATN Syntactic Parser', in The 18th Annual Meeting of the Association for Computational Linguistics and parasession on Topics in Interactive Discourse. Proceedings of the Conference, June 19-22, 1980, Philadelphia, pp. 113-115.

(7) R. Kaplan, J. Bresnan: Lexical-Functional Grammar: A formal system for grammatical representation, to appear in J.Bresnan (ed.), The mental representation of grammatical relations, MIT Press.

(8) M. Kay: 'Functional Grammar', in Proc. of the 5th meeting of the Berkeley Linguistic Society, Berkeley 1979, pp. 142-158.

(9) S.Kwasny, N. Sondheimer: 'Relaxation techniques for parsing ill-formed input', in AJCL 7, 2 (april-june 1981),

pp. 99-108.

(10) E. Wanner, M. Maratsos: ´An ATN approach to comprehension´
in M. Halle, J. Bresnan, G. Miller (eds.) Linguistic
Theory and Psychological Reality, MIT Press 1978.

(11) W. Woods: ´Transition Network Grammars for Natural lang-
uage Analysis´, in CACM 13-10, New York, 1970.