



www.jpis.az

15 (2)

2024

ICT problems of the Azerbaijani language, the Azerbaijani language problems of ICT

Rasim Alguliyev¹, Yedgar Jafarov², Ramiz Aliguliyev³

^{1,2,3} Institute of Information Technology, B.Vahabzade str., 9A, AZ1141, Baku, Azerbaijan

¹r.alguliev@gmail.com, ²yedgar.jafarov@gmail.com, ³r.aliguliyev@gmail.com

¹<https://orcid.org/0000-0003-1223-7411>, ²<https://orcid.org/0000-0001-8205-8000>, ³<https://orcid.org/0000-0001-9795-1694>

ARTICLE INFO

<http://doi.org/10.25045/jpis.v15.i2.01>

Article history:

Received 01 February 2024

Received in revised form

01 March 2024

Accepted 06 May 2024

Keywords:

Language and ICT

Computational linguistics

Language digitization

National corpus of languages

Natural language processing

ABSTRACT

The article is dedicated to the intersection of language and technology, a crucial topic in our contemporary world. The rapid advancements in Information Communication Technologies present both new opportunities and challenges for national languages. As information technologies permeate various aspects of our lives, encompassing communication, the vitality and progression of language occur within this technological landscape. To delve into these dynamics, the article examines the case of the Azerbaijani language. The research is bifurcated into two key areas. The first addresses obstacles hindering the language from fully leveraging Information Communication Technologies capabilities, while the second explores limitations impeding Information Communication Technologies from effectively utilizing the language's potential. The article provides a thorough analysis of the issues identified in both aspects and proposes solutions to overcome these existing challenges.

1. Introduction

A digital divide, in other words, an uneven distribution of computer technologies and Internet access among different regions and population groups has been maintaining its relevance for many years as one of the global problems of the information society.

For many years when the issue of the digital divide has been discussed at the global level, the main focus has been on the technical aspects of this issue, and relatively little attention has been paid to the language issues. In recent years, taking into account the realities created by the globalization processes and the rapid development of technologies, sensitivity to the language aspect is increasing, and discussions in this direction are expanding.

The problem of digital divide from the technical point of view according to some criteria,

such as technical supply of the population, including the number of computers, other portable devices, mobile phones per capita, Internet access in the country etc. is not at a threatening level for Azerbaijan [www.azstat.org, 2024]. In terms of the language aspects of the information society, this problem is quite relevant for Azerbaijan, and the relevance of the issue is increasing over time.

Obviously national languages such as Azerbaijani have been seriously affected by globalization processes in recent decades. The English language, based on such strong foundations, as huge economy, financial, information, knowledge resources, including influential media, wide communication circle, etc. is gradually weakening the position of the national languages. The increasing dominance of English, as well as a small number of other global languages, in the Internet and information communication technologies (ICT)

environment in general increases this effect. Therefore, facilitating the use of ICT in vernacular languages and finding both content and tool solutions in this regard is a necessity today. Because millions of people speak these languages, including Azerbaijani, communicate with each other, express their thoughts and ideas orally or in writing, and pass them to the future through various sources. A significant part of the national languages are the official languages of individual states. The use of ICT in their own languages is a human right of individuals who are speakers of these languages, and it is important to ensure these rights. Moreover, since information technologies are included in most areas of our life where communication acts are realized, language existence, processing, protection, development, transmission to new generations, etc. occur precisely in the environment of technologies. Therefore, the future of national languages depends urgently on the level of their adaptation to the technological environment.

2. Related work

Considering the urgency of the mentioned issue, a Manifesto for Open Language Technology was adopted in 2019 signed by a group of scientists. The signatories of the manifesto highlighted common principles that all suppliers, manufacturers and distributors in the software and digital services industry should support. These are as follows:

Open localization: all software must be localizable regardless of the manufacturer.

Open interfaces: all programming interfaces associated with the language should be open by default.

Open resources: all language resources should be open and publicly available with the permission of the language community.

Accessible standards: international language standards must be respected, fully implemented, and applications updated regularly [openlt.org, 2019].

The signatories of the document expressed their belief that adherence to the above principles would lead to a better world for indigenous and minority languages and agreed on the importance of moving forward to ensure indigenous peoples' self-determination in the digital world. We believe that the creation of various technical solutions for the use of vernacular languages on computers and the Internet will strengthen the functionality of these languages in the ICT environment.

In recent decades, scientific research on this topic is also expanding. An important part of the conducted research covers ICT-related problems of indigenous languages of peoples living in African countries, Canada, and Australia. Dalvit L. (2010), Djite P. (2005), Dror I.E. (2008), Osborn D.Z. (2010), Begay W.R. (2013), Bell J. (2013), Bow J. (2021), Elliot R. (2021), Gallan C.K. (2016), Geneen I. (2018), Johnson G.M. (2016), Junker M.O. (2018), M. Koole and K. Lewis (2020), L.J. Brar and N. Roihan (2021), Patton J. (2018), Quintrileo E.J. (2019), Soria J. (2016) and others have relevant studies on this topic.

3. Materials and methods

In this research we use a descriptive method. The descriptive method in linguistics focuses on the description and analysis of language elements and language problems based on the analysis of certain facts. Extensive analysis of the topic based on this method helps to form a clear picture of the problem and find optimal solutions.

Currently, although the Azerbaijani language is represented at a certain level in the technological environment, this representation is not of decisive importance in communications. Such a situation is connected with certain limitations. Some of the limitations are related to the use of ICT opportunities in the Azerbaijani language. Another part of ICT itself is the restrictions that prevent the use of the possibilities of the Azerbaijani language, that prevent technologies from successfully crossing the borders of the local language environment and entering this environment. Therefore, we can approach the existing problems mainly in 2 aspects:

- *ICT problems of the Azerbaijani language;*
- *Azerbaijani language problems of ICT.*

Considering both areas in the context of the Azerbaijani language is somewhat conditional. Because the issues analyzed in this article, the difficulties to be solved, the proposals put forward for their elimination and the expected perspectives also apply to other Turkic languages and many national languages in the world in general.

4. ICT problems of the Azerbaijani language

Digital transformations are taking place in almost all spheres of social life outside the will of the peoples who speak their mother tongue, virtual realities are created by applying ICT. In this reality, new language manifestations are shaped, since the Azerbaijani language cannot keep up with these

rapid processes, ICT problems of the language appear. These problems are multifaceted.

An important part of the ICT problems of the Azerbaijani language is related to national domain names. As the Internet of Things, i.e., the idea of marking each material entity that surrounds us with its own domain name in the virtual space, gradually becomes a reality during the 4th Industrial Revolution, this problem becomes more noticeable.

The need to use the English (Latin) alphabet during the creation of domain names gradually makes spelling and grammatical mistakes in national languages the standard. Another problem is that when choosing and registering a domain name, taking into account the difficulty of using such letters as "ə", "ö", "ü", etc., registrars prefer not to use these letters. Thus, those letters are gradually discriminated and lose their position.

In order to eliminate such problems, the creation of domain names with the national alphabet of the Azerbaijani language should be started. Successful implementation of this work will help eradicate a number of problems and open up new opportunities.

Although Latin script is used in Azerbaijan, for the creation of national domains, taking into account the presence of vowels and consonants such as "ə", "ö", "ğ", "ş", "ç", "ı", "ü" adapted to the internal laws of our language in our alphabet, a number of technical issues need to be resolved.

One of the ICT problems of the Azerbaijani language is related to the issues of transliteration (replacing scripts of one writing system with the scripts or combinations of scripts of another writing system). With the exception of one or two languages, there are no transliteration interfaces between the national scripts of other nations and the scripts of the Azerbaijani language. Currently, there is no standard transliteration converter that determines which letters and combinations of letters correspond to the corresponding letters in the Latin scripts of the Azerbaijani scripts when they are converted to other languages or vice versa. In this regard, the creation of a national standard transliteration converter is also among the issues awaiting resolution. Such a problem is primarily manifested in domain names. Since domain names are given in English scripts, certain misunderstandings arise when Azerbaijani names are converted to this script in many cases. Thus, when Azerbaijani names are written in English, it is necessary to replace some letters. For example, it is possible to name sites about Baku, Karabakh, Shusha, Aghdam with different domain names: baku.az, baki.az, karabakh.az, shusha.az,

aghdam.az, agdam.az, agdam.az, etc. Substitution of letters can lead to chaos in the virtual world. In order to eliminate this confusion created in the virtual space around the world, there are standards related to transliteration. Specific Latin scripts "ə", "ö", "ü" etc. of the Azerbaijani alphabet lack in the Latin script used for English, however, there are no standard rules and requirements in Azerbaijan for how to transliterate these letters when creating domain names based on words containing these letters. Therefore, everyone must transliterate these words as they think convenient. For example, they adapt them as "ə-e", "ü-u", "ş-sh", "g-kh" and so on. This problem makes the creation of a transliteration system in Azerbaijan one of the most important issues to be solved. Thus, domain names made or to be made from hundreds of words in the spelling dictionary of the Azerbaijani language pose a great threat to our mother tongue in the virtual world. The phonetic spread in domain names also causes semantic inaccuracies (for example, www.əhsən.az - www.ehsan.az). For example, when someone wants to create a site related to any diseases, natural disasters, social problems, etc. and chooses the name as www.bəla.az to guide in advance on its content, it will not be possible to form the domain in this way. In this case, it is necessary to create the domain either www.bela.az or www.bala.az by transliterating the vowel "ə" with "e" or "a". And these, as seen, are appropriate for completely different contents" [Jafarov Y. et al, 2020].

Currently, since there are no special signs in the Latin scripts for the Azerbaijani orthography, which express the letters ç, ə, ğ, ı, ö, ş and ü, appropriate solutions are applied to avoid problems related to these letters when using information technologies. Although the global players of the computer and software market did not take into account the alphabets of many languages until 2000, a chaotic situation reigned in this area, but in 2000, after the alphabet symbols of all nations of the world were encoded by the Unicode standard, a single standard was formed. Azeri (Cyrillic) and Azeri (Latin) keyboard drivers have also been added to Windows to create text in the Azerbaijani language, which is based on Unicode. Thus, Unicode standards for the Azerbaijani language have taken their place in the information technology environment. Ubiquitous application of Unicode standards and the availability of tools supporting it are currently the most important areas of software development.

However, the introduction of the Unicode system is still not sufficient to fully solve the problems related to national alphabets, including

the Azerbaijani alphabet. Since some specific software does not support the Unicode system, there are certain difficulties related to the national alphabet when creating the appearance (interface) of programs in the Azerbaijani language.

It is precisely because of such problems that it is impossible to create e-mail addresses in the Azerbaijani language, as well as to send SMS messages through a mobile phone in full accordance with the orthographic norms of the language without distorting specific letters that are not in English. Therefore, state institutions, companies, and individuals are forced to violate the relevant spelling norms of our state language when creating e-mail addresses and sending SMS messages to citizens and customers.

In addition, due to the influence of user habits formed over many years, it is still very common to write ch instead of ç, e instead of ə, g instead of ğ, o instead of ö, sh or ch instead of ş, u instead of ü. Keyboards are currently adapted to the Azerbaijani alphabet by replacing the corresponding letters and signs on them with specific scripts in our alphabet. Some users do not use such keyboards, and if it is not necessary to write the text without spelling errors, they use the equivalents of the corresponding letters according to English grammar, for example, they write sh or ch instead of ş, thus violating the orthographic norms of the language. The widespread use of alphabet symbols in this way among the representatives of the young generation, who are the leading subjects of virtual communications and who benefit from technologies, threatens the future of the language.

It should be taken into account that in the digital age, it is important for every individual to have the ability and skill to write and send messages in native language to keep this language alive. Therefore, it is an important task to achieve the production of keyboards brought to the country based on the Azerbaijani scripts layout, to ensure the presence of the national alphabet layout in the keyboard system of mobile phones, and this issue should be kept under strict control.

The UN International Standards Organization has created three-letter codes for more than seven thousand languages. Web page and software developers use these codes, called ISO 639-3, to help computers and other devices learn to distinguish between different languages. Without an ISO code, it is impossible to translate operating systems and software interfaces, or even to take advantage of tools such as Google translate. The ISO code provides the relevant conditions for

receiving this support. Azerbaijani language also has ISO 639-3 code. Since issues related to ISO code, fonts, and keyboards have already been resolved to a certain level, the main tasks related to the use of the Azerbaijani language in ICT are improvement, implementation of control measures in this area, development of automatic programs that correct and edit words, sentences, and texts in the new strategic period, enriching relevant bases for maintaining accuracy and fluency when using machine translation programs, and creating speech-to-text and text-to-speech converting programs [Microsoft Corporation. 2004].

One of the ICT problems of the Azerbaijani language is that the digital resources related to various aspects of the language are not integrated into the e-government environment and online services are not provided in a "single window" mode accessible to all.

The main prospective project in this field is the construction of the architectural-technological foundations of the system of national languages on the cosmopolitan e-government platform. Due to such a socio-technological architecture, the transformation of national languages into a part of the e-government platform and their place in this technical platform will greatly contribute to the preservation and development of languages in the globalization era [Alguliyev R.M., Aliguliyev R.M., Jafarov Y.M., Yusifov F.F., Gurbanova A.M. 2021].

New methods, algorithms, approaches should be developed, and a number of applications should be established in order to make national languages a part of the e-government and to ensure the natural development of this platform on all segments. This includes solving the problems of localization and globalization, documenting languages, assessing the language situation in the country by analyzing texts, intellectual analysis of language use, feedback, spelling, orthoepy, transliteration, terminology, dialectology issues, names and surnames, vulgarisms, building language industry service systems, Creation of the National Corpus of the Language, etc.

The lack of national, corporate and individual corpora of the Azerbaijani language is one of the main ICT problems of our language. The presence of the national corpus of languages in the virtual environment is of great importance. When we say the national corpus of a language, we mean an information search system aimed at collecting relevant texts in electronic form in any specific language. In the national corpus, the language is

fully recorded at a certain historical stage, that is represented by all its styles, genres, literary language, dialects, etc. [Mahmudov, M.A., 2002]

The corpus characterizes the language with all its genre and stylistic variations, dialects and professional lexicon at a certain stage or at all stages of its existence.

The national language corpus is a set of natural language texts stored in an electronic medium, arranged in a certain order, providing material for linguistic research on various linguistic phenomena and aspects. The texts that make up the corpus and the reference to them are made according to certain rules. The first multifunctional databases and dictionaries were created in 1956 in the United States. Many dictionaries have been developed based on electronic databases, such as Webster's English Dictionary in 1996. In the 1960s and 1970s, the first corpora appeared in connection with the development of computer technology (Kennedy 1998). The number of words there was limited. The famous corpora known at that time are The Brown Corpus, Lancaster-Oslo/Bergen Corpus, London-Lund Corpus.

At present, many countries of the world have compiled national corpora for the world's leading languages. Of course, these corpora are not the same according to their scientific and technical indicators. The British National Corpus is considered the world's most respected and exemplary corpus. In the family of Slavic languages, national corpora of Russian and Czech languages are specially distinguished.

The current experience in the field of creating national language corpora stipulates the acceleration of the preparation of the national corpus of the Azerbaijani language, as well as the realization of joint projects between the Turkic countries for the formation of the national corpus of Turkic languages.

The "State program on the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country" envisages the creation of a common information base of the Azerbaijani language as the state language, the preparation of electronic frequency, terminological dictionaries, the strengthening of activities for the creation of Internet resources, electronic and interactive textbooks in the Azerbaijani language, the creation of an electronic portal of manuscripts, the creation and development of internet technology, machine translation systems and other modern applied-

linguistic technologies to ensure wider use of the Azerbaijani language [www.president.az, 2013].

The Decree of the President of the Republic of Azerbaijan dated July 17, 2018, "On a number of measures related to ensuring the wider use of the Azerbaijani language in the electronic space" embraces the implementation of the project "Electronic platform of the Azerbaijani language for foreigners" in order to facilitate the wider use of the Azerbaijani language in the electronic space, its accessibility for people interested in this language and to facilitate their learning [www.president.az, 2018]. All this shapes the legal basis for the creation of the national corpus of the Azerbaijani language.

In order to generate the theoretical and practical base of the national corpus of the Azerbaijani language, formal analysis systems, formal morphological, syntactic, semantic analysis and synthesis programs should be developed. Formal analysis and synthesis systems are considered the most important components of national language corpora. Therefore, it would be more appropriate to start the preparation of the corpus from them. It is possible to build other components of the system only on the basis of a perfect and thorough formal analysis. The problem of eliminating homonymy, especially at the morphological, syntactic and semantic levels, is considered one of the main obstacles in the creation of national language corpora.

The texts included in the corpus should contain all styles of the Azerbaijani language at the same level and volume. According to the principles of creating a national corpus of languages, the larger the volume, the more reliable and comprehensive the corpus is. Modern information technologies allow to increase the volume of national language corpora in any framework. Unquestionably, it would be more useful to use optimal positioning methods here. The volume may be overwhelming, but the language is underrepresented. For example, the entire corpus can be filled with newspaper and magazine material, while other styles and areas of the language can be left out of the corpus. Therefore, the correct planning of the structure of the corpus is considered very important, and it is considered appropriate to include the following texts in the national corpus of the Azerbaijani language. Fig. 1 illustrates the structure of the National Corpus of the language.

In addition to what is shown here, the corpus can be entered new texts and materials from different fields.



Fig. 1. Structure of the National Corpus of the Language

As mentioned in the existing studies, any text written in Azerbaijani and related to the Azerbaijani language can be added to the corpus if it represents the language.

The national corpus is intended as an open system. In this regard, it is possible to make changes in its structure and functions as necessary. Not all sub-corps of the national corpus work at the same time. As sub-corpses are formed and their support bases are created, their activities begin. At the first stage, several sub-corpses of the corpus become functional, their interrelationships and demand act as a stimulus for the creation of new sub-corpses [Mahmudov M., 2013].

Obviously, electronic library service is currently operating in our republic. The electronic version of the literature published in the Azerbaijan National Library, the library of the Azerbaijan National Academy of Sciences, the libraries of higher schools, many field libraries, scientific-research institutions, and higher schools is created and uploaded on the Internet. Besides, there are websites of almost all newspapers and magazines published in our country. Prepared texts are placed on these pages in electronic format. Moreover, our country has thousands of electronic information portals, electronic information resources of various central and regional state, private, non-governmental organizations, commercial entities, individual citizens. All television programs, films, music samples, etc. broadcast on the airspace of Azerbaijan are posted on special internet platforms in audio-video format and can be viewed. It is possible to use all of these as a mass material in the national corpus of our language. However, the mentioned ones are uploaded on the Internet in their full version without

preliminary processing, without polishing in accordance with the requirements of the national corpus. This makes it difficult to use them for various pragmatic purposes. National corpus materials should be entered into the system not directly, but after a certain processing stage (linguistic processing, marking, etc.). Therefore, it is necessary to create specialist groups, allocate necessary funds and start large-scale projects.

Within the framework of solving the ICT problems of the Azerbaijani language, there is a need to digitize a large number of resources (dictionaries, atlases, etc.) related to our language, and create websites related to the language. For many years, various encyclopedic dictionaries and separate types of philological dictionaries, i.e., explanatory, terminological, bilingual, orthographic, orthoepic, dialectological, etymological, historical-ethnographic dictionaries have been published in the Azerbaijani language. Furthermore, linguistic atlases have also been published that preserve vocabulary of the people alive in the regions and represent the facts of the language and dialect not only from the lexical point of view, but also from the phonetic-grammatical point of view. Obviously, linguistic atlases are a set of systematic maps showing the linguistic landscape of separate regions and the distribution area of certain language phenomena. They appear at a certain stage of the development of dialectology, when a lot of materials are collected shaping a general idea about the dialectal features of the language. The preparation of the first dialectological atlases among Turkic-speaking peoples started in Azerbaijan, and rich experience has been accumulated in this field. Currently, there are initiatives to integrate some of the existing dictionaries into the electronic environment for both scientific and commercial purposes. Platforms such

as azerdict.com, obastan.com are some of them. The intensity of applications to these platforms shows that there is great interest in them. However, a very limited part of the documentary resources formed in the Azerbaijani language is represented on these platforms. There is a need to expand such platforms with new resources, making them a part of e-government, digitizing linguistic atlases, toponyms, family names, names, surnames, etc. and to create new digital linguistic atlases and conduct human geography analyzes based on them.

“Human geography is a scientific field that studies societies, their characteristic features, their activities, the social way of thinking specific to this society in relation to the geographic environment and national spatial data they are connected to” [Jafarov Y., 2016].

When determining the relationship between the historical-geographic features of the territory and the population structure, revealing the natural and human factors that affect the distribution of settlements by region, and following these effects in the modern era, language facts enable the detection of serious cryptographic information. The creation of such digital databases, on the one hand, will help the development of the language, and on the other hand, it will create an opportunity to conduct analyzes with various pragmatic purposes.

Communication plays a vital role in the development of language as well as in its gradual decline. Today, the virtualization of communication processes between people has created new trends in language. An important part of these trends appears in social network communication, which has become an integral part of our modern life. The most characteristic features of this, as in oral speech, are the deviation from language norms in the social network environment, the tendency to concise expression of ideas, the substitution of symbols, images for words, sentences and texts, the creation of a number of new words, the acquisition of new meanings of existing words, etc.

All these are serious problems that have a negative impact on the development of the literary language and pose threats to the Azerbaijani language. Therefore, there is a need to implement some measures to increase the level of language literacy of citizens. Development of a special functional module related to language sanitation in e-government can help to solve this problem. This module applies electronic editing, proofreading, text creation programs, continuous

monitoring of deviations from language norms in electronic mass media, i.e., televisions, radio.

E-government platform also creates certain opportunities for adjusting the orthographic and orthoepic norms of the language. If a special service related to the spelling and orthography of the language is formed on this platform, it will solve numerous existing problems through various integration solutions, the formation and regular updating of databases, the identification and elimination of characteristic distortions, the organization of public discussions about individual words and their spelling, the rules for writing neologisms, providing users with online consulting services about spelling and orthography rules, and improving online dictionaries.

Like other languages of the world, the literary language of Azerbaijan is based on the International Phonetic Alphabet. Currently, more than half of the world's languages use this system and have created their own orthoepic dictionaries. In order for correct pronunciation norms to gain functionality in Azerbaijan, first of all, a single source reflecting exact normative rules, i.e., an orthoepic dictionary should be created and integrated into the e-government platform.

Special attention should be paid to the development of students' speaking ability in teaching processes and exam strategies, this factor should be taken into account in the evaluation of their educational achievements, and to this end, the opportunities of the relevant service on the e-government platform should be used. It is very important to provide special subsystems in the e-government platform for people with physical disabilities. To this end, surdo-interpreter (sign language) programs, voice-to-text, text-to-voice applications, etc. need to be integrated into the system. All this will contribute to social inclusion, sustainable development, integration of relevant social groups into society.

5. Azerbaijani language problems of ICT

As we mentioned, there are a number of serious problems related to the use of the opportunities of the Azerbaijani language in ICT. These problems do not allow ICT solutions to benefit from the richness of our language and fully enter the national language environment. We conditionally call these problems Azerbaijani language problems of ICT.

One of such problems is serious limitations in the application of natural language processing and text analysis methods and tools in the Azerbaijani language. Natural language processing has now become one of the main development realities in most countries of the world. This scientific field, now more commonly known by the abbreviation NLP (Natural Language Processing), is a subcategory at the intersection of computer science, artificial intelligence and linguistics.

NLP is the analysis of the structure and fixed rules of natural languages for technical purposes and their reprocessing according to these purposes. The solutions created on the basis of these analyzes enable automatic translation of texts, development of question-and-answer platforms, automatic speech and comprehension of given commands, synthesis of conversations, extracting a brief overview of the text, data protection, processing, etc. Due to the widespread use of information technologies and social networks, written communication is applied more widely today than ever before. It integrates tools such as auto-correction, editing, and even voice-to-text applications into written communication platforms. All these help to analyze the written text and check language rules like natural language processing programs. In particular, speech and command understanding software will eliminate input devices such as keyboards and mice for human-computer communication in the future. Such a development will lead to new and revolutionary changes in machine-human communication, enabling computers to be used by more people [Kurdi, M.Z., 2017].

NLP is mainly applied in the following areas: Sentiment analysis; Chatbot; Speech recognition; Machine translation; Pronunciation check; Searching for suitable keywords for the CEO; Data extraction; Advertisement compliance, etc.

It should be noted that researches related to text analysis have been started since the late 50s of the last century. In connection with technological progress, development trends in this field have expanded since the beginning of the 80s, and since the beginning of the 21st century, development processes have accelerated. Text analysis is an intersection of various sciences and covers several fields as information retrieval, data analysis, machine learning, statistics, applied linguistics, etc. Currently, approximately 80 percent of all information available in various resources is text information. From this point of view, text analysis has a very important socio-economic value, and

this value will increase gradually.

Text analysis includes classification and clustering of texts, creation of word groups for texts, sentiment analysis of texts, analysis of ideas in texts, formation of text summaries, internal modeling of the text itself, etc. Methods such as information retrieval, word frequency distribution, pattern recognition, marking, information extraction and even visualization are used here [Nabiyev, V.V., 2013].

At present, it is impossible to apply these technologies, which are widely used in many languages of the world, in the Azerbaijani language and to benefit from these technologies for speakers of the national language. The main reasons for this are that the ontology of the Azerbaijani language, the necessary number of thesaurus dictionaries and frequency dictionaries of various purposes have not been prepared in our mother tongue. Linguo-statistics, which is currently widely used in world linguistics, determines which units are found frequently or very rarely in speech and text. Analysis of the functionality of language units ensures the discovery of certain regularities in the processes of formation of speech expressions. Frequency dictionaries created on the basis of this regularity are also a list of selected words, word-forms or word combinations in the studied text. In the dictionary, the numbers indicating the frequency of use of the units in the texts are recorded. A frequency dictionary compares words by frequency of occurrence. In this regard, the scope of application and use of frequency dictionaries is very wide. Frequency dictionaries show a number of features determined by the ratio of any text to vocabulary and individual words in the text with precise measurements [Genee, I., 2018]

Thesauruses are the product of many years of research in the field of ideographic lexicography. They include the main features of ideographic dictionaries. Thesauruses systematize the entire lexical fund of the language from a paradigmatic point of view. Determining the semantic relations of words plays a very important role in these dictionaries. Thesaurus-type semantic dictionaries contain information that includes the semantic relationships of words. Thesauruses differ from each other according to the character and field of the words that make up it (professional lexicon), the depth and direction of the hierarchy, the type and quantity of used paradigmatic relations, as well as the formulation methods. Thesauruses are open-ended and continuously enriched by

additional paradigmatic relations, taking into account practical necessity.

The number of such dictionaries in our language is very limited and almost at the level of early initiatives. In addition, sentimental analysis and vulgar words database has not been created in the Azerbaijani language. There is no national corpus of the language. Necessary tools such as stemming, which consists of creating morphological variants of the word root, tokenization, which involves dividing the text into different words, and lemmatization, which combines different flecional forms of the same word, have not been developed. All this plays a

very important role in NLP, including the language of chatbots. For the mentioned reasons, ICT cannot penetrate widely into the sphere of use of the Azerbaijani language.

If the indicated problems are solved, then it is necessary to summarize and classify texts, speeches, audiovisual content in the Azerbaijani language, to recognize and analyze human speech in our language using a computer, to synthesize artificial speech from Azerbaijani texts on the computer, to develop human-computer (robot, drone, etc.) dialogue systems (chatbot, etc.). Table 1 presents all directions for solving these problems.

Table 1. Problem solving directions

▪ Determining the emotional tonality of texts and speeches (opinion mining, sentiment analysis, etc.)
▪ Creating quality computer translation programs between the Azerbaijani language and other languages of the world
▪ Developing the technologies to ensure automatic detection of new terms entering the Azerbaijani language through online channels
▪ Enabling uninterrupted monitoring of the Azerbaijani language in the virtual environment.

Table 2. The result of the implementation of successful projects in the specified directions

▪ Objective and effective decisions will be made regarding language development.
▪ A national information search engine will be created in the Internet environment.
▪ Information security in the Azerbaijani language virtual space, including protection of children from harmful text and audio-video content will be provided.
▪ Cybercrime will be effectively combated.
▪ Contribution will be made to the formation of the national e-government in Azerbaijan, and the provision of various online services related to the Azerbaijani language to citizens. Great support will be given to the preparation of national anti-plagiarism programs for Azerbaijani-language texts.

Table 2 is compiled as a result of the implementation of successful projects in the indicated direction.

In short, by solving numerous problems related to the Azerbaijani language in the era of digital transformations, it is possible to ensure the preservation, development and wider application of the Azerbaijani language in the online environment.

6. Discussion

At present, although separate organizations and individuals develop separate language-related software products for various purposes, including commercial ones, these products cannot be widely used for objective reasons. In order for these products to prove themselves, it is necessary to integrate them into e-government systems and to form an ecosystem of the Azerbaijani language

based on client/server technologies as a whole. In this case, free unlimited services related to the Azerbaijani language, including spelling, terminology, dialectology, translation, etc., will be available to anyone who needs them in the virtual space.

- Taking into account that these technological innovations will make important contributions to the language policy of the state, i.e., it will ensure its use, protection and development as an official state language, as well as the establishment of translation and transliteration with other languages, it is appropriate for the state to provide political, legal, infrastructure and financial support to these projects.

- It is very important to establish e-government and its functional components according to the requirements of the Azerbaijani language. The functional components of the Azerbaijani language on the digital platform should be created to interact with each other in the e-government

environment, as well as provide citizens with uninterrupted online language services (terminological, orthographic, vulgarism blocking, translation, data, etc.).

- It is extremely necessary to use international standards for teaching and evaluating the Azerbaijani language in our country. Thus, citizens' language skills should be assessed based on national standards in terms of writing, reading, listening and speaking, and the results of the assessment should be represented in their personal resumes (CV).

- In order to support the language policy of the state, a system of linguistic demography should be established as a part of the e-government environment, sociolinguistic analyzes should be conducted on various aspects, and necessary proposals should be formulated.

- During the implementation of this concept, the need to protect not only the state language of Azerbaijan, but also the languages of different ethnic groups living in our country and having their own languages should be taken into account, their languages should be documented, relevant online language ecosystems should be created for ethnic minorities within the framework of the multicultural policy of the Azerbaijan government, these ecosystems should have an open architecture, should be constantly improved and developed in accordance with technological challenges.

7. Conclusion

Each of the prospective projects related to all mentioned issues should be performed in multifunctional teams: because linguists have limited professional knowledge and skills in the technological field, and ICT specialists' linguistic knowledge is not at a professional level. In order to overcome these limitations, linguists, ICT specialists, product owners, project managers, etc. should establish multi-functional teams. In addition, computer linguistics specialty, department, etc. should be organized in higher and secondary education institutions. Personnel in this specialty are being trained in higher education institutions of most advanced countries of the world. In order to eliminate the shortage of personnel in this field in our country, young people should be assigned to universities and research centers in foreign countries to acquire professional knowledge and skills in the fields of computer linguistics and language engineering.

There is a need for modern challenges in computer linguistics in the Azerbaijani language, the preparation of textbooks in accordance with international practice, and the translation of high-quality textbooks and teaching aids available in foreign languages on this topic into our language. In general, "ICT problems of the Azerbaijani language, the Azerbaijani language problems of ICT" should be recognized as a scientific priority, and systematic research should be conducted in this field. Based on the international experience, the system of teaching and evaluating the Azerbaijani language should be rebuilt with the application of ICT.

In order to support ongoing efforts in all these fields and to develop coordination and cooperation between them, steps should be taken, various grants, tax benefits, technical equipment, training, etc. should be provided for those who contribute to this work. Moreover, relations between linguists and ICT specialists working on various projects in Turkic-speaking countries should be strengthened, they should be involved in joint projects. Forums and other such large-scale events related to the ICT problem of Turkic languages, Turkic language problems of ICT should be organized, and cooperation with small projects on information exchange in the field of language should be encouraged.

Solving numerous problems related to the Azerbaijani language in the era of digital transformations will create real opportunities to ensure the protection, development and wide application of our language in the online environment.

References

- A Manifesto for Open Language Technology. 2024 <https://openlt.org/>
- Alguliyev R.M., Aliguliyev R.M., Jafarov Y.M., Yusifov F.F., Gurbanova A.M. (2021). From a cosmopolitan e-government to a national e-government: prospects for creating Azerbaijani language services in a virtual environment // Information society problems, , 12(2), 3–25 (in Azerbaijani)
- Begay, W.R. (2013). Mobile apps and indigenous language learning: New developments in the field of indigenous language revitalization [Masters Thesis, University of Arizona]. <https://www.semanticscholar.org/paper/Mobile-apps-andindigenous-language-learning%3A-NewBegay/2d2975a469b614e4e1ba4eed6c1e1e40b4eda60f>
- Bell, J. (2013). Language attitudes and language revival/survival. *Journal of Multilingual and Multicultural Development*, 34(4), 399–410. <https://doi.org/10.1080/01434632.2013.794812>
- Bow, C. (2021). Entanglements of digital technologies and Indigenous language work in the Northern Territory

- [Doctoral, Charles Darwin University and the Australian National University]. <http://hdl.handle.net/1885/233083>
- Dalvit L (2010). Multilingualism and ICT Education at Rhodes University: An Exploratory Study. Unpublished Ph.D. thesis, Rhodes University.
- Decree of the President of the Republic of Azerbaijan on a number of measures related to ensuring wider use of the Azerbaijani language in an electronic space. 2018 <https://president.az/articles/29492>
- Djite P (2005). From Liturgy to Technology: Modernising the Languages of Africa. *Language Problems and Language Planning*, 32(2), 133-152.
- Dror IE (2008). Technology enhanced learning: The good, the bad and the ugly. *Pragmatis and Cognition*: 16(2), 215-223.
- Elliot, R. (2021). Technology in Language Revitalization. In J. Sallabank & J. Olko (Eds.), *Revitalizing Endangered Languages: A Practical Guide* (pp. 297–311). Cambridge University Press. <https://doi.org/10.1017/9781108641142.008>
- Genee, I. (2018). The Blackfoot Language Resources and Digital Dictionary project: Creating integrated web resources for language documentation and revitalization. *Language Documentation & Conservation*, 12, 274–314.
- Heritage Languages and the Well-being of Speakers, 2023. <https://www.linguapax.org/en/linguapax-review-en/>
<https://www.azstat.org/MetaDataG/>
- Jafarov, Y.M. (2016). Issues of geonomastics and human geography // Materials of the Fourth International Scientific Conference of Young Researchers (Qafqaz University) dedicated to the 93rd anniversary of the national leader Heydar Aliyev, (Baku), p. 158 (in Azerbaijani)
- Jafarov, Y.M., Gasimova R.T. (2020). Some problems in the design of domain names with the Azerbaijani alphabet on the Internet. *Journal of Information Society Problems*, 11(1), 64-74 (in Azerbaijani)
- Johnson, G. M. (2016). Technology use among Indigenous adolescents in remote regions of Australia. *International Journal of Adolescence and Youth*, 21(2), 218–231. <https://doi.org/10.1080/02673843.2013.823553>
- Junker, M.O. (2018). Participatory action research for Indigenous linguistics in the digital age. In *Insights from practices in community-based research: from theory to practice around the globe*, 319. De Gruyter Mouton. <https://doi.org/10.1515/9783110527018-009>
- Koole, M., & Lewis, K. (2020). Mobile learning as a tool for indigenous language revitalization and sustainability in canada: framing the challenge (pp. 520–532). <https://doi.org/10.4018/978-1-7998-0423-9.ch028>
- Kurdi, M.Z. (2017). *Natural Language Processing and Computational Linguistics 2: semantics, discourse, and applications* / M.Z.Kurdi. ISTE-Wiley, 318 p.
- Quintrileo, E.J. (2019). Technology's role in Mapudungun language teaching and revitalization. *AlterNative: An International Journal of Indigenous Peoples*, 15(3), 205–216
- Li, J., Brar, A., & Roihan, N. (2021). The use of digital technology to enhance language and literacy skills for Indigenous people: A systematic literature review. *Computers and Education Open*, 2, 100035. <https://doi.org/10.1016/j.caeo.2021.100035>
- Mahmudov M.A. (2002). System of formal text analysis. Baku, "Elm", pp. 39-40 (in Azerbaijani).
- Mahmudov, M.A. (2013). *Computer linguistics* / M.A. Mahmudov. - Baku: Science and Education, (in Azerbaijani)
- Microsoft Corporation. 2004. "Microsoft Enables Millions More to Experience Personal Computing Through Local Language Program." *PressPass - Information for Journalists*, March 16. <http://www.microsoft.com/presspass/press/2004/mar04/03-16LLPPR.asp>
- Nabiyev, V.V. (2013). *Yapay Zekâ: Problemler, Yöntemler, Algoritma* / V.V.Nabiyev. - Seçkin Yayıncılık, 792 s.
- Nabiyev, V.V. *Artificial Intelligence: Problems, Methods, Algorithm* / V.V.Nabiyev. - Seçkin Publishing, 792 (in Turkish)
- Osborn D.Z. (2010). *African Languages in a Digital Age: Challenges and Opportunities for Indigenous Language Computing*. Cape Town: HSRC Press.
- Patton, J. (2018). New app hopes to revitalize, teach Indigenous languages. *Global News*. Retrieved from <https://globalnews.ca/news/4288160/indigenous-languageappwikwemikong-first-nation/>
- Soria, C. (2016). What is digital language diversity and why should we care? In J. Cru (Ed.), *Linguapax Review 2016—Digital Media and Language Revitalisation*. Linguapax International.
- State Program 2013 on the use of the Azerbaijani language in accordance with the requirements of the time and the development of linguistics in the country in terms of globalization <https://www.president.az/articles/7744>