

AUTOMATION OF RESEARCH AND RECONSTRUCTION OF LANGUAGES

Nikolaev Y.A.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Lazarenko A.M. – senior lecturer

Annotation. A project to automate the research of languages and their reconstruction is considered. The project includes the creation of a computer program, the capabilities of which will include: storing linguistic material in a structured format and analyzing this material, classifying languages and determining the relationship of languages, creating a generalized mixture of selected languages (which will help linguistic scientists in the reconstruction of proto-languages).

Keywords. language learning, research optimization, machine approach, new concept, software, analysis, sounds, graphic symbols, data collecting, program accuracy, dialects, language relationship, saving, arithmetic mean, AI-created, proto-languages, reconstruction.

People have always been interested in the topic of language learning. However, it is one thing to learn a language for personal use as a means of communication, and quite another thing to research it at a deep level, as a phenomenon and a clear system. Language research involves the great resources and strength of many linguists. And in order to reduce costs and optimize the research process, the use of a machine approach was proposed. The concept of the software that is proposed is as follows: a program that, based on the collected data, classifies a language according to its characteristics, determines which language family it belongs to, analyzes sounds and how they are reproduced by the vocal apparatus, it follows from the features of the program, that it will divide the sound recording into phonemes and decode the phonemes into graphic symbols, after which you can listen to each individual phoneme, since it will be output to a separate sound file.

To accomplish all of the above, several fundamental steps were developed and proposed:

- Collect the language material of native speakers who live preferably in the same area (village or settlement).
- Develop a unified algorithm for classifying languages.
- Analyze all the languages of the world and hone the accuracy of the program in well-researched languages.
- Research small and poorly researched languages and dialects in order to determine the degree of their relationship with other languages and for their further revival and preservation from the final loss.

On the example of the evolution of word “four”, you can see how all Indo-European languages are related to each other.

At the stage when there will no longer be a shortage of linguistic material, it is planned to create an arithmetic mean between all researched languages using AI. Artificial intelligence, based on the analysis of changes in languages from region to region and their general characteristics, will bring the data and structure of all languages to the average and most general state. It will be a mixture of all the languages researched. This artificially created language will help scientists research and reconstruct proto-languages.

References

1. Эволюция слова «четыре» от протоиндоевропейского к различным современным европейским языкам: [Electronic resource]. – URL: https://www.reddi.com/r/europes/comments/46q37t/the_evolution_of_the_word_for_four_from/