

The Role of Information Technology in the Development of Linguistics

Isayeva Narmina Shua

Azerbaijan Technical University, Baku, AZ1134, Azerbaijan

**Corresponding author. Email: narmina.isayeva@mail.com*

ABSTRACT

The article discusses the importance of information technology and its role in society, its impact on the science of linguistics. Our modern society is built on information technology. Information technology is currently being applied to all fields of science. One of them is linguistics, one of which is called computer linguistics. It is well known that the collection, systematization and analysis of language materials have historically been difficult. Thanks to the use of information technology, this process has become much easier. The creation of language corporations, in particular, has become the most important issue. Thus, the analysis, study and learning of the vocabulary, dialects, phonetic, lexical and grammatical features of specific languages has become an easy process with the help of language corpuses. The main step in learning a language is to learn its vocabulary and meaning. Therefore, the dictionary block of language corpuses must be set up first of all. The dictionary block is a system that can be applied to several dictionaries at the same time and clearly illustrates the meaning of the word.

Keywords: *information society, information technology, linguistics, new areas of linguistics, language corpuses*

1. INTRODUCTION

The development of the ever-changing society is accelerated by information technology. This rapid development has an important role in the development of linguistics as in all areas of science. Thus, the representation of language in an electronic environment, and the more effective effect of scientific research on the electronic method, makes it necessary to introduce information technologies into the language. The ideology of the information society has been the subject of scientific-theoretical research for more than 50 years. E.Toftler, D.Bell, M.Castells and other prominent scholars, who played an exceptional role in the formation of this ideology, sought to define this new phase of human history from various aspects [1, p.3]. Therefore, there are various approaches to the ideology of the information society. The main goal of the information society is to create a wealth of information both to solve the problems facing humanity and to ensure the development of society. Against this background, fundamental research is being conducted to investigate the role and place of language in the virtual space [2, p.789; 3].

When we approach to the concept of information society from a technological point of view, we see that the computer and the Internet are the main technical means of collecting, storing, processing, transmitting and other important functions [1, p.5].

2. METHODS

2.1. The concept of computational linguistics

Computer linguistics began to develop in the 60s of the 20th century. Computer linguistics differs from other areas of applied linguistics in terms of processing of language materials on computers. In modern times, computer science, which is one of the fields of science that operates as an independent field, uses many new concepts and terms. Computer linguistics, which is a very broad concept, combines different fields of science. These include machine translation, computer dictionary, the creation of national machine funds and national corpuses, and so on. These areas, in turn, are broad concepts and combine several specific areas.

Since computer linguistics is a science field that develops a mathematical model of natural languages in human intellectual activity, the basis of this field is the use and development of language for computer systems. In this regard, the issues learned by computer linguistics can be grouped as follows:

- 1) Processing of the natural language (This provides syntactic, morphological, semantic analysis of the text.);
- 2) Automatic translation systems of texts;
- 3) Selection of facts, information, information from texts;
- 4) Creation of expert systems- knowledge system;
- 5) Creation of questionnaires (dialogue);

- 6) Optical recognition of symbols. (For example: Fine Reader);
- 7) Automatic speech recognition;
- 8) Automatic speech synthesis;
- 9) Development of information retrieval systems;
- 10) Corpus linguistics (including the creation and use of electronic corpus of texts);
- 11) Creation of electronic dictionaries [6, p.5].

Each of the areas of computer linguistics is new areas that require in-depth scientific research. The advancement and development of computer technology, becoming the most important and integral part of the human community opens new dimensions in the face of the rapidly evolving computer science.

2.2. Language corpora

One of the most relevant and widespread areas of computer linguistics is corpus linguistics. In linguistics, the corpus has different definitions.

According to the Russian linguist — V. P. Zakharova, the corpus is a German origin word (*das Corus*), a collection of texts used for linguistic analysis. These texts contain thousands, millions of words, and are stored in a computer's memory. Most modern corpora are systematized. This means that the texts are systematized according to their characteristics, that is, genres and dialects. The corpus is a collection of systematized, computerized texts. The corpus is created to learn any language. Other sets of languages are used for other purposes. The composition of the corpus should support the purpose of language learning [7, p. 5].

Prof. M. Mahmudov, in explaining corpus linguistics and the language corpus that are its subject, notes that corpus linguistic predetermined texts representing any particular language, aimed at systematic collection in electronic format, learns the issues about creation of information retrieval systems and their use in research. Within a certain period of time, language is fully represented by all genres, styles, literary language, dialects, grammar, dictionaries, and so on in the national corpus of a specific language. According to the author, the corpus is a plurality of natural language texts that provide researchers with operational and accurate information about the various language events and facts stored on the electronic carrier [4, p.44].

The use of the word “national” in corpus linguistics is related to the distinction of British English in one of the earliest corpora. Since there are many variants of the English language used in Britain, the word “national” has been added to emphasize the English language used in Britain, in order to separate it from other options. This tradition was later preserved, and when the corpus of other languages was created, the word “national”, which was in the name of the first corpora, was preserved and used in others.

The information search system, based on the electronic collection of texts in a particular language, is called corpus. The language of the national corpus, at a certain stage of its existence, represents all of its variants. It is

used for scientific research and teaching of the language. Researchers create working tools for linguistic research in the field of corpus linguistics and, with the help of these tools, they undertake the study of linguistic events and the clarification of their meaning.

Corpus linguistics is a field of linguistics dealing with the creation and use of text corpora. This term first appeared in the 1960s on the creation of corpus. Since the 1980s, corpus linguistics has entered a more enthusiastic development phase, which in turn has spurred the development of computing techniques. In those years, the language corpus was classified as a collection of texts that were collected according to certain principles, marked with certain standards and equipped with a special search engine.

3. RESULTS

Computer dictionary should be learned in this context as an area of computational linguistics. The term also includes the term lexicography, which was also fashionable and overused.

Computer dictionary differs from ordinary dictionaries by the following characteristics:

- 1) The procedure for accessing the dictionary and obtaining information based on certain queries is simpler and faster.
- 2) It is possible to refer to several dictionaries at the same time, to clarify the meaning of the word as soon as possible, and to draw a general result.
- 3) Each of these dictionaries has the ability to highlight any language and compare it with other dictionaries.
- 4) Ordinary dictionaries remain unchanged until they are printed, and the number and explanation of the words cannot be changed. Computer dictionaries are open, dynamic systems. It is possible to add new words to such dictionaries and extract archaic ones [5, p. 804].

As it is known, computer lexicography creates linguistic base of computer dictionaries, lexicographic works and are an area of application linguistics focused on software.

The main tasks of traditional and computer lexicography are to determine the structure of dictionaries, vocabulary articles, and to develop the principles of compiling different types of dictionaries. Traditionally, the dictionary consists of collected words and their explanations. Electronic (automatic, computer) dictionaries are a collection of words in computer format. These words are used by humans and are considered to be a major part of more sophisticated computer programs (for example, machine translation).

Automatic dictionaries for human use are computer versions of the most popular traditional dictionaries. For example, “Explanatory Dictionary of the Azerbaijani Language” (4 Volumes), “Spelling Dictionary of the Azerbaijani language”, “Dictionary of abbreviations of Azerbaijani language” etc. [9].

Automatic dictionaries are dictionaries that are in a specific machine format intended for use on computers or

for computer programs of texts. In other words, there are two different types of automatic dictionary:

- 1) Automatic dictionaries for human use;
- 2) Automatic dictionaries designed for text processing programs.

Articles in automatic dictionaries intended for human use It differs from other dictionaries by its structure, such as machine translation, automatic reference system, data retrieval, etc. Often, computer versions of common dictionaries are used as automatic dictionaries there is a computer version of the English Explanatory Dictionary, as well as S. I. Ozhegov's dictionary in the software market.

Russian linguist, lexicologist, researcher of norms of Russian literary language, doctor of philology Ozhegov Sergei Ivanovich (1900-1964) prepared "Dictionary of Russian language" in 1949. In 1990, the 22nd edition of this dictionary was published. Since 1992, "Explanatory Dictionary of Russian language" along with the N.Y.Shvedova prepared, but it is known by the name of dictionaries S.I.Ozhegov [10].

In the current online version of Explanatory Dictionary of Russian language, S.I.Ozhegov implemented three types of search engines:

1. Search by alphabet of the dictionary book version, which is known to us – is placed at the right side of the page. This is a complete analogue of the search system known to most of us from S.I.Ozhegov's book. You are writing the first letter, and you are looking at the words that are in this letter until you find that word, with 50 words each time. Then you click that word to look up the dictionary article with the explanation of the word found.

2. Searching for a dictionary article which is exactly related to the word or phrase. If you remember the exact writing form of the word you are looking for, enter it into the search engine, the online dictionary search system will give you an immediate explanation of the dictionary article you are looking for.

3. Switch to online expanded search. If you do not remember the exact writing form of word and phrase, but you know a few start letters, in such condition use the search engine for the word part. To do this, go to the "Advanced Search Engine" page in the upper menu.

The "Explanatory Dictionary of Russian language" consists of 80,000 words and phrases (with headline words and derivative words). The words and phraseology included in the explanatory dictionary applies to the Russian literary language as well as specific language spheres that are interconnected with it. Simple speech lexicon, which is observed in literary works and spoken language, is also widely represented in the explanatory dictionary. Dictionary articles include information on word descriptions, structure characteristics, polysemy, examples of literary and spoken language, word combinations and grammar characteristics. The dictionary articles in S.I.Ozhegov's dictionary reflect the information about phraseological associations created from any word or related to it in one or other side.

This type of automatic dictionary replicates the structure of dictionary articles in simple dictionaries simultaneously. However, they have many advantages over paper dictionaries. Many dictionary users use the automatic dictionary of simple dictionaries according to their purpose. For example, it is possible to select all the adjectives or verbs in the dictionary automatically as a list.

4. DISCUSSION

Research and analysis shows that information technology has a direct impact on the development of linguistic science. Due to these new technologies, changes in the language or changes that need to be made can be tracked in a short period of time. All this shows that special attention should be paid to the development of computer linguistics as a science.

The creation of national language corpus that incorporates all the peculiarities of the language should be one of the primary tasks facing linguistics.

Language corpus, in turn, is subdivided into phonetic subcorpus, lexicographic subcorpus, subcorpus of dialect texts, and so on, which facilitates the search process for linguistic researchers. Language corpus helps researchers in the process of research, learning the language, speech recognition by means of sound corpus and in teaching process.

The creation of a national language corpus is a pressing issue. The issue of developing a block of dictionaries within the national language corpus is important as an integral part of the computer dictionary.

The possibilities of modern information technologies and mechanisms should be widely used in the formation of glossary blocks and in the development of its software.

REFERENCES

- [1] Aliquluyev, R.M, Valehov, J.F, Mahmudov, R.Sh. Some aspects of the formation of Internet journalism. Express information. Information Society Series. Baku: Information Society, 2008, 32 p.
- [2] Habibova, K.A., Jafarov, Y.M. Language policy in the virtual space //AEBMR-Advances in Economics Business and Management Research: Proceedings of materials1st International Scientific Conference "Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth" MTDE 2019 Volume: 81, p.:789-792. DOI: <https://doi.org/10.2991/mtde-19.2019.161>
- [3] Habibova, K. A. The Specificity of Everyday Discourses // AEBMR-Advances in Economics Business and Management Research: Proceedings of the First International Volga Region Conference on

Economics, Humanities and Sports (FICEHS 2019), Vol. 114, Atlantis Press, Paris, 2020, p.: 335-340. DOI: <https://doi.org/10.2991/aebmr.k.200114.079>

[4] Mahmudov, M.A. Computer Linguistics. Baku: Science and Education, 2013, 356 p.

[5] Mammadova, R.H., Guluzade, S.F., Aliyeva, N.A. The issues on the optimal structure of lexicography block in national language and processing of its software. // AEBMR-Advances in Economics Business and Management Research. (Proceedings of "1st International Scientific Conference "Modern Management Trends and the Digital Economy: from Regional Development to Global Economic Growth" MTDE 2019", Yekaterinburg, Russia, 14-15 April, 2019), vol: 81, Atlantis Press, Paris, 2019, p.: 802-805, DOI: <https://doi.org/10.2991/mtde-19.2019.164>

[6] Mirzayeva, M., Veliyeva, K. Theoretical problems of machine translation. Baku, "Science and Education", 2016, 120 p.

[7] Zakharov, V.P. Corpus Linguistics. The methodical allowance. St. Petersburg, Saint Petersburg State University, 2005, 48 p.

[8] Castells, M. The information age: economy, society and culture / Per. c English Subsystem. Ed. O. I. Shkaratana. - M.: GU WShE, 2000. 608 p.

[9] <http://korpus.azerbaycandili.az/>

[10] <http://www.ozhegov.org/>