

INTERACTION WITH A LIMITED OBJECT DOMAIN -

ZAPSIB PROJECT

A.S.Narin'yani

Computing Center, Siberian Branch USSR Ac. Sci.  
630090 Novosibirsk, USSR

**Abstract.** The report presents the basic principles of the ZAPSIB project aimed at the development of a modular series of linguistic processors designed for natural language (NL) interaction with applied data bases. The general structure of the ZAPSIB processors and functions of the main modules are discussed, as well as technology of the project including problem of processors adaptation to an object domain of the interaction.

1. Basic principles

Launching the project the authors\*) were aware of specialities of commercial systems which are principally different in many respects from the experimental programs developed as their prototypes at the beginning of the applicational direction of our NL works.<sup>1,2</sup>

This position was implemented in the basic principles of the project which could be formulated as follows:

- (a) Giving up the realization of any "generalized" scheme of interaction (an "average" user ↔ an "average" object domain). No scheme of that kind is possible in principle: a customer's demands could differ decisively on the main parameters of the interaction, such as
- limitedness of the NL syntax ;
  - contents and complexity of the object domain;
  - the lexicon size;
  - the computer's resource;
  - the efficiency of L-processor, etc.

For some of the parameters the limits of those demands can vary up to 100, 1 000 or even 10 000 times. In this spectrum of diversity it is not possible to extract one or two dominant stereotypes - practically every customer needs his own L-processor, adequate to his special conditions and interaction domain.

This situation determines the strategy of the project: it programs the development of not one but a series of L-processors with the same general structure whose basic modules are realized as sequences of successively extending and compatible

\*) The project being carried out by A.I. Laboratory of the Computing Center of the Siberian Div. of the USSR Acad. Sci.

versions. Implementation of this principle is supposed to provide more adequate choice of L-processor configuration with regard to a particular user.

(b) Each L-processor is to be partitioned into the universal and adaptable parts. The latter covers all the information depending on the domain of application and includes

- the data base structure: object, their attributes and relations;
  - the lexicon of interaction domain, including the vocabulary, standard word-complexes and denotations.
  - the syntax of the formal language of the system the L-processor works with.
- To specificate the adaptable part of L-processor during its "tuning in" the object domain, the processor's modules are completed with special means. For a better effectiveness of the adaptation a professional carrying out this process is provided with a high-level declarative language and a set of specialized metaprocessors which compile the "outer" specification into the inner representation.

The complex of these metaprocessors composes the STEND system which is constructed specially to ensure maximal comfort and effectiveness of adaptation procedure (fig.1).

(c) Shortcomings of the traditional "syntactical analysis → semantical analysis" sequence are well known:

- This scheme enables to process only "syntactically normal" texts. Any violation of the norm (which is rather rule than exception for a mass user) leads to faults.

- In principle this scheme is based on assumption of existence of a "complete" formal NL model. But no such a model has been elaborated up to the moment and most probably it will not be available during nearest ten years.

- Even rather rough approximations of the model being developed recently are cumbersome, expensive and too efficiency-killing for a commercial type system. Semantically-oriented analysis of text based on maximal utilization of semantic "foundation" of a message and using syntax information as locally as possible for elimination of superfluous meanings, seems free of the mentioned shortcomings and much more adequate as a model of understanding process.<sup>2,3,4</sup>

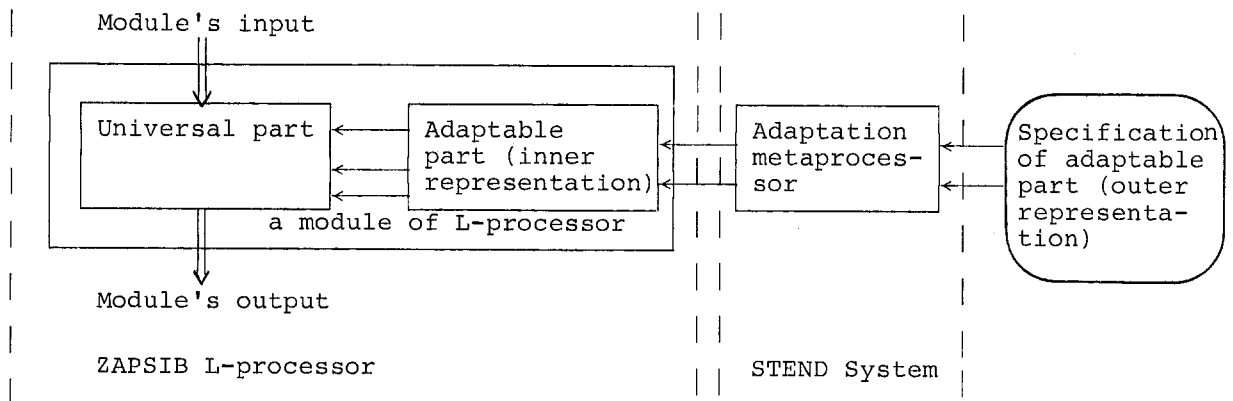


Fig.1. A module of a ZAPSIB L-processor and the scheme of its adaptation through the STEND system.

The sphere of applications of the approach is limited now to restricted object domains, and 'user - applied data base' interface is one of the most actual examples of such a problem.

For realization of the semantically-oriented analysis the ZAPSIB L-processors are completed with special means enabling to specify and use detailed data about the interaction domain.

(d) The main procedure of the analysis is organized as a non-deterministic bottom-up parse process, one- or multi-variant, depending on the processor version. This organization corresponds optimally to chosen formal apparatus based on the notion of `component` which generalizes the means of dependency and constituents grammars.

## 2. General scheme of ZAPSIB L-processors

The minor versions of ZAPSIB L-processors being under development now have the general scheme (fig.2).

Preprocessing module includes

- lexical analysis which decomposes the string of input text and divides it into words, numbers of various notations and letter-digital denotations;
- assemblage of word-complexes, i.e. standard combinations of lexemes which are used as an integrate semantic unit under further stages of analysis (War and Peace, International Federation of Information Processing, etc).

Main process operates with a system of rules, each of them being production realized in a high-level context-dependent grammar. The system includes special means to control partial ordering of the rules application. The level of the grammar and control means depends on the L-processor version. At the module's

output one or more (in a case of ambiguous result of the analysis) acyclic parse graphs are formed.

Postprocessing comprises three stages:

- elimination of the local ambiguities with the help of global information about the text meaning formed up to the end of the parse;
- synthesis of the text semantic representation according to the parse graph;
- generation of the output representation of the text meaning in the user's system formal language.

Model of interaction domain incorporates all the semantic and pragmatic information concerning the interaction domain and necessary for the operating of all other modules.

Feed-back with the user serves, if necessary to specify the user's intentions and verify the results of the analysis. The ZAPSIB strategy regards applying to the user as an extreme measure in the most urgent cases.

Each of the main modules is in its turn a complex of modules and this provides sufficient flexibility and compatibility of different versions of the modules.

## 3. Technology of the project

For the development of individual modules as well as "assembled" configurations we use a two-stage technological cycle:

- (1) Creation of the working pilot program in the very high-level SETL language;
  - (2) Transferring the SETL-program into the instrumental language (PL/1).
- Such a technology helps to cut down efforts on the development of the universal part of the software up to three times.

Special attention in the project is paid to automation of the adaptation procedure of the L-processor to the user's object

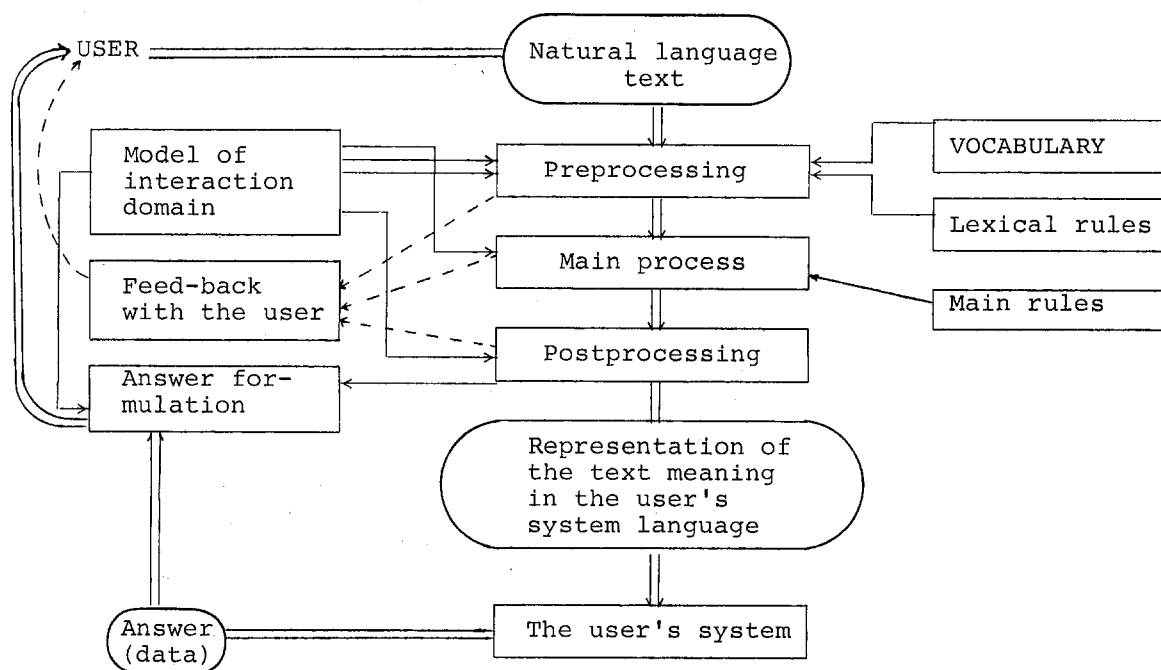


Fig. 2. General scheme of ZAPSIB L-processors.

domain. The adaptation is expected to be realized on the pilot "L-processor - data base" tandem by means of the STEND system.<sup>5,6</sup> Provided with a set of specialized dialogue means the system enables to carry out procedure by direct interaction with any of the L-processor modules.

#### References

1. Нариньяни А.С., Яхно Т.М. Контекстно-зависимая грамматика с разрывными составляющими в системе восходящего анализа. - В кн.: Взаимодействие с ЭВМ на естественном языке. - Новосибирск: 1978, с.157-165.
2. Левин Д.Я., Нариньяни А.С. Экспериментальный минипроцессор: семантически ориентированный анализ. - В кн.: Взаимодействие с ЭВМ на естественном языке. - Новосибирск: 1978, с. 223-233.

3. Narin'yani A.S. AI Work in the Computer Center of the Siberian Branch of the USSR Acad. of Sciences. - In: Machine Intelligence, Ellis Horwood Ltd. 1979, V.9.

4. Нариньяни А.С. Лингвистические процессоры ЗАПСИБ (Часть 1 - задачи проекта) . - Новосибирск, 1979. - 22 с. (Препринт/ВЦ СО АН СССР, 199).

5. Нариньяни А.С. Лингвистические процессоры ЗАПСИБ (Часть 2 - Общая схема и основные модули). - Новосибирск, 1979.- 48 с. (Препринт/ВЦ СО АН СССР, 202).

6. Левин Д.Я. СТЕНД - система адаптации лингвистических процессоров. - Новосибирск, 1980. - 29 с. (Препринт/ВЦ СО АН СССР, 238).