CZU 81'37

PROBLEMATICS OF THE SYSTEMS OF UNDERSTANDING NATURAL LANGUAGESAS

Zinaida CAMENEV
Free International University of Moldova (ULIM)
Anatol POPESCUL
Technical University of Moldova

Artificial intelligence (AI) researchers are concerned with developing computer systems that simulate human intelligence using techniques of semantic processing. AI is divided into three independent research areas: Natural Language Processing (NLP), robotics and expert systems. The communication is dedicated to the systems understanding natural languages. The problems necessary for the elaboration of such systems are: the domain of knowledge, the process of knowledge representation, organization of control of actions, organization of knowledge and access to it. If taking each of the problems apart we can say that the selection of the types of presentations of knowledge is unique but the means of presentations are different. Because of a big volume of informational data it is necessary to introduce an interpretation regime and an archives service. In such a case the system must possess metaknowledge of three types: metaknowledge of the environment, metaknowledge about the language of communication and metaknowledge about the participants of communication. The management of the intercourse system is performed by such methods as „Top-down" and „Bottom-up". As it is impossible to prepare all the necessary reactions of the system, the only way out consists in the fact that the system itself realizes what happens and is able to express it in terms understood by the user. Such selfconsciousness of the system sometimes is very difficult to organize because of a big volume of data. It all depends on the access of knowledge which must be given in alphabetic order, small units must be transformed into bigger ones called blocks. New processed information is received by concluding rules and facts stored in the system by means of formal, specialized and metaconclusions. These bring to a good system understanding natural languages which then must be translated by means of the computer programs.

Keywords: *artificial intelligence, natural languages, human intelligence, systems understanding, representation of knowledge, intercourse system, metaknowledge, access to knowledge, concluding rules and facts, machine intelligence.*

During the elaboration of the systems (models) understanding the Natural language it is necessary to solve the following questions:

I. **Which** internal presentation of the knowledge is used?

a) *The domain of knowledge.* What domains of knowledge may be presented and what may not?

b) *The completeness of properties.* Are the properties of the objects evident or not during their presentation in the system?

c) *Basic concepts.* Are there any existing basic concepts (atomic ones)? How are they?

d) *Variety of representations.* Is it possible to present one and the same fact in different ways and by different presentations?

II. **Practicalness**. Is the system designed for work under real conditions?

a) *Introduction of the knowledge.* Is the system accommodated for the introduction of initial knowledge by big volumes requested by the practice, and also for the operational modification of knowledge emerging in the process of the exploitation of the system?

b) *Metaknowledge.* Is the system oriented to the work with the user not knowing in detail the structure of the knowledge and its contents? Can the system announce about its knowledge?

c) *Errors.* What does the system do during wrong processing of the input information?

III. **Actions**.

How will the system change knowledge in behavior (=action)? What actions are admitted? How are they presented in the system?

a) *Modification* and *concordance* in the surrounding world with the changes in the knowledge of the system. How is the concordance of changes provided in the environment with the changes in the knowledge of the system?

b) *Frame problems.* How does the system determine what consequences will the fulfilled action have?

c) *Planning and implementation*. How does the system represent the hypothetical (planned) and real (implemented) sequence of actions?

IV. **The process of Representation.** How is the input task represented in the knowledge of the system?

a) *Assimilation.*

How does the system transform the input task into an existing internal structure of the knowledge?

b) *Accommodation.*

How may the internal structure of the system be changed to cope with the input task?

V. **Organization of Control.**

How does the system organize itself to solve the task?

a) *Control of the System.* How does the system organize its moduli and resources during the solution of the task?

b) *Selfconsciousness*. What does the system know about its current condition and its knowledge? How does it get this information?

VI. **Organization** of Knowledge and Access to Knowledge.

How does the system determine which knowledge is appropriate in the given moment of time?

a) *Connectedness*. How are the elements of knowledge interconnected?

b) *Mechanism of Access.* What mechanism is used for the selection of appropriate knowledge?

c) *Confrontation*. How are different structures of data compared to determine their equality and similarity?

VII. **Conclusion**. How can facts be received by the system without using information from the surrounding world?

I.1. Problems of the Knowledge Representation

The essence of this problem consists in the expression of the given domain of knowledge in the selected presentation and the determination of those aspects of the knowledge domain which cannot be represented. It is necessary to solve how

the objects and the interrelations existing in the real world are reflected in the system of the units of presentation and relation. It is necessary to note that the selection of a certain presentation of the given world leads to the fact that some interrelations will be evidently filled up but the others will not. For example, if we shall present the picture in the way of black and white dots then the interrelation "be on the left of" or "be on the right of" in the given picture may be not given evidently, i.e. they are received by means of some procedures of the conclusion. But, if the picture is represented as a semantic net then the similar relations may be evidently expressed.

2. Basic Notions. Practically universal recognition received the opinion about the fact that knowledge must be expressed in the way of indivisible units of meaning (=atoms, or primitives) (Schank; Wilks). Differences exist in the question: what is an atom and what the quantity of these atoms is. Some consider that atoms may be the canonic words (Filmore), the others consider that these must be the units of meaning from which the words consist of (Schank). It is necessary to note that such a point of view doesn't reflect the real situation. In reality the task of the science consists in forming new basic concepts including the way of laying out of indivisible units before the basic notions are divided into smaller ones. Thus, the system **must not have fixed basic notions in perspective**, their structure must be changed in the process of system functioning, i.e. what was an atom yesterday, today it may get an internal structure. It is easy to see that the description of complex objects and events cannot be given in the terms of the unique set of atoms.

Presentation of knowledge as atoms, more elementary than words, simplifies the conclusions. But we must remember that the smaller the atoms the more complex it is to compare it with the internal knowledge. Thus, the division into atoms simplifies the conclusion but complicates the presentation of the input information in the internal presentation.

3. Completeness of the Particularity. Presentation owes the completeness of the particularity A, if for any object that has the particularity A this fact is represented in the system evidently. The presence of the completeness of the particularities in an evident way is the guarantee of the absence of this particularity, in general. But to provide the completeness of the particularities it is rather hard in the dynamic world, besides, the achievement of the completeness (during the dynamics) is connected with great computer loses.

4. Variety. It is generally known that the complexity of solution of the problem, to a great extent, depends on the fact how the task is presented. The presence of different presentations (according to the type and means) in a system allows to combine the advantages of different forms of presentations. Under the way of representation of fact A we shall understand that internal system of language in which knowledge A is presented. Under the type of presentation of fact A we shall understand the affirmation in some internal languages corresponding to fact A. Thus, for example, fact "the object X is in the place Y", may be presented by different means (e.g. in the calculation of predicates) and in different types (in the way of coordinates of place Y or by the indication of the place of the object X in relation to the place of some known object). In the process of the usage of different presentations it is necessary to solve at least three problems: the selection of the presentation, the transformation of one presentation into another, combinability of presentations.

5. Selection of Presentation. In systems with different presentations the concrete fact may be presented by some types and means. In connection with this the system must have a mechanism determining in what way it is necessary to present some private fact. For example, the fact about the location of some object A in the place of B may be demonstrated by the correlation of absolute coordinates of B with the object A, or by correlation of the place of object A with the place of some known object C or by the indication of relative coordinates of A concerning to D. So the demonstrated fact may be shown by different means. In a concrete system the basic problem is the selection of the type of presentation as the means of presentation is usually unique for the whole system or is predetermined by the presented knowledge (e.g. for the dictionary of the system it is used one way of presentation, for the information about the external world – another).

6. The compatibility of presentations. The variety of presentations brings to the fact that one and the same information may be stored in the system in different ways. If one of the types is changed then the other type must be controlled for compatibility with the first. For example, let the system know that object A is found relatively to object B "on the left of" and the information comes (on the coordinates) about the movement of object A. It is clear, that the given relation "on the left of" must be counted in relation to the movement of object A. The alternative solution of the given problem may be the selection of the main type of presentation that is only changed during the entrance of new facts.

7. Transformation of presentations. In connection with the fact that selection of this or that presentation (both the type and the means) determines the grade of complexity of the processing of the input information and modification of knowledge, it is necessary to have a mechanism realizing the transformation of one presentation into another. Sometimes, however, information from one presentation cannot be transformed into another. (e.g. from the fact "A on the left of B" and the knowledge of coordinate B we cannot get coordinate A).

II.1. Practicality

The given aspect characterizes the suitableness of the system for the practical (industrial, but not experimental) usage in the tasks of the real grade of complexity. The basic particularities of the industrial systems in comparison with the experimental ones are the following: a big volume of informational data provision (ID), user's lack of knowledge of the detailed structure of the **ID** and its contents (in connection with its big volume) and inadmissibility of the situation, when the system cannot process the input information without explaining the causes of failure.

2. Introduction of the knowledge. In connection with the big volume of ID the problem of the input introduction of knowledge and their modification is transformed into a difficult and complicated task. The complexity of elaboration and modification of ID is made difficult because of the fact that it is necessary to prepare the ID in the same way as the program. It leads to the necessity of usage of an interpretation regime and the introduction of archives service.

3. Metaknowledge. At present it is impossible to create systems possessing complete data even in a single domain not to speaking about the creation of the systems of communication with the user in the natural language. That is why the system must know and be able to announce about the fact what it knows from the subdomain of knowledge that the user is interested in at the given

moment. In order that the answer to the user's demand about the knowledge of the system not to include and enumerate all the known data the system must possess metaknowledge. For the system to be able to answer the question about its knowledge metaknowledge must be presented evidently, i.e. in a way accessible for research. Metaknowledge of the system must include knowledge of three types: about the environment, about the language of communication and about the participants of communication. Metaknowledge about the language must contain facts about the input and output languages, including the description of the limits and the data structures. Metaknowledge about the environment must describe the structure of the thematic domain, the facts which are given in the system. In order to simplify and establish the mutual understanding with the user the metaknowledge must include data from the domains adjacent to the domain known to the system.

4. Errors. The systems understanding the existing natural language and elaborated at present cannot answer any input information. The causes of failure in the formation of the answer are the following: the user made a mistake in the syntax of the natural language; the user included an ambiguous or incorrect demand; the input information cannot be processed in connection with the limited possibilities of the system, etc. For successful exploitation the industrial system must be able to report to the user about the causes of the failure, direct him/her to periphrasing of the input information and demand the missing information from him. We shall underline that in the systems understanding the natural language the result of such an approach is the necessity of a dialogue regime.

III.1. Actions

The given aspect characterizes how the correspondence among the actions in the real world is established, on the one hand, and the statistical structures, on the other. Let us note that both the data and the programs of the system present the static knowledge. The dynamics is presented by the process of fulfilling the program.

2. Modification and agreement. The main problem during the presentation of the action in the system consists in the fact of modeling the aim of changes called by the action. We shall name the model of action **an operation.** To model the action is possible to be modelled in different ways.

Let us explain the essence of the two main approaches on the basis of an example. Let the moment T_1 is "The book is on the table". In the moment T_2 we begin to move the book in the direction of the edge of the table. In the moment T_3 the book appears to be pushed from the table (and falls down). The condition of the world in the moments T_1 and T_2 is not difficult to show. Really, in the moment T_1 the world is static, but beginning with the moment T_2 till T_3 (where $T_i<T_3$) the change of the condition of the world is expressed by the change of two (out of three) coordinates of the book in the direction of the movement (i.e. the standard reason of the action of movement). However, during the moment T_3 the standard presentation of the movement is not reflecting adequately the reality. In the given example with the object "book" it is advisable to connect the standard determining procedure, with the fact if this object is kept and if not then the falling of the object is modeled.

The essence of the problem of agreement consists in taking into account that the concrete presentation is not adequately reflecting the actions of the real

world in the modeled one. Presentation may allow such operations that are not possible in reality. For example, if in a concrete presentation the sizes of the objects are not considered then in the process of modeling the objects may be moved in such a way which is not possible in reality in connection with the space length of the objects.

3. The Problem of Borders. The essence of the problem (Raphael) consists in the fact of how the system may effectively determine the factors that must be measured as a result of some action and which may not. This problem bears a principle character reflecting the interconnections and interconditionality in the modeled world.

4. Planning and Fulfilling. In the simplest model of the world representing only its current condition the time may be presented inadequately in the terms of changing the knowledge about the world. In a more complicated model of the world it is possible at the same time to present some different conditions of the world characterized by different time (present, past and future). Such a presentation allows to remember a certain sequence of the events existing both in reality and hypothetically. In the process of representing the sequence of events we must connect the time with every fact during which this fact was real. Planning may be presented for example as well as the real sequence of actions. Under planning we understand search of some sequence of actions by the system leading to the desired condition of the world.

IV.1. The Process of Representation

The process of reflection of the input information on the internal representation is performed on the basis of the system knowledge. System knowledge directs and limits many possible interpretations of the input information.

1. *Assimilation*. The problem of assimilation includes: The fact of diversity of structures (information and quantity) found undoubtable in the surrounding world. This diversity also exists in the natural language. On the other hand, it is generally accepted that the quantity of structures in the intellectual systems must be considerably less than in the surrounding world. Thus, assimilation is based on the determination of the reflection of the external diversity of the input information in the limited number of internal structures takes place.

2. *Accommodation.* The concept of assimilation naturally leads to accommodation. Accommodation (Moore) is understood as the modification of the internal presentation of the system designed to extend the circle of the solved by the system tasks. It is possible to say that accommodation implies acquisition of new possibilities in the process of assimilation and, namely, this differentiates it from assimilation.

V. Organization of control

The present aspect concerns the questions of the solution of the input task organized by the system. The management of the intercourse system must combine such methods as “Top-Down” and “Bottom-Up”.

1. *Selfconsciousness*

In order the system to realize the intercourse with the user it must be able to realize its and the user's statements in terms known to the user. As it is known it is impossible to prepare all the necessary reactions of the system to the possible errors. The only way out consists in the fact that the system itself realized what

happens and could express it in terms understood by the user. The differences between realization and selfconsciousness are rather essential. The realization of the environment is equal to the understanding of input information in the natural language in the systems of processing the natural language. In the process of the realization of the environmental world the information comes to the input of the natural language rather evidently. Quite different is the situation with the selfconsciousness. The information about the processes of the system is presented on the lower level in the way of sequences of machine commands. One of the possible ways of the solutions of the problem of selfconsciousness is the interpretation of the work of the system at the level of abstraction determined by the terminology understood by the user of the system. It is necessary to mark that the question of the determination of the level of abstraction is complicated and not yet elaborated.

It is possible that the solution of the problem of the levels of selfconsciousness consists in the interpretation of the work of the system at some levels of abstraction. But if the system determines incorrectly the degree of abstraction the user is interested in then the user is giving, for example, the repeated question "why?" it will show the system the increase of the level of abstraction but asking the question "how?" will show the decrease of the level of abstraction.

In order to avoid the frequently appearing confusion of selfconsciousness and metaknowledge we shall mark the following. Metaknowledge is the knowledge of the system about its data, i.e. data both of the known by its facts and by the processes. Metaknowledge means statics. Selfconsciousness is the process as a result of which the system realizes its current condition. Selfconsciousness is a process, i.e. a dynamic one.

VI. Organisation of knowledge and Access to Knowledge

The usage in the necessary moment of the needed knowledge is an indicator of the intellectuality of the system. In the problem of access to knowledge we may chose three aspects: connecters of data, the mechanism of the access to the data and the means of comparison of the data, suggested by the mechanism of the access. The access to data is the operation inverse to the operation of remembrance of data. These operations imply contrary demands to the organization of data. In reality in order to find the access you must know where the object is. However, in order the object to be in the definite place it is necessary to pay certain efforts (-time). A typical example may serve the work with the vocabulary. In order to find the word rapidly in the dictionary the words must be given (for example, in an alphabetic order), however, putting the vocabulary in order to increase the time of the memorizing of the new word. On the other side, if the vocabulary is not put in order then the time of access is greatly increased, but the time of the time of memorizing of the new word is decreased. We shall speak only of the access because this operation is made more often.

1. *Connectedness*

The majority of researchers came to the conclusion that for the successful search of knowledge at the local data it is necessary to organize small units into bigger units than tops and arcs (in the semantic nets) or formulae (in logical calculations). These units are called rather differently: block, knot or chunk. We shall use the term "block" not the "knot" to avoid confusion with the "node" (top)

in the semantical nets. In the structures of the data it is possible to choose two types of links among the elements: external and internal. The syntactical analysis is an example of a program fulfilling the finding of the substructures of the sentences in grammar.

Syntactic Comparison. At present the forms (but not the contents) of two blocks are being compared. In order to simplify the process of comparison all the forms are kept in a unique (canonical) presentation. The comparison is considered successful if as a result the forms (sometimes called samples) are identical.

Parametrical comparison. The result of the syntactical comparison is binary. The samples are compared or not. In the parametric comparison the parameter determining the degree of comparison is introduced.

Semantic comparison. In this case the forms of the elements are not compared but their functions are on the contrary.

Forced comparison. Its essence consists in the fact that one structure is considered from the point of view of the other. In contrast of the usual comparison, here (in the principle) we can get a positive result. The question consists in the "force" of constraint. Special procedures can fulfil constraints connected with the structures. If the indicated procedures cannot establish the success possible under such conditions the considered structures can be examined as being compared.

Conclusions

To store all the facts in the system in reality is both non-effective and impossible. The process of getting new information goes with the help of the concluding rules and facts, is evidently stored in the system. Sometimes we use the word "deduction" in order either to stress the infirmity of the conclusion or to avoid confusion with the very word "conclusion" in the meaning of "withdrawal to display". We distinguish the following types of conclusions: formal conclusion, specialized conclusion, metaconclusion.

1. The Formal conclusion. The present type of conclusion is used during the registration of the facts of the modelled world into some formal language. Thus, a part of the statements corresponding to the initial data, are considered as axioms and the aim of the task is considered as a theorem the equity of which is necessary to establish or disprove on the basis of axioms and rules of the formal system.

2. Specialized conclusion. In a number of systems in order to get certain facts specialized rules or procedures are used. For example, in the world representing the location of objects on the space by means of an evident task of two coordinates a special procedure may take out the fact: object X is left to the object Y. The specialized conclusion has the following peculiarities in comparison with the formal conclusion:

- In the given case is evidently shown the facts and their use. And, therefore, the retrieval process (the overcome) practically is not present.
- Effectivity (the absence of retrieval) leads to some hardness as the framework of the application of the rules is limited. On the other side, namely, the restriction simplifies (or removes) different types of control of the rules applied.

– In the process of realization of a specialized output in the way of procedures it is not possible to get any intermediate results and substantiation of the final result. The result is recognized to be correct without any proof. The unique guarantee of correctness is the fidelity of the procedure.

3. Metaconclusion. Conclusions of the given class are rather important because they are widely used by the man in the process of the intellectual activity. Usually metaconclusions are considered: the inductive conclusion, conclusion on the basis of analogy, the conclusion based on the knowledge about himself.

Inductive Conclusion. The inductive conclusion using some multitudes of facts (A) forms the basis for the general rule. The general rule must be compatible with A, but it is not obligatory to be correct. The rule may be rejected in the process of the appearance of additional information. The usage of the general rule allows in a more compatible way to represent information.

Conclusion on the Basis of Analogy. The essence of the conclusion on the basis of analogy is treated in the following way: If between two situations a certain criterium of analogy is found then the result belonging to the first situation, is spread on the second situation.

Conclusion Based on the Knowledge itself. In order to answer some questions there is no necessity to look for a concrete fact. Sometimes it is enough to know what the system knows (or doesn't know). The essence consists in the following: let some objects of the given class possess the peculiarity X, let the object of the same class, about which it is known enough or even much more, not to contain indications about the presence of the peculiarity X. Then, probably, the given object doesn't possess the peculiarity X.

All these conclusions bring to a good working system understanding, as it was planned, natural languages leading to correct processing of natural languages and to their translation by means of machines.

Bibliographical References

Filmore, Ch.T. "The Case for Case". *Universals in Linguistics Theory.* New York: 1968.

Minsky, H. "A Framework for Representation Knowledge". *The Psychology of Computer Vision.* New York: McGraw-Hill, 1975.

Moore, T. and Newell, A. "How Can Merylin Understand?" *Knowledge and Cognition*. Baltimore: Lawrence Erlbaum Associates, 1973.

Raphael, B. "The Frame Problem in Problem Solving Systems". *Artificial Intelligence and Heuristic Programming.* Edinburgh: University Press, 1971.

Schank, R. C., Abelson R. "Сценарии, планы и знание*". Труды IV Международной Объединенной конференции по искусственному интеллекту*, 1975, т. VI.

--- . *Conceptual Dependency: A Theory of Natural Language Understanding Cognitive Psychology.* 1971, Vol. 3, nr. 4.

Stuart, J., Rusself, P. Norving. *Artificial Intelligence. A Modern Approach*. New Delphi: 2005.

Turafsky, Daniel, James H. Martin. *Speech and Language Processing. An Introduction to Natural Language Processing. Computational Linguistics and Speech Recognition*. Third Indian Reprint, 2004.

Wilks, Y. A. *Grammar, Meaning and the Machine. Analysis of Language.* London: Routledge and Kegan Paul, 1972.