

## ON AN APPROACH FOR DESIGNING LINGUISTIC PROCESSORS

Radoslav Pavlov, Galia Angelova

Laboratory of Mathematical Linguistics, Institute of  
Mathematics with Computer Center, Bulgarian Academy  
of Sciences, P.O.B. 373, 1090 Sofia, Bulgaria

The present paper discusses the principles of designing a system for man-computer dialogue in natural language. The system is being elaborated at the Laboratory of Mathematical Linguistics at the Institute of Mathematics with Computer Center of the Bulgarian Academy of Sciences.

The described system requires:

1. A formal description of the syntax of basic nuclear structures of the natural language sentences which has to be used during the process of syntactical analysis of the users' phrases in a natural language;
2. Techniques for representing in the computer the knowledge about a given problem area. The obtained description has to be used during the process of semantical analysis of the users' phrases in a natural language;
3. A dictionary containing the terms used in the chosen problem area. The dictionary is compiled by specialists in the given problem area;
4. A dictionary containing the service vocabulary (independent of a particular problem area) - word-groups as "less than", "greater than", "equal to", "and", "or", "as... as" etc.
5. A basic software system - a data base management

system (for example, a relational DBMS), or a program package (for example, BMDP - bio-medical data processing). These ready software products service users of a given problem area and they usually have their own language for describing and processing the data.

The considered system is viewed as a superstructure over the given software system and it performs:

1. Translation from a language close to the natural one into the internal software system's language for description or processing the data;
2. An effort to analyse the correctness of the user's request in terms of the given problem area description and messages when discovering mistakes;
3. When ambiguities happen a dialogue is carried out in order to make the user's request unambiguous.
4. Processing the results from the basic software system and final shaping of the messages in the context of the user's request.

Such a system is called a linguistic processor (or linguistic preprocessor) and is considered as a dynamic system which can be generated over various problem areas and various basic software systems.

By "a natural language as a tool for the man - computer dialogue" in this paper we mean the following:

- a fixed list of appropriate natural language phrases is not assigned;
- restrictions are not imposed on the grammatical structures of the sentences the man enters into the computer;
- man has to interact with the computer in a correct natural language.

Dictionary of the terms  
of the problem area

Structural description  
of the problem area

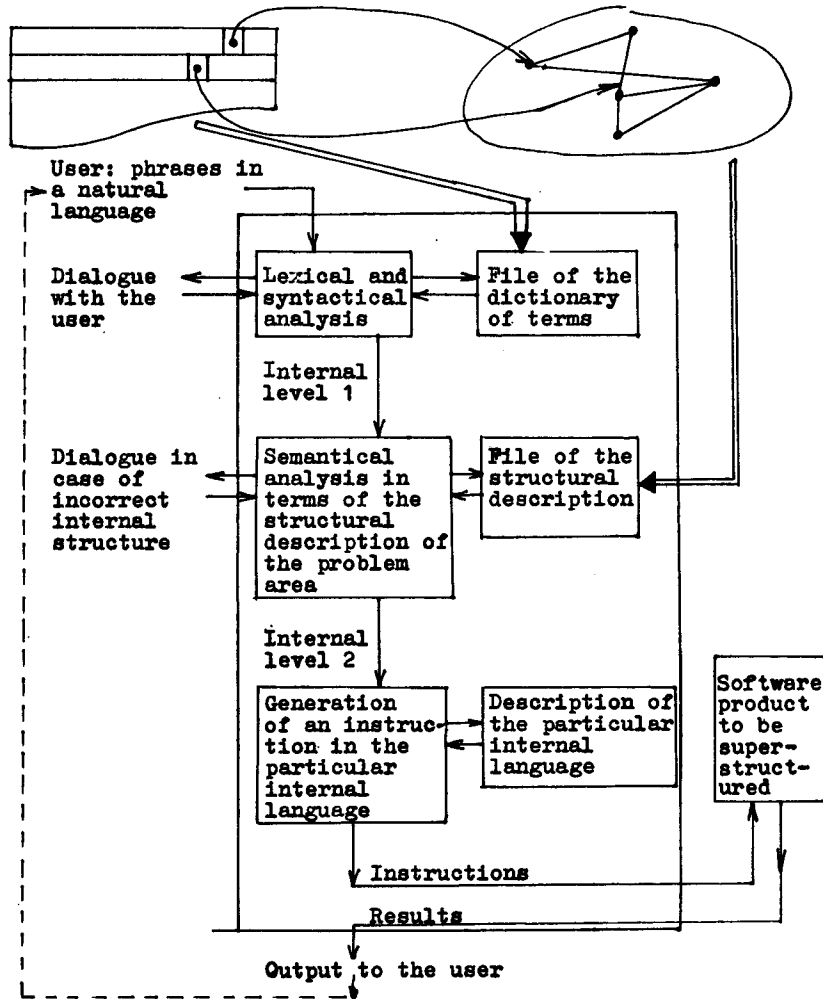


Fig.1. Functional relationships between different parts of the linguistic processor and the problem area descriptions.

The describe approach for creating linguistic processors can be qualified as a semantic - syntactical one since semantic analysis in terms of the given problem area description has a leading role in the process of "understanding" of the input phrase in a natural language, while syntactical analysis has an auxiliary role. The structural description of the problem area is a network designed for the representation of the problem area semantics and is viewed as a static, immobile and determinative tool for "understanding" and accomplishing the dialogue. The syntactical analysis does not presume a full scanning of the input sentence. It involves analysis of syntactical dependences at different levels with respect to the problem area complexity and, naturally, to the input phrase complexity. By means of an elaboration of the formal description of Bulgarian syntax, an improvement of the techniques for syntactical analysis is envisaged.

When working out the principles of construction of linguistic processors, some peculiarities of the Bulgarian language have been taken into account, namely, its flexional character, the absence of a case system, free word-order, etc.

The approach proposed for the realization of the man - computer dialogue in a natural language allows to process problem areas with different complexity. Linguistics processors at two levels are being elaborated at the Laboratory of Mathematical Linguistics - for access of users - nonprogrammers to a relational data base in Bulgarian language and for access of users - nonprogrammers to the facilities for statistical analysis proposed by the program package BMDP in Bulgarian language.