# "FREE" ORDER LANGUAGES: AN EXPERIMENTAL LEXICON BASED PARSER

Cristiano Castelfranchi, Domenico Parisi, Oliviero Stock

Istituto di Psicologia, Consiglio Nazionale delle Ricerche
Via dei Monti Tiburtini, 509, 00157-Roma

Most of the work on natural language understanding (NLU)
has been done on English. English is a language with relative-
ly rigid word order, a characteristic that has influenced all
NLU systems proposed so far. These systems have dedicated only
minor attention to problems that are of major importance in
languages with a freer ordering like Italian. Work on parsing
partially ungrammatical sentences in rigid languages (Hayes
and Mouradian, 1980; Charniak, 1981) bears some similarity to
work on parsing of "free" order language. In both cases an
exclusively top-down model seems inappropriate. For instance,
when considering an incomplete sentence it may still be advis-
able to proceed in building some representation. The result-
ing structure will then be inserted within a larger cognitive
structure. The same bottom-up way of proceeding must necessa-
rily be part of an NLU system for "free" order languages.

An ATN (Woods, 1970) type system, like the one developed
for Italian (Cappelli et al., 1978), shows definite limitat-
ions even when it is furnished with heuristics for strategy
selection (Ferrari and Stock, 1980) based on adaptation to a
coherent text. Other ideas such as passing information in the
case of failure (Weischedel and Black, 1980) or relaxing con-
ditions on arcs in certain circumstances (Kwasny and Sondhei-
mer, 1979) do not seem to be satisfying solutions for our
problem.

- 65 -

In other hand, we do not agree that syntax be given a subordinate role, as Schank's theoretical approach implies (though in Schank and Riesbeck's parser, ELI (Riesbeck and Schank, 1978), syntax has a more important role than would be expected). In any event, without enough syntax it becomes hard to analyze complex sentences and to explain a number of psycholinguistic phenomena.

2. WEDNESDAY, the system proposed here, is the core of an extended mechanism we are developing and implementing in LISP (fig.1) (see Parisi and Castelfranchi, in press). It is an analyzer with semantic output based on word interpretation. The semantic information brought in by each word is progressively connected to get at the sentence's meaning according to syntactic constraints and expectations. Syntax is a set of instructions directly concerned with assembling semantic units.
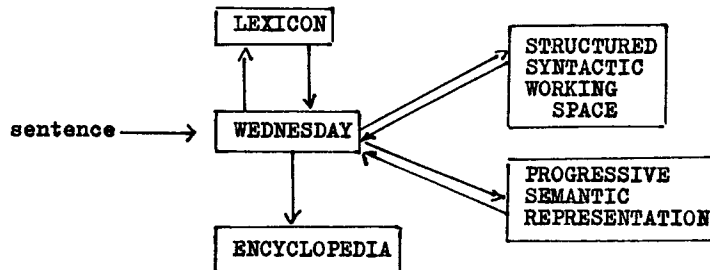


Fig. 1

What is characteristic of WEDNESDAY is that syntax is not a separate component, but is distributed throughout the lexicon. Lexical entries are composed of a declarative part and a procedural part. The procedural part is made up of syntactic instructions designed to assemble the declarative (semantic) part of the entry with the declarative parts of the other words in the sentence. The syntactic assembling process

allows the construction of the sentence's semantic network to
be carried out progressively and so it can also operate on in-
complete sentences. Furthermore, the system's output has a
format which is homogenous to that of the knowledge base
(encyclopedia). This means that as the sentence analysis pro-
ceeds, it is natural to have the knowledge base controlling
the assembly process top-down, in cooperation with the lexical-
-syntactic analysis. Clauses in multi-clause sentences are
conceived as restricted search spaces for the assembly instruc-
tions carried by the words, and the same is true for noun
spaces (noun phrases). Part-of-speech word categories are
interpreted as procedures for opening and closing spaces as a
function of context. Therefore, part of the process is depend-
ent on a recursive (push-down storage) mechanism. On the other
hand there is a semantic sentence memory that is transparent
to the closed-level system.

Being lexically based, WEDNESDAY can deal in a natural
way with idiosyncrasies typical of many words. It can also
deal with flexible idioms, i.e. idioms that can vary in morpho-
logy, word order, syntactic construction, semantic additions,
and synonyms. Their recognition is governed by the individual
lexical entries and takes place at the assembling level.

Word disambiguation is treated in a non-deterministic
way. Syntactically based disambiguation (e.g. the word that)
attempts each reading of an ambiguous word in turn and cancels
inappropriate readings by testing them against syntactic well-
- formedness criteria. Encyclopaedia based disambiguation
(e.g. the word bank) works through the activation of encyclo-
paedic nodes by the words in the sentence. It selects that
reading of an ambiguous word which activates the node "better"
connected to the other activated nodes. It should be noted
that, our approach, in comparison to Small's work on Word
Expert Parsing (Small, 1980), is characterized by (a) non-
-determinism, (b) a more systematic recourse to syntax, (c) a
separation of syntactically based word disambiguation from
encyclopaedia based word disambiguation.

- 67 -

References

Capelli A., Ferrari G., Moretti L., Prodanof I., Stock O.
"An ATN Parser for Italian: Some Experiments". In
Proc. of the Seventh International Conference on Com-
putational Linguistics. Bergen 1978.

Charniak E. "Six topics in Search of a Parser". Proc. of the
7th IJCAI. Vancouver, 1981

Ferrari G., Stock O. "Strategy Selection of an ATN Syntactic
Parser". Proc. of the 18th Meeting of the Association
for Computational Linguistics. Philadelphia, 1980.

Hayes P., Mouradian G. "Flexible Parsing". Proc. of the 18th
Meeting of the Association for Computational Linguistics.
Philadelphia, 1980

Kwasny S. and Sondheimer N. "Ungrammaticality and Extra-
grammaticality in Natural Language Understanding Syst-
ems". Proc. of the 17th Annual Meeting of the ACL.
San Diego, 1979

Parisi D. e Castelfranchi C. La comprensione del linguaggio.
Turin, Boringhieri, in press.

Riesbeck C. and Schank R. "Comprehension by Computer: Expect-
ation Based Analysis of Sentences in Context". Research
Report 78. Yale Univ., 1978.

Small S. "Word Expert Parsing: A Theory of Distributed Word-
Based Natural Language Understanding". TR-954; Dept.
of Computer Science, University of Maryland, 1980.

Weischedel R. and Black J. "If the Parser Fails". Proc. of
the 18th Annual Meeting of the Association for Comput-
ational Linguistics. Philadelphia, 1980.

Woods W. "Transition network Grammars for Natural Language
Analysis". Comm. of the ACM, Vol. 13, no. 10, 1970