

Using Mathematical Methods and Information Technologies in Linguistics

Bikesh Revovna Ospanova

Karaganda State Technical University, Mira blvd, 56, 100027, Karaganda, Kazakhstan

Abstract: Linguistics is one of the most computerized and mathematized humanities that is proved by the presence of such fields of science of language, as mathematical linguistics, linguistic statistics, applied linguistics, computer linguistics. Present day linguistic studies require using information technologies and mathematical methods of information processing for the revealing of linguistic phenomena essence. Mathematics is a universal language of science and a powerful means of solving applied problems in a lot of fields of science. Thus, the most urgent problem of present day linguistic education is the use of mathematical and information technologies in carrying out scientific research.

[Ospanova B.R. **Using Mathematical Methods and Information Technologies in Linguistics.** *Life Sci J* 2014;11(4s):373-375]. (ISSN:1097-8135). <http://www.lifesciencesite.com>. 67

Keywords: linguistics, information technologies, mathematical methods, interaction, entropy, structure, language, text, theory of information, study.

1. Introduction

With rapid development of information technologies mathematics becomes more often an effective instrument of studying linguistic objects, namely: there increases the volume of statistical linguistics and other information requiring mathematical processing and computational experiments. Mathematical methods penetration into linguistics is connected, first of all, with the development of the basics of theoretical and practical language base requiring more accurate and objective methods of analyzing language phenomena. At the same time the use of mathematical methods in systemizing, measuring and generalizing the linguistic material permits linguists to carry out deeper studies of the applied language aspects and the text structure.

The main part

The interaction of linguistics with other sciences, for example, with mathematics, biology, cybernetics, informatics, etc. can be performed only when using the mathematical apparatus. Mathematics penetration into linguistics is especially deep in connection with searching for a solution of the linguistic analysis problems at the level of information systems.

Interesting studies with the use of mathematical methods and computer technologies are being carried out within the frames of such complex scientific trends, as "Computer and mathematical modeling", "Applied linguistics", "Quantitative linguistics". In linguistics there are born new synthetic structures combining into an integral unity a number of conceptions and helping, finally, to build a general theory of language as a result of a lot of experimental studies.

A special role in the course of this interaction is played by mathematics. It is not only the instrument that helps to solve a diversity of problems but it permits to use a number of ideas and approaches developed in other fields of human knowledge. This can take place only based on interdisciplinary studies, where it is necessary to use mathematical methods and up-to-date means of information processing. In other words, the generalizing point is mathematical modeling the language objects based on computer technologies.

Linguistic-mathematical methods are successfully used in solving both theoretical and practical problems: theory of grammar, theory of artificial languages, automatic processing, translating from one language to another, etc. theoretical studies in this field were carried out by Piotrovski R.G. [1], Potapova R.K. [2], Boguslavskaya V.V. [3], Bayevski I.V. [4], Sadur V.G. [5], Gorodetski B.Yu. [6]. From the unified positions there were described the basic methods and ways used for linguistic models statistical processing, for developing formalized analogues with their realization using information technologies. They wanted to show how the widening of the arsenal of mathematical methods in studying the language structure for obtaining various information opens the prospects for building linguistic-mathematical models based on computer technologies, since a wide use of automation means and its efficiency increase depends in a large extent on the quality of the carried out experiments.

The use of theory of information in the vocal expression was developed by such foreign scientists, as R. Jacobson [7] and the employees of the Prague School. Theory of information became a field of active work for linguists who tried to use some concepts and methods of theory of information in

solving the problems of linguistics. Special attention was attracted to the Shannon's measure of entropy, or "information of choice". Entropy measuring was considered as a possible means for quantitative description of language processes and structure. The work in this direction was continued by Greenberg [8], Osgood [9], R. Kehler [10], E. Yanch [11] and others. For example, E. Yanch in his work presents an original conception of synthesizing natural sciences and humanities, and R. Kehler reconstructed a language model in the unity of its structure and dynamics.

For the productive work with information a person needs a certain level of information culture. This is an ability to work purposefully with information and to use it for obtaining, processing and transferring through up-to-date means and methods. For example, automation of the language material processing takes place with the use of mathematical calculations, where there are combined "hard sciences" and "humanities" which become a fertile soil for new discoveries in the field of linguistics and informatics.

Electronic translation from one language to another is a quickly developed branch of information technologies. In spite of the fact that computer translation cannot be compared with a translation done by a man, the machine became an integral helper in translating large volumes of the text. It is considered that in the nearest future there will be developed more perfect translation systems based, first of all, on the text semantic analysis.

The occurring of information technologies, especially the Internet, led to certain changes in the language that serves this sphere. These language changes are so scaled and global that some linguists speak of occurring a network, or an electronic language. English linguist D. Crystal in his book "Language revolution" [12] called the Internet occurring one of facts (in his opinion, the most significant) that led to revolutionary transformations in the language. When generalizing a number of works on studying electronic communication, there can be separated several trends:

- studying the common tendencies in the network language, such as complication of some and simplification of other speech means connected with the plan of expression, content and plan of the Internet pragmatic intentions;
- studying the genres and/or discourse formats of the Internet, i.e. typical forms of speech presenting a functional style with a certain list of constitutive signs: electronic mail, electronic talks, chats, electronic bulletin boards (BBS) and computer conferences.

Besides, D. Crystal in his monograph "Language and the Internet" underlines a number of genres calling them *broad Internet-using situations* [13]: electronic mail, synchronous and asynchronous chats including BBS; virtual worlds (MOOs, MUDs, MUCKs, MUSES, etc.); WEB-texts to which there are referred electronic texts with hyper- and linear structure (e.g. texts in *pdf* format).

Thus, in the present day world linguistics became a foundation for developing information technologies. While informatics is a vigorously developing field of the human activity, the union of mathematics and linguistics will continue playing its role in the development of science. Besides, with occurring computer technologies mathematics becomes an integral part of almost all the elements of the human activity. At present mathematical methods are effectively used in all branches of science: linguistics, pedagogy, medicine, psychology, economy, etc.

There is continued studying a language by the methods of statistics that permits to define its qualitative characteristics more accurately. It is important the most spirited hypotheses of a language to fins their mathematical prove. It should not also be forgotten that without developing the internal experimental and theoretical studies in a concrete science, namely in linguistics, it is impossible to advance only by the way of formal using mathematical means. To build a substantial mathematical model in linguistics it is necessary to strip the language laws, there are needed, first of all, substantial primary linguistic hypotheses.

In the present day information world linguistics widens the sphere of its interests. Statistical laws of the language and text functioning have always been a linguistic subject. Now they are becoming more and more relevant, as well as the analysis of the units that are separated on the basis of these statistical laws. Alongside with single texts dealt with by linguists before, the object of linguistics become text collections, information flows as objects of a new information area.

At present the interaction between mathematics and linguistics is more often used in practice. For linguistics a very important measure is the language entropy. The language entropy is a common measure of the probabilistic-linguistic ties in the language text. When calculation entropy there is needed an object mathematical model. The object of study is a text that is built with the use of the language system and that contains a mathematical model. Entropy is a measure of uncertainty. Therefore in any definition of entropy there is present a mathematical calculation of the text and its description, as well as additional

characteristics of the model permitting to measure uncertainty.

For example, as a result of the carried out experiment there was developed a linguistic-mathematical model of the text information-entropic characteristics built on the basis of the fundamental law of preserving the sum of information and entropy using Shannon's formula. In the general characteristic of the text entropic-information analysis there was used Shannon's statistical formula to determine the text perfectness, harmony:

$$H = -\sum_{i=1}^N p_i \log_2 p_i \quad [14]$$

where p_i is probability of detecting some unit of the system in their multitude N ;

$$\sum_{i=1}^N p_i = 1, \quad p_i \geq 0, \quad i = 1, 2, \dots, N$$

The entire entropic-information analysis was carried out on the wide different-genre text material belonging to different styles. This permitted to develop a linguistic-mathematical model for calculating the indicators of determination and stochasticity of the text in Russian and Kazakh for the complex comparative study.

Thus, the need for studying the laws of the text structure within the frames of mathematical modeling was conditioned by the urgent tendencies of interaction between pedagogical innovation technologies and the creative by its meaning character of the educational system in the course of the prior national project in the field of linguistics, mathematics and information.

Conclusion

We believe that the present day condition of the linguistic science is impossible without addressing to the possibilities of mathematical and computer modeling. In this aspect there are widely developed both general theoretical and practical studies in the field of linguistics with the use of models based on a diverse mathematical apparatus, without which involving it is impossible to understand completely the language nature.

The process of linguistics approaching hard sciences impacts the unification of the conceptual and categorical apparatus of science. The use in the linguistic science of mathematical concepts does not substitute the conceptual apparatus of this science but only assists its improvement and development of

some its sections by widening the frames of the traditionally qualitative approach to the linguistic problems.

Thus, mathematical methods used in such scientific-practical fields of linguistics, as linguistic synergy, theory of vocal communication, linguistic didactics, theory of script, theory of translation, etc., increase the accuracy of the results and give a possibility of more objective explanation of language phenomena, they contribute into modernizing the system of higher vocational education od mathematical and humanitarian trends, define the prospects of scientific research at the intersection of adjacent fields of hard sciences and humanities.

Corresponding Author:

Dr. Ospanova,
Karaganda State Technical University, Mira blvd, 56,
100027, Karaganda, Kazakhstan

References

1. Piotrovski R.G. 1996. Linguistic synergy. Sc.-tech. information. Ser.2. Information processes and systems. No 12.
2. Potapova R.K. 2002. New information technologies and linguistics. – Moscow.
3. Boguslavskaya V.V. 2008. text modeling: linguistic-and-sociocultural conception. Analysis of journalistic texts. – M.: LCI,– 280 p.
4. Bayevski V.S. 2001. Linguistic, mathematical, semiotic and computer models in history and theory of literature. – M.: Slavonic culture languages. – 336 p.
5. Sadur V.G. 1983. Vocal communication with electronic-computing machines and problems of their development // Vocal communication: problems and prospects, Moscow.
6. Gorodetski B.Yu. 1989 Computer linguistics: modeling language communication // Innovations in foreign linguistics. Iss. XXIV. Computer linguistic. Moscow.
7. Jacobson R. 1964. Developing a language target model in European linguistics in the period between two wars// Innovations in foreign linguistics. Iss. IV. M., 374 p.
8. On language: selected writings of Joseph H. Greenberg. 1990. Ed. by Keith Denning and Suzanne Kemmer. Stanford: Stanford University Press.
9. Charles E. 1990. Osgood & Oliver Tzeng (eds), Language, Meaning, and Culture: The Selected Papers of C. E. Osgood. Praeger Publishers.
10. Kohler R. 1993 Elements of synergetic linguistics. Proceedings of the st Quantitative linguistic conference. Dordrecht.
11. Jantsch E. 1992. Die Selbstorganisation des Universums // Vom Urknall zum menschlichen Geist. Munchen.
12. Crystal D. 2001. Language and the Internet / D. Crystal. – Cambridge : Cambridge University Press.
13. Crystal D. 2004. The Language Revolution. D. Crystal. – Cambridge : Policy Press.
14. Shannon C. 1963. Works on theory of informatics and cybernetics. Under ed. of R.L. Dobrushin and O.B. Lupanov. M.: Publ. Foreign Literature. 827 p.