Control and Cybernetics

vol. **25** (1996) No. 2

LISP application for modelling of user's utterance for computer system

by

Jadwiga Ratyńska* and Jacek Wojutyński**

* Technical University in Radom,
Poland

** Scientific Centre for Terotechnology in Radom,
Poland

Abstract: LISP is one of the functional languages. It is used as a programming language for modelling of user's utterance. The process of modelling is aided with computer. LISP is one of few languages which can be applied in the field of artificial intelligence. The identification of information expressed by user's utterance is realized according to grammar rules. D-C grammatical processor is used for this purpose. An analysis module is also used for automatic generation of proper grammatical forms. It is possible to perform the analysis correctly if topically limited language area is considered.

1. Introduction

LISP, treated as an instrument for performing calculations on symbolic structures can be used in many fields relative to information processing, for instance in natural language processing.

In case of artificial intelligence applications dealing with research and modelling of objects relative to intelligent behaviour, instead of symbolic transformations we have to do with building of calculation models based on suggested theories. Programs written in LISP make possible to analyze such models.

The efficiency of a natural language processing model considerably increases if topically oriented natural language area is used, Ratyńska, Wojutyński, Homenda (1990), Pustejowsky (1993). LISP is very well suited for description of models relative to natural languages.

The character of LISP helps the programmer to concentrate on the description of a problem, connections among elements of a system, flow of information etc.

It is logical to describe the methodology of dialogue in natural Polish language in LISP. Therefore programs written in LISP provide user models which are indispensable for natural language processing.

2. Natural language as language of access to conversational system

Natural language is one of the most useful instruments for communication between user and computer. Polish language like most languages used by a man is an irregular set of words and rules for building of proper forms. This irregularity is the cause of essential difficulties in generation of formal language description.

There is no uniform mathematical model of Polish language and formalization of this language meets exceptional difficulties. It makes impossible to establish full language control considering the informatic description of full language area.

Polish language limited to a dialogue topic, especially to the topic of exact sciencies, limits vocabulary and also variety of utterance building the syntax becomes less complicated, Godlewski, Kołkowski, Lipiński (1983).

An application of natural language limited to a special topic has many advantages. First of all it is connected with limitation of vocabulary and semantics which are two basic elements of natural language. Grammar understood as syntax, semantics and inflexion is also limited to defined piece of information, David, McAllester, Givan (1992), Ratyńska, Wojutyński (1994).

Those properties make possible to formalize the natural language in the dialogue relative to an established piece of information. The following properties result from the analysis of utterances relative to the language area which is topically oriented, Ratyńska (1989):

1. in analyzed sentences we deal with repetition of words and expressions,
2. it is possible to separate the given positions on which defined words appear,
3. a subset of sentences can be separated differing from each other in words on given position.

The above properties make it possible to introduce the idea of formally limited language area (FLLA). FLLA can be described in a formal way. This means that FLLA can be described by means of mathematical model assigned to this area.

There is a possibility to organize a dialogue of standard value according to a scheme or set of information flow schemes between user and computer. The scheme or set of schemes taking part in dialogue between man and computer is called dialogogram, Godlewski, Kołkowski, Lipiński (1983).

Dialogogram is a network of information flow which controls construction of utterances in Polish natural language (PNL) by user. The utterances are identified and understood by computer having at disposal field vocabulary (FV).

An application of LISP increases the efficiency of utterance identification. Each formally limited language area has its own field vocabulary and dialogogram. Computer identifies user's utterance by use of dialogogram and vocabulary.

3. Grammatical model of the computer system input structure

The Backus normal form (BNF) can be used for description of morphology, syntax or semantics of language. Notation of syntax has the shape of syntactic definitions.

The syntactic definition contains the following elements:

1. the name of defined sequence between the signs $\langle\rangle$,
2. sign ::= is interpreted as "equal by definition",
3. the sequence of symbols and metalanguage variables are separated by sign | interpreted as "or".

The application of this notation for description of Polish language inflexion intends to reduce the grammatical description of a word to the sequence of a few symbols describing the formal parameter (FP) as follows:

$$FP = \langle a\ b\ c\ d\rangle$$

where:
a determines the part of speech,
b determines gender,
c determines singular or plural number,
d determines case.

Upon introduction of the following notation:
RZ symbol of noun,
PM symbol of adjective,
ZM symbol of pronoun,
LI symbol of numeral,
CZ symbol of verb,
PK symbol of adverb,
PI symbol of preposition,
SP symbol of conjunction,
it is possible to write in BNF:

$$\begin{aligned}
\langle word\rangle &::= \langle inflected\rangle|\langle uninflected\rangle\\
\langle inflected\rangle &::= \langle RZ\rangle|\langle PM\rangle|\langle ZM\rangle|\langle LI\rangle|\langle CZ\rangle\\
\langle uninflected\rangle &::= \langle PK\rangle|\langle PI\rangle|\langle SP\rangle\\
\langle gender\rangle &::= m|z|n\\
\langle number\rangle &::= P|Mn\\
\langle case\rangle &::= M|D|C|B|N|Msc
\end{aligned}$$

where:

m masculine (gender),
z femine (gender),
n neuter (gender),
P singular (number),
Mn plural (number),
M nominative,
B accusative,
D genitive,
N instrumental,
C dative,
Msc locative.

The syntax model of Polish language can be also described on the basis of output data of the morphological model. Grammatical categories of particular metalanguage variable are determined by grammatical processor which plays the rule of declension-conjugation processor (D-C processor).

D-C processor generates grammatical forms of analyzed metalanguage variable on-line. This is the grammatical processor which generates all grammatical forms of words according to the grammar of Polish language.

A program simulating the work of D-C processor has been written in LISP. Construction of the rule set for metalanguage variable is considerably simplified owing to LISP. A fragment of rule set for adjective $\langle PM \rangle$ is written in the following way:

$(setf\, RULES - PM, ($
$(\text{"}a\text{"}\ (\text{"}PM + z + P + M\text{"})$
$(\text{"}ą\text{"}\ (\text{"}PM + z + P + B\text{"}\ \text{"}PM + z + P + N\text{"})$
$(\text{"}ej\text{"}\ (\text{"}PM + z + P + D\text{"}\ \text{"}PM + z + P + C\text{"}\ \text{"}PM + z + P + Msc\text{"})$
$))$

To improve the efficiency of the D-C processor work it is necessary to assure the cooperation between D-C processor and field vocabulary. All words connected with given language area are placed in field vocabulary in the basic form, for example noun in nominative, verb in infinitive etc.

The words are grouped in proper subset according to the kind of metalanguage variable. It is connected with easier access to the basic forms of metalanguage variables. D-C processor works according to the algorithm shown in Fig.1.

The declension-conjunction processor is used in analysis as well as in synthesis module which is shown in Fig.2.

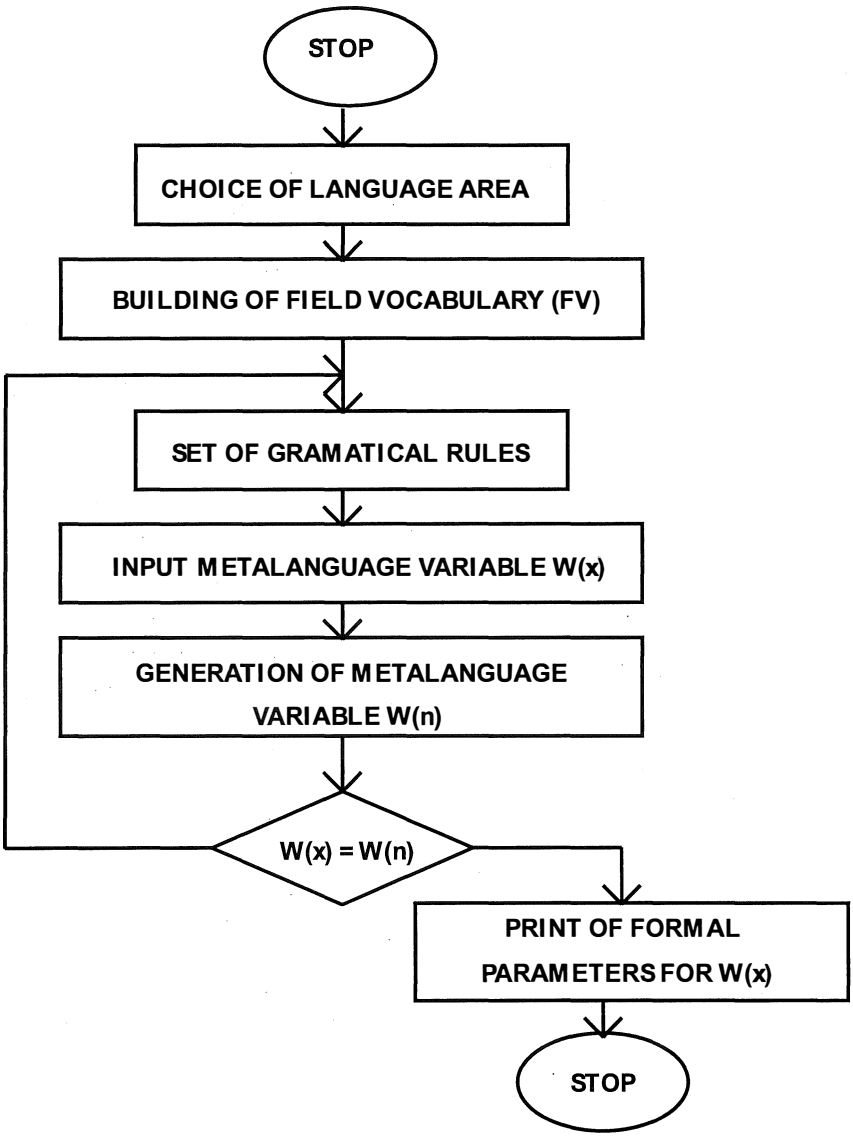

Figure 1. The algorithm for the D-C processor

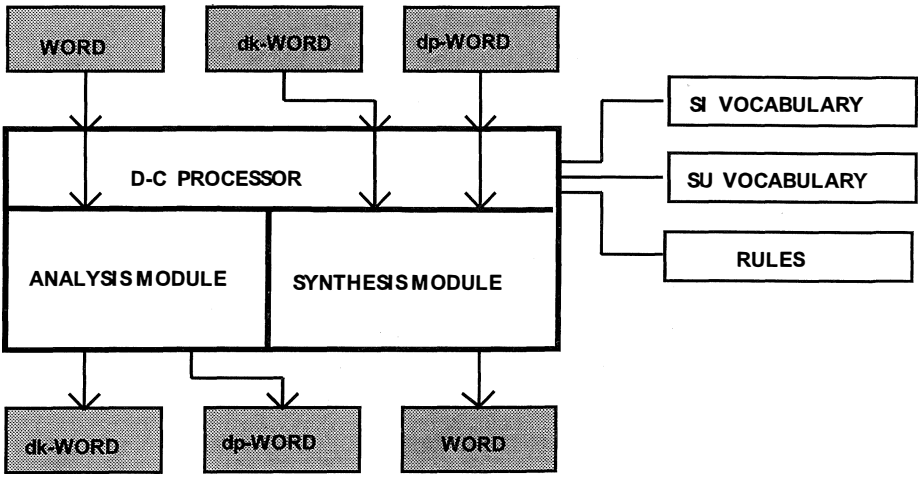

Figure 2. Scheme of D-C module

SI VOCABULARY vocabulary of inflected words
SU VOCABULARY vocabulary of uninflected words
dp-WORD basic form of the word
dk-WORD inflected form of the word

4. The analysis module

A precisely defined basic form of a word, formal parameters, syntactics and epistemological structure are the basic substrate of the analysis module. The module is presented in Fig.3.

In Fig.3 a rectangle denotes the function of the analysis module, an single oval – the imprecisely defined data, and a double oval – the precisely defined data.

The epistemological structures determine information flow paths relative to the actual dialogue. In this way the graph is generated. The words playing the roles of object, subject etc. in the sentence are placed in nodes of that graph.

The following steps must be done to obtain the precisely defined epistemological structure. First of all the grammatical module including information as to which grammatical forms of words play the roles of subject, object etc. in sentence must be generated. This module is marked by S-F in Fig.3.

The S-F module is used for carrying out the analysis marked by ANL_sf in Fig.3. The S-F module acts on dk-WORD received at the output of the D-C. The information gained is not precisely determined because the analyzed dk-WORD can play a different role in the sentence, for instance subject, object etc.

Then the grammatical module called reduction module – RED – is used. This module consists of position ranges, determining the possible sequence of words (used as subject, object etc.) in sentence. By use of that module the precisely determined sequence of analyzed parts of speech is obtained.

To determine the epistemological structure the grammatical module called S-P module is used. In this module, according to the Polish grammar, all possible epistemological structures assigned to the tested language area are placed. By use of that module the precisely defined epistemological structure is obtained from the system marked ANL_sp in Fig.3.

Except from the path connected with precise determination of the epistemological structure, the paths of precise determination of the inflected word parameters (ANL_dk), the basic word forms (ANL_dp) are also generated.

The precise definition of dk-WORD is carried out in the following way: at the output of D-C processor the inflected form of the analyzed word (dk-WORD) is obtained.

By use of grammatical module marked SYNFLEX in Fig.3 determining the common part of ANL_dk and ANL_sf more limited information relative to the dk-WORD is received but the information gained is not precisely determined.

To receive precisely defined form of dk-WORD the grammatical module containing agreement ranges marked R-Z in Fig.3 is used. The R-Z module gains information relative to possible grammatical connections among words. Owing to this module the precisely determined grammatical form of word signed by dk-WORD is received.

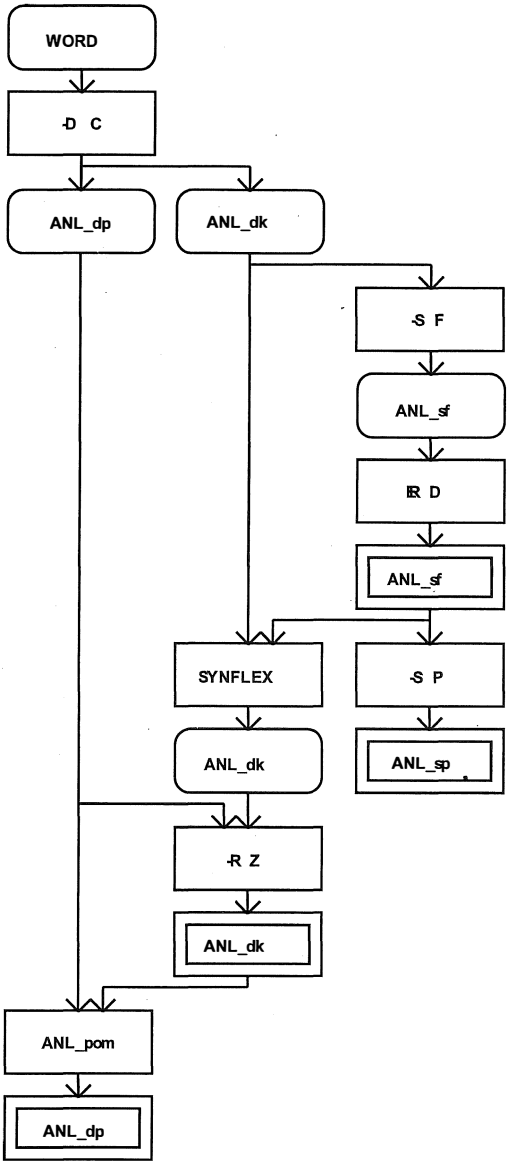

Figure 3. The analysis module

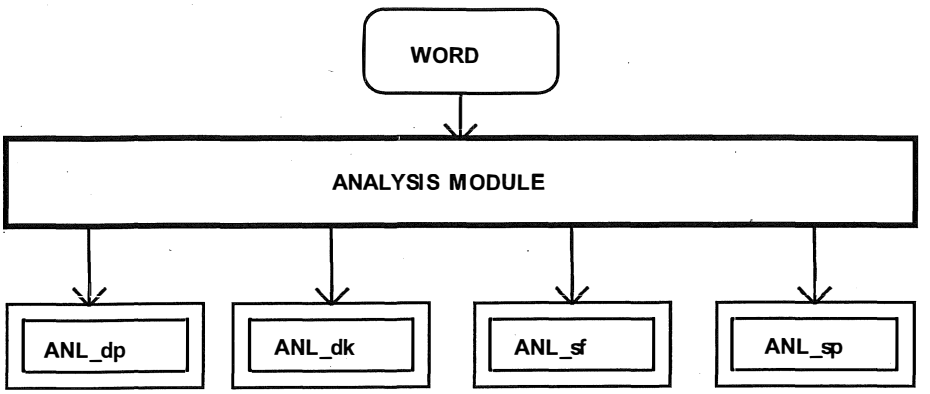

Figure 4. Automatic generation of linguistic description of a metalanguage variable

In the output of the D-C processor the basic form of word marked by dp-WORD is also given but this form is not determined in the precise way.

To determine the basic word form precisely the auxiliary analysis module marked by ANL_pom in Fig.3 is used. This module uses information relative to the precisely determined grammatical form of dk-WORD. In the output of the ANL_pom module the precisely determined word basic form – dp-WORD is received.

The following precisely defined forms of word are obtained after using the analysis module: the basic form, inflexion, syntactics and epistemological structure of sentence. The results of the analysis module work are shown in Fig.4.

The analysis module is also used for automatic generation of knowledge base marked K-BASE in Fig.5. K-BASE consists of words to which the basic form and inflected parameters are assigned.

5. Summary

LISP belongs to the functional languages. The big advantage of functional languages consists in the fact that the value of expression depends on the value of its components and not on calculations. The value of expression does not change in a given contex either.

LISP consists of two calculation mechanisms: functional language and environment which stores all objects used by a program. This is one of few languages that can be used in the field of artificial intelligence (AI).

Generation of grammatical forms of input structure on-line is very important for identification of information expressed by properties of natural language texts, Winston (1984), David, McAllester, Givan (1992), Palmer (1993).

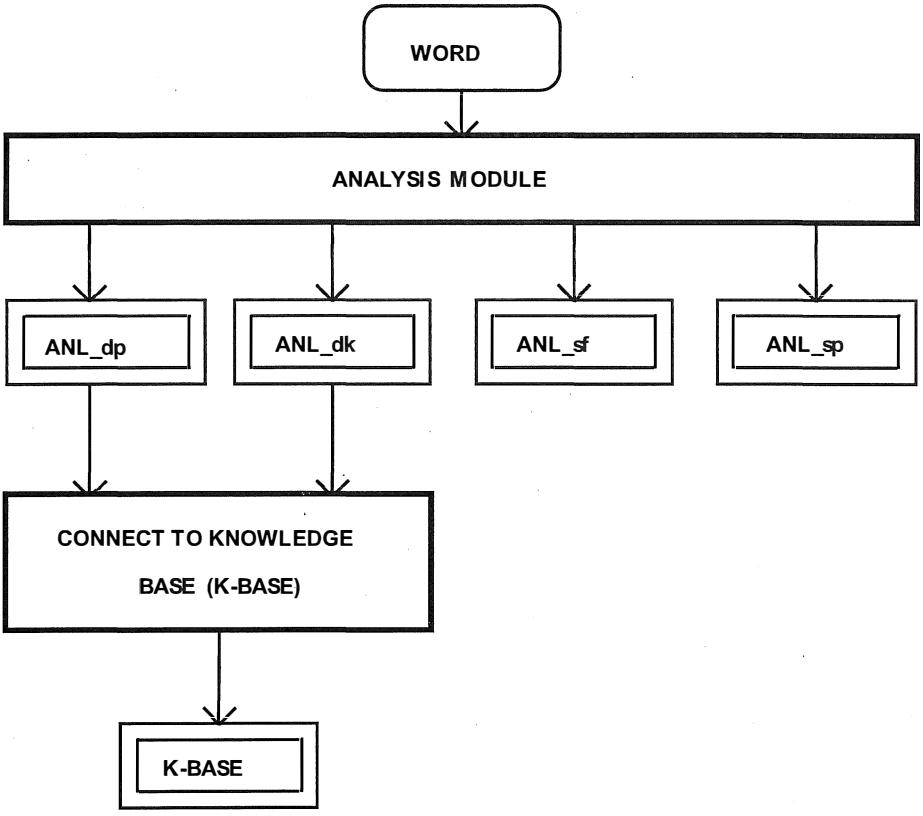

Figure 5. Automatic generation of knowledge base (K-BASE)

This identification is carried out in several steps. In the beginning correctness should be checked of input utterance consisting of the metalanguage variable conform to the grammar.

Grammatical D-C processor is used for this purpose. This process is very difficult and complicated. However, it is possible to make an identification and interpretation of user's utterance in proper way if topically limited language area is considered.

References

GODLEWSKI, M., KOŁKOWSKI, L., LIPIŃSKI, E. (1983) *Problems of Algorithmization and Automatization of Didactic Process.* PWN, Warsaw (in Polish)

LEVIS, M. and NORTON (1983) Automated analysis of instructional text. *Artificial Intelligence*, **20**, 3.

WINSTON, P.H. (1984) *Artificial Intelligence* (second edition). Addisson-Wesley, Reading, Massachusetts.

RATYŃSKA, J. (1989) Dialogue organization in topically limited language area. *Conference on Computerization of Didactics*, Łódź (in Polish).

RATYŃSKA, J., WOJUTYŃSKI, J., HOMENDA, W. (1990) Expert system with natural language communication. *International Symposium on SYSTEMS-MODELLING-CONTROL*, Poland.

DAVID, A., MCALLESTER, GIVAN, R. (1992) Natural language syntax and first-order inference. *Artificial Intelligence*, **55**, 1.

PUSTEJOWSKY, J. (1993) Lexical knowledge representation and natural language processing. *Artificial Intelligence*, **63**.

PALMER, M.S. (1993) The KERNEL text understanding system. *Artificial Intelligence*, **63**.

RATYŃSKA, J.,WOJUTYŃSKI, J. (1994) Identification of language phrase in topically oriented language area. *Informatyka*, Wyd. Polit. Łódzkiej, Łódź.