Journal of Engineering and Applied Sciences 12 (Special Issue 4): 6864-6869, 2017
ISSN: 1816-949X

On the Language Handling Technology and its Peculiarities

Alexander G. Zhikharev, Sergey I. Matorin,
Irina A. Kupieva, Stanislava B. Smirnova and Anatoliy I. Boiko
Belgorod State National Research University, Belgorod, Russia

Abstract: The problem of understanding speech is considered as an integral part of the speech data processing. A model of the process of understanding signs is given. A new methodology for representation of knowledge in a computer is described on the basis of the system-object approach "node-function-object". This algorithm is described according to the lacks of other natural language handling technologies. Also, a linguistic marker is viewed from the point of view of its using in recognizing and understanding the natural language's sign.

Key words: Unit, function, object, UFO-approach, linguistic marker, natural language, language sign recognition

INTRODUCTION

Man and computer are the concepts that are rapidly getting closer and closer. At the present stage of the mankind development, it has already become difficult to imagine our everyday life without digital devices. At a rapid pace, automated information processing systems are being introduced into all spheres of human activity, so a person has to interact with a computer, transfer information to it for processing and receive corresponding results. This interaction (human vs. computer) can be realized traditionally in the form of formal commands given by a person through special devices that are easy to be understood by a computer. At the present time, however, the speech interaction of a person and a computer becomes more important which could be provided by solving several tasks.

Recognition of speech which as a result forms an array of signs in a computer corresponding to the spoken audio signals representing words and expressions.

Understanding of speech that results in the formation of the expected response by using computer based on the previously formed array of signs formed by the expected result of human speech effect.

Full-scale speech processing is possible only as a result of solving both mentioned problems. At the same time, understanding of speech, i.e., Natural Language (NL) for the present moment continues to remain the problem unsolved.

SIMULATION OF UNDERSTANDING OF LANGUAGE

What happens when a person hears speech? Here, two situations are possible. The person has previously acquired knowledge associated with the concept represented by recognized from the speech signs. Thus, there appears different bonds in the person's mind by means of operating the given knowledge. For example, suppose a person recognized the word table. In doing so, he/she has the knowledge that the table is a piece of furniture that consists of several parts (cloth and four legs) and is used for the convenient arrangement of objects. There are images of the table and the objects associated with it: chair, work, cabinet, etc. All these processes occur due to the fact that earlier people acquired knowledge in the given subject area and at the moment of pronunciation of the sounds composing the word table, he has knowledge of the corresponding concept. If, for a person, the recognized word is not associated with a known concept, then the processes of obtaining knowledge about the concept represented by the recognized word are started.

At present, almost all computer systems based on knowledge (KBSs) are organized according to the algorithm that was described above in general terms. There is a certain knowledge base which contains knowledge of a specific subject area, presented in a certain way and implemented various associations and mechanisms of inference, through which "thinking", "knowledge" and "learning" are simulated.

We simulate the process of understanding the NL. As a result, we get a model whose context level contains information on the interaction of the voice data source with KBSs. This model is shown in Fig. 1. As, it can be seen from Fig. 1, at the entrance to the information processing system there can be either speech data corresponding to the very concept or speech data corresponding to information about the concept (in the event that the original concept is not known to the

Corresponding Author: Alexander G. Zhikharev, Belgorod State National Research University, Belgorod, Russia

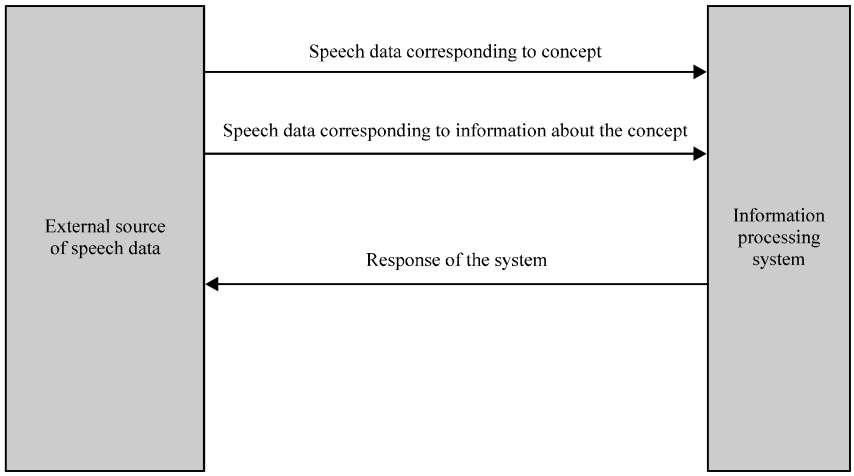

Fig. 1: Context model of natural language understanding

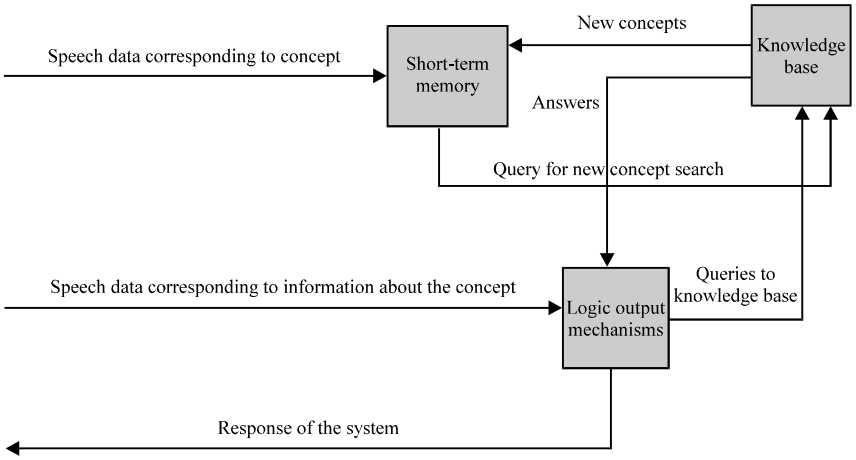

Fig. 2: Reaction model of the system to the irritation with the speech data

system) and the output will be the response of the system to the external "irritation" with speech data. Model of the KBS reaction is pictures in Fig. 2 to the external irritation with the speech data.

As it was mentioned above, when verifying the voice data, it is necessary to check whether there is any knowledge associated with it. To do this, short-term memory serves as the content indexer of the knowledge base, i.e., there are indexed all the concepts used in the knowledge base for a quick search and comparison of concepts that correspond to the speech data from an external source of information. Output mechanisms are the interface between an external source of voice data and a knowledge base, i.e., with the help of these mechanisms and the knowledge contained in the knowledge base is operated. The knowledge base is a long-term repository of knowledge in which it is presented in a certain way. The whole work of KBSs depends on the way knowledge is presented: its effectiveness, speed and quality.

The approach to understanding and strategy of its modeling allows us to consider the result of understanding as a result of the cognitive information process. This approach to the result of understanding allows you to choose for its modeling such an object, the model of which will provide an imitation of understanding by means of electronic computer technology.

To justify the method of understanding modeling proposed in accordance with the adopted strategy, it is necessary to justify the choice of the modeling object and for this it is necessary to determine what is the result of the understanding of the signs of the NL that is in this case of speech by the person.

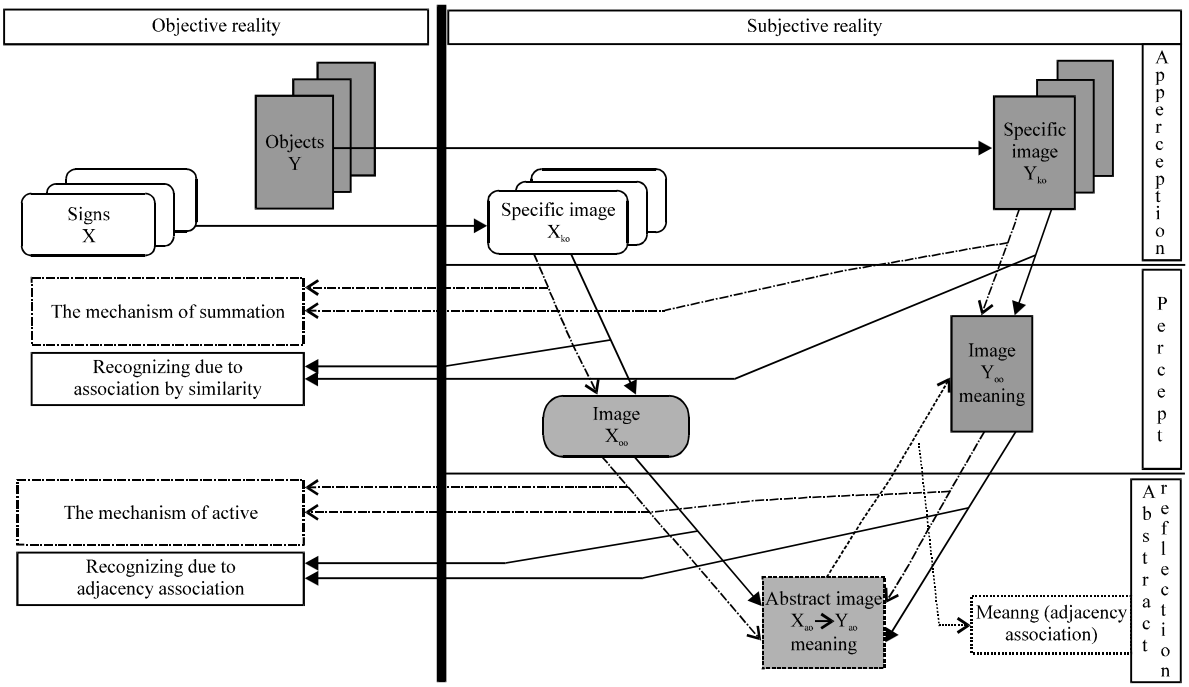

Fig. 3: The sign understanding as cognitive informational process

To solve this problem, a scheme for understanding a sign (Matorin, 1997) was developed, based on the data of experimental psychology, neuropsychology and cognitive science. This diagram illustrates the cognitive information process of understanding the meaning of the sign of the NL (Fig. 3).

The process of understanding a sign includes the stages of its reflection in the mind of a human and proper understanding. Despite, the unity of these stages, there are fundamental differences between them. Let us first consider the process of reflection of any material object in the mind of a person. Under, the influence of the material object Y on the sense organs, there are sensations in the mind, when "adding" at the level of perception, the image of a particular object is formed (Y_{ko} is a concrete image). With repeated perception of this object or objects of this type, the mechanism of summation operates. As a result, repeated, general, sensually perceived attributes of objects are fixed in the human mind which leads to the formation of a generalized image (Y_{oo}) of objects of a given type at the presentation level.

With the accumulation of the individual's experience of reflecting objects of the world and when a certain level of generalization is reached, the mechanism of active search begins to operate and an abstract image (Y_{ao}) of objects of this class or the concept of this class of objects at the level of abstract reflection (theoretical or verbal-logical thinking) is formed. Thanks to the logic by which the mechanism of active search works, the abstract image contains essential features of the reflected class of objects, i.e., in it a person learns the essence of objects of the material world.

If an individual has formed the abstract image of a certain class of objects, then in the process of recognizing the object of this class from Y_{ko}, Y_{oo} is excited due to association by similarity and then because of the adjacency association Y_{ao}. As a result of the identification, the specific perceived object Y is correlated with the corresponding object class. The mechanisms of logical conclusion and active search that operate can increase the degree of generalization and the degree of abstraction of images.

At the first stage of the process of understanding, the sign of X as a result of its reflection and identification as an object in the human mind, an accurate, generalized and abstract images are excited X_{ko}, X_{oo} and X_{ao}, respectively. The essence of the sign lies in the designation of another object, so one can assume that the sign known to a human will be reflected in the mind exactly to the level of the abstract image.

At the second stage of the process of understanding the sign, a transition is made by association along the contiguity (meaning) from X_{ao} to the image of the object Y. Thus, the understanding of the sign of X.

The transition from the image of the sign (X_{ao}) to the image of the object (Y_{ao}) is reduced to the use of the

association arising in the consciousness in the process of forming the concept (object Y) by the contiguity between the given sign for the abstract image of the given object and the system of other signs, i.e., abstract images or concepts. The whole difference between X_{ao} and Y_{ao} is that X_{ao} is a single sign and Y_{ao} is the same sign but included in the system of other signs.

The transition from the image of the sign (X_{ao}) to the image of the object (Y_{ao}) is reduced to the use of the association arising in the consciousness in the process of forming the concept (Y_{ao}) by the contiguity between Y_{ao} and Y_{oo} which works when identifying objects of the class.

The images of the object are the meanings to which according to the adjacency association, a transition from the image of the sign is made. Generalized meaning is formed on the sensory level, abstract is on the abstract level of reflection, respectively. Thus, the specific meaning is the image (Y_{ko}) of a specific object in human mind at the perception level and the generalized meaning is the generalized image (Y_{oo}) of a given type of objects at the level of representation, the abstract meaning (concept) is the abstract image (Y_{ao}) of the class of objects at the level of the theoretical or verbal-logical) thinking.

Analysis of the functions of participants in the sign situation allows us to consider the image of the denotate, i.e., the meaning, as the main result of the understanding process which allows us to justify the proposed approach to the modeling of understanding. If the process of understanding the signs of NL lies in "reconstructing" their meaning as the result of this process, then the formalization of understanding must be carried out by formalizing the meaning of the sign. Moreover, such formalization is possible, since the analysis of cognitive information processes in the understanding of signs shows that the meaning is reduced to a constructive and tangible phenomenon the image of the object indicated by the sign. The image of the object of real reality in the human mind as can be seen from the scheme of understanding the signs of NL has quite definite functions in the system of the given information process, the structure, properties and mechanisms of formation.

Thus, the formalization of understanding and the concretization of the object of modeling are possible in the approach to understanding as to the procedure for the formation of a sense (the image of the object the denotate) which is the result of understanding. Consideration of the understanding of the meaning of the sign as a process of initiation (excitation) of the image of the object designated by this sign, through association by contiguity from the image of the sign, makes it possible to practically solve the problem of modeling understanding. This is due to the fact that the model of understanding with our approach is a model of the image of the object, structurally related to the image of the sign, capable of being excited under the influence of signs entering the computer input. The concretization of understanding and its result allows you to select a simulation object for which it is possible to create a software product in the computer's memory.

REPRESENTATION OF KNOWLEDGE IN A COMPUTER IS THE BASIS OF THE SEMANTIC PROCESSING OF SPEECH

The problem of understanding the NL is closely related to the problem of knowledge representation in computer systems, since from the way information about the surrounding world is represented in the computer all the possibilities for its processing depend.

The most common methods of representing knowledge in computer systems are: network method; production method; frame method. Each of the listed methods as a rule is oriented to the description of any one side of the modeled system (structural network, functional product, object frame). Thus, modern methods of representation of knowledge do not allow to comprehensively describe a certain fact which often makes it difficult to solve problems related to the understanding of computer systems of the NL. Therefore, the task of creating a way of representing knowledge which would integrate the capabilities of traditional methods is relevant.

The Unit-Function-Object approach (UFO approach) is a modern graph-analytical technology for visual modeling of systems (Matorin, 2002; Matorin *et al*., 2007) in which any system is represented as three interrelated aspects (and is considered, therefore as an UFO element): unit is an intersection of incoming and outgoing system links (structural characteristic of the system); function is a process in which incoming communications of the system are transformed into outgoing links (functional, dynamic characteristics of the system).

Object is an entity that realizes the performance of a function (object, substantial characteristic of the system). Each UFO element can have a hierarchical structure, since any node, any function and any object can consist of the corresponding elements which also represent units, functions and objects (UFO elements) (Matorin *et al*., 2015; Zhikharev *et al*., 2015). With the help of such a hierarchy, you can easily describe any system, taking into account its structural, functional and object characteristics. In fact, the configurations of

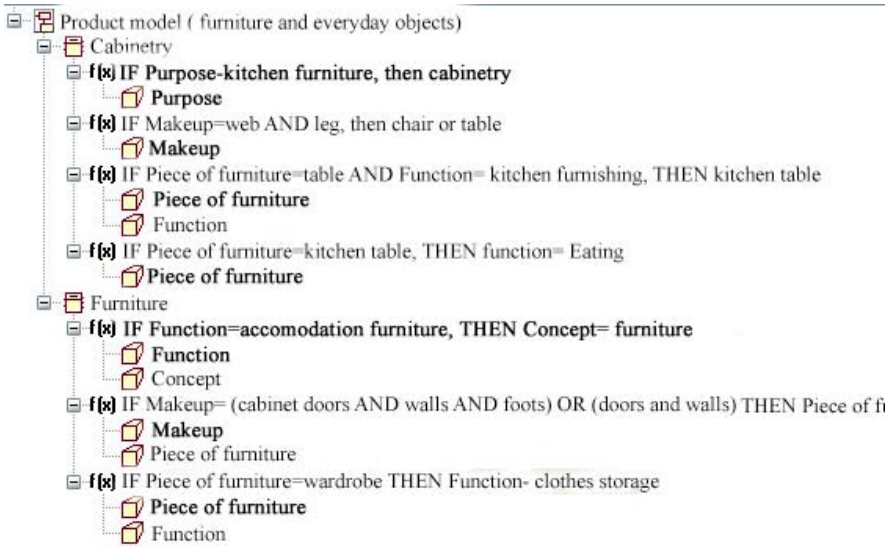

Fig. 4: Production hierarchy (furniture)

UFO-elements are a means for storing knowledge about some phenomena which as it was shown, characterize these phenomena from the point of view of both the structure and the function and the object simultaneously. Thus, the representation of knowledge with the help of nodes, functions and objects gives a complete description of the presented knowledge, their structural, functional and objective characteristics.

In modeling the understanding of speech, a description of the structure of spoken concepts is indispensable. It is very difficult to implement a logical inference mechanism or any other operations with concepts. To solve this problem, it is necessary to store the rules of inferences, i.e., meta-knowledge. These rules are conveniently presented in the form of products that can also be modeled using the UFO approach. For example, consider the production system for the subject domain. A fragment of the product hierarchy in the node-function-object notation is shown in Fig. 4.

In case of a description of a system of products, nodes act as elements of grouping rules on a particular topic, functions-define rules pertaining to a specific node, objects-describe the structure of the rule, i.e., each property of the rule (appointment, composition, piece of furniture, etc.,) corresponds to the object in the UFO notation. Links between nodes transmit values of rule properties, for example, the value of the "composition" property.

Normally a sign is viewed in NL by means of recognizing special linguistic marker. In linguistics, a marker is a free or bound morpheme that indicates the grammatical function of a labeled word, phrase or sentence. Most often, markers occur as either a cliches or form-building affixes. In analytical languages and agglutinative languages, markers are usually easily distinguished. In inflectional languages and polysynthetic languages, this is not always the case. For example, in Latin, a strongly inflected language, the word amo ("I love") is denoted by the suffix -o which shows the modality, the real voice, the first person, the singular, the present tense. Analytical languages have a relatively limited number of markers.

Markers should be distinguished from the linguistic concept of significance. An unmarked form is a basic "neutral" form of the word is usually used as a dictionary of the lemma, for example, in English nouns in the singular (for example, a cat instead of cats) and for the infinitive (e.g., is instead, ate and had eaten). Unmarked forms (for example, the nominative in many languages) tend to have fewer markers but this is not true for all languages, except, for example, Latin. On the contrary, the marked form can have a zero affix such as the genitive case of the plural of certain nouns in Russian (for example, a boot). In some languages, the same forms of the marker have several functions, for example, when they are used in cases and declensions (for example, in Latin).

When studying the problem of formalizing linguistic knowledge with respect of punctuation, it became clear that the functions and rules for the use of punctuation marks attracted special attention of researchers at the first stage of work on the problem of automation of translation. Later, interest in this issue decreased and gave way to key issues of syntax. Thus, punctuation in the applied aspect turns out to be among the least studied questions.

An obligatory condition for the application of methods of cybernetics in this or that field of knowledge is the preliminary formalization of the regularities to be studied in the form of a (more or less developed) sign system.

Consideration of the system of concepts helps to establish that formalization is the basis for the processes of algorithmization, programming and computerization of linguistic material.

In the study, we are talking about the importance of formalization in linguistic studies, presented theoretical material on the problem of formalization in the domestic and foreign linguistics and different points of view on the possibility of formalization in linguistics. Already in 1894 F. de Saussure came to the conclusion that the fundamental relations between units of language can be expressed regularly with the help of mathematical formulas. In linguistics, two opposing approaches to the problem were formed. Some scholars do not recognize the possibility of formalizing linguistic descriptions. According to A. Tarski, language can not be classified as a class of formalized systems. Other scientists present a different view. Yu.S. Stepanov, for example, explains the reduction of existing linguistic descriptions to a schematic and economical form by "the emergence of formalization in the depths of tradition", the researcher divides R.V. Pazuhina compromise position which is that formalization is possible in the event that it is used in the application of the formal description of linguistic facts are used with regard to their incomplete adequacy to achieve limited practical purposes. On this basis, it is quite natural to talk about the possibility of formalizing the language punctuation system and in particular, the structural organization of a simple sentence with the word "it" between the main members and the dash in front of it which is possible thanks to the automation of the punctuation rules. It is necessary to reckon with that; that fully formalized can be meaningful only those areas that are basic in terms of their logical structure and that formalization is more possible where the content is expressed clearly outlined thematically and is characterized by a certain standardized.

CONCLUSION

It can be concluded that the storage of knowledge by means of nodes of functions and objects gives a detailed and complete description of the facts, phenomena, etc., i.e., concepts of natural language. The formalization of linguistic knowledge as a method of describing and studying a language is largely a logical consequence of the results to which structural linguistics has come. Structural description is the highest degree of formalization. Therefore, the implementation of the knowledge representation method based on the UFO approach (Zhikharev *et al*., 2016; Kondratenko *et al*., 2017) will allow to integrate the traditional ways of presenting knowledge into one universal way which will improve the efficiency of not only understanding the speech by the machine but also working with knowledge in general.

ACKNOWLEDGEMENTS

Researchers are supported by grants of the Russian Foundation for Basic Research: 16-07-00460-a and 16-07-00193-a and the President's Grant for the state support of young Russian researchers of PhD of the Russian Federation MD-2687.2017.6.

REFERENCES

Kondratenko, A.A., S.I. Matorin, A.G. Zhikharev, A.N. Nemtsev and I.N. Riabtceva, 2017. Application of logical output means on ontologies to UFO models of subject domains. J. Eng. Appl. Sci., 12: 1347-1354.

Matorin, S.I., 1997. [On modeling of intellectual understanding of the language of business communication]. Master Thesis, Nuclear Threat Initiative, Washington, USA. (In Russian)

Matorin, S.I., 2002. [On a new method of systematic analysis, coordinated with the procedure for object-oriented design (In Russian)]. Cybern. Syst. Anal., 1: 118-130.

Matorin, S.I., A.G. Zhikharev, K.V. Korchagina and T.V. Saitseva, 2015. About the development of new tools system-object simulation process and traffic. Intl. J. Appl. Eng. Res., 10: 45292-45296.

Matorin, S.I., O.A. Zimovets and A.G. Zhikharev, 2007. [Information management technology based on the system-object approach node-function-about bulletin of the National Technical University]. Master Thesis, Kharkiv Polytechnic Institute, Kharkiv, Ukraine. (In Russian)

Zhikharev, A.G., S.I. Matorin and N.O. Zaitseva, 2015. About perspectives of simulation technological processes functioning with using system-object approach node-function-object. Intl. J. Appl. Eng. Res., 10: 31363-31370.

Zhikharev, A.G., S.I. Matorin, O.A. Zimovets, M.S. Zhikhareva and V.I. Rakov, 2016. The simulation modeling of systems taking into account their internal parameters change. Intl. J. Pharm. Technol., 8: 26933-26945.