

REPUBLIC OF AZERBAIJAN

On the rights of the manuscript

ABSTRACT

of the dissertation for the degree of Doctor of Philosophy

**OPTIMAL PLACEMENT OF LANGUAGE MATERIAL AND
USING POSSIBILITIES IN ELECTRONIC SPACE**

Specialty: 5704.01 – “Language theory”

Science: Philological sciences

Applicant: **Maya Intigham Heydarova**

Baku – 2026

The work was performed at the department of AI and Computer linguistics of the Linguistics Institute named after Nasimi of ANAS.

Scientific adviser: Doctor of Philological Sciences, Professor
Masud Ahmad Mahmudov

Official opponents: Doctor of Philological Sciences, Professor
Ilham Mikayil Tahirov

Doctor of Philosophy in Philology, Associate Professor
Sevinj Salim Mammadzadeh

Doctor of Philosophy in Philology
Jala Eldar Mammadova

Dissertation council FD 1.06 of the Supreme Attestation Commission under the President of the Republic of Azerbaijan operating at the Institute of Linguistics named after Nasimi, Azerbaijan National Academy of Sciences.

Chairman of the Dissertation council:

 Doctor of Philological Sciences, Professor
Nadir Balaoglan Mammadli

Scientific secretary of the Dissertation council:

Doctor of Philosophy in Philology, Associate Professor
Sevinj Yusif Mammadova

Chairman of the scientific seminar:

 Doctor of Philological Sciences, Professor
Gulsum Israfil Huseynova

INTRODUCTION

Relevance and development of the topic. In the contemporary globalised world, characterised by the comprehensive development of information and communication technologies and the Internet, the emergence and development of economic, political, scientific and cultural relations between people has entered a new stage in terms of both quality and quantity. The global development in the field of information technologies is exerting a positive influence on linguistics, as is the case in all fields of science. The advent of machine translation, electronic dictionaries, national language corpora, natural language processing systems, the creation of formal grammars and speech recognition systems, algorithms for analysing and synthesising natural text, and other such issues has given rise to a new scientific direction, namely computational linguistics. It is important to note that, *“Computational linguistics is a very broad concept and includes different areas. It includes problems of artificial intelligence, modeling of communication and acquisition, modeling of plot structure, presentation of texts using hypertext technology, machine translation, computer lexicography, creation of national machine funds and national corpora for individual languages, etc.”*¹

At the current stage, computational linguistics is more manifested in the process of applying artificial intelligence systems to language. Presently, various types of information resources, language corpora, electronic dictionaries, large language models, and text analysis programs have been created in the world. The possibilities of using such electronic resources are improving and updating day by day. Along with all this, one can also encounter negative situations in this direction. In particular, the widespread use of scientific articles, monographs, and dissertations with the application of artificial intelligence also makes it necessary to take

¹ Mahmudov, M.Ə. Kompüter dilçiliyi / M.Ə.Mahmudov. – Bakı: Elm və təhsil, – 2013, – s.27

measures against such programs. In this sense, in terms of applying innovations in world practice to Azerbaijani linguistics, writing such dissertations is distinguished by its relevance.

In contrast, in the contemporary era, language material, which functions as a carrier and transmitter of various forms of information, is obtained through the placement of texts with various contents and purposes in cyberspace.

In the modern information society, there are no prohibitions or restrictions in terms of the collection, processing and transmission of information. Conversely, the optimal placement of language material in the Internet space, as the most significant and essential means of communication in human society, and the beneficial use of it by users in accordance with their interests, should also be scientifically analysed. It is imperative to conduct research and provide recommendations regarding the development and preparation of certain regulatory tools and rules. The aforementioned elements can be characterised as aspects that determine the relevance of the topic.

In recent years, a significant volume of theoretical and practical work has been undertaken in Azerbaijan in this field. However, given the daily creation of new software and the development of various databases, it is evident that there is still a considerable amount of work to be done in this field within our country. The orders issued by the President of the Azerbaijan Republic, Mr. Ilham Aliyev, once again demonstrate the significance of this field ².

The placement of language material, including materials related to the Azerbaijani language, in electronic space and its status have been studied by many researchers from various aspects. Examples of such research include the works “Computer linguistics” ³ and “National corpora of Turkic languages” ⁴ by Azerbaijani scientists M.Mahmudov, “National corpus and lexical infrastructure: the

² Heydərova, M.İ. Azərbaycan dili elektron məkanda // – Bakı: Dilçilik İnstitutunun əsərləri, – 2018. № 2, – s.187-200.

³ Mahmudov, M.Ə. Kompüter dilçiliyi / M.Ə.Mahmudov. – Bakı: Elm və təhsil, – 2013. – 356 s.

⁴ Mahmudov, M.Ə. Türk dillərinin milli korpusları / M.Ə.Mahmudov. – Bakı: Elm və təhsil, – 2018. – 392 s.

experience of building the azerbaijani language lexical database “Introduction to computer linguistics”⁵ by N.Mammadli, M.Mahmudov, I.Tahirov by K.Valiyeva, “Theoretical problems of machine translation”⁶ by M.Mirzaliyeva and K.Valiyeva, “Ways to solve computer linguistics problems”⁷ by K.Valiyeva, by R.Mahmudova, “Semantic analysis in machine translation”⁸ by J.Jafarova, “Principles of automatic dictionary compilation in educational computer programs”⁹ submitted by A.Alizadeh for the degree of PhD in Philology, “National language corpora and principles of their formation”¹⁰ by S.Mammadzadeh, “Principles of word presentation and explanation in electronic dictionaries”¹¹ by J.Mammadova, as well as among foreign researchers, the works of A.Antopolsky’s “Linguistic information resources”¹², T.Lando, O.Mitrenina, I.Nikolaev’s “Applied and computer linguistics”¹³, Y.Marchuk’s “Computer linguistics”¹⁴, I.Bolshakov’s and A.Gelbukh’s “Computational linguistics: models, resources,

⁵ Məmmədli, N. Milli korpus və leksik infrastruktur: Azərbaycan dilinin söz bazasının yaradılması təcrübəsi. / N. Məmmədli, M. Mahmudov, İ. Tahirov. – // Bakı: Elm və təhsil, Dilçilik araşdırmaları, – 2025. № 2. – s.3-15.

⁶ Mirzəliyeva, M., Vəliyeva, K. Maşın tərcüməsinin nəzəri problemləri / M.Mirzəliyeva, K.Vəliyeva. – Bakı: Elm və təhsil, – 2016. – 120 s.

⁷ Vəliyeva, K.A. Kompüter dilçiliyi məsələlərinin həlli yolları / K.A.Vəliyeva – Bakı: Zərdabi Nəşr MMC, – 2022. – 104 s.

⁸ Cəfərova, G.N. Maşın tərcüməsində semantik təhlil / G.N.Cəfərova. – Bakı: Elm və təhsil, – 2023, – 232 s.

⁹ Ализаде, А.А. Принципы составления автоматического словаря в компьютерных программах обучения: / дисс. на соискание ученой степени доктора философии по филологии / – Баку, 2023. – 175 с.

¹⁰ Məmmədzadə, S.S. Milli dil korpusları və onların formalaşması prinsipləri: / filologiya üzrə fəlsəfə doktoru dis. / – Bakı, 2018. – 149 s.

¹¹ Məmmədova, J.E. Elektron lüğətlərdə sözlərin təqdimi və izahı prinsipləri / filologiya üzrə fəlsəfə doktoru dis. / – Bakı, 2024. – 149 s.

¹² Антопольский, А.Б. Лингвистические информационные ресурсы: монография / А.Б Антопольский. – Москва: ИНИОН РАН, – 2022. – 464 с.

¹³ Ландо, Т.М., Митренина, О.В., Николаев, И.С. Прикладная и компьютерная лингвистика / Т.Ландо, О.Митренина, И.Николаев. – Ленанд, – 2016. – 320 с.

¹⁴ Марчук, Ю.Н. Компьютерная лингвистика / Ю.Марчук. – Москва: АСТ: Восток – Запад, – 2007. – 317 с.

applications”¹⁵, J.Nerbonne’s “Linguistic databases”¹⁶, V.Zakharov’s and S.Bogdanova’s “Corpus linguistics”¹⁷, A.Stefanovich’s “Corpus linguistics: A guide to methodology”¹⁸, D.Crystal’s “Language and the Internet”¹⁹ and others can be cited.

Object and subject of research. The object of the study is language materials placed in electronic space or on the Internet in various forms (electronic dictionaries, text analysis programs, linguistic databases, national language corpora, etc.). The subject of the study is the investigation of the optimal placement and use of these language materials in electronic space.

Aim and tasks of the study. The aim of the study is to investigate, analyse and develop appropriate recommendations in this direction regarding the optimal placement and use of language materials on the Internet. For the purposes outlined herein, the following tasks are to be given primary consideration:

- to determine the functions of electronic space in the communication process;
- to analyze the forms of communication in electronic space;
- to classify language materials placed in various forms on the Internet;
- to analyze the importance and usefulness of language materials placed on the Internet;
- to determine the placement principles of various language materials in the Internet space;

¹⁵ Bolshakov, I., Gelbukh, A. Computational linguistics: Models, Resources, Applications / I.Bolshakov, A.Gelbukh. – Mexico: IPN-UNAM-FCE, – 2004. – 186 p.

¹⁶ Nerbonne, J. Linguistic Databases / J. Nerbonne. – Cambridge: Cambridge University Press, – 1998. – 243 p.

¹⁷ Захаров, В.П., Богданова, С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов / В.П.Захаров, С.Ю.Богданова. – Иркутск: ИГЛУ, – 2011. – 161 с.

¹⁸ Stefanowitsch, A. Corpus linguistics: A guide to the methodology / A. Stefanowitsch. – Berlin: Language Science Press. – 2020. – 510 p.

¹⁹ Crystal, D. Language and the Internet / D.Crystal. – Cambridge: Cambridge University Press, – 2006. – 318 p.

- to classify electronic dictionaries and analyze the tradition of their placement in the Internet space;
- to classify linguistic databases and determine their role in research;
- to harmonize the characteristics of the placement of language corpora prepared in Azerbaijan with the experience of foreign countries;
- to determine the functions of electronic space in the communication process;
- to analyze the forms of communication in electronic space;
- to classify language materials placed in various forms on the Internet;
- to analyze the importance and usefulness of language materials placed on the Internet;
- to determine the principles of placing various language materials on the Internet;
- to classify electronic dictionaries and analyze the tradition of their placement on the Internet;
- to classify linguistic databases and determine their role in research;
- to harmonize the characteristics of the placement of language corpora prepared in Azerbaijan with the experience of foreign countries;
- to clarify the possibilities of representing the Azerbaijani language on the Internet and to classify language materials related to the Azerbaijani language;
- to examine the current state of electronic dictionaries and educational materials intended for the Azerbaijani language and to determine their features;
- to prepare recommendations on the placement of the national corpus of the Azerbaijani language on the Internet.

Research methods. The descriptive, statistical, comparative, and system-structural methods were used in the research.

The main provisions of the defense: The provisions put forward for protection include the following:

– when the internet, electronic communication and exchange are widespread, the huge amount of information materials, which include hundreds of millions of written texts, audio clips, images and other illustrative tools collected in electronic space every day, must be regulated and placed according to the user's requirements.

– scientific analysis of the statistical parameters of the language material to be placed on the Internet in terms of quantity and quality should be taken as the basis for the optimal placement process;

– increased attention should be paid to the creation and posting on the Internet of electronic versions of dictionaries, textbooks, and other popular printed products widely used by the scientific community, students, teachers, and other individuals. Electronic dictionaries and libraries should be developed using the latest technologies and constantly updated;

– language materials posted on the Internet - linguistic databases collect information about the world's languages in one database, which facilitates the conduct of linguistic research;

– one of the main tools for placing and using language materials on the Internet is linguistic corpora or national language corpora. Marking up texts in corpora is of great importance;

– the development of a phonetic database is considered an important source for research in other fields, in addition to linguistics;

– localization of computer programs should be carried out taking into account the national characteristics of peoples;

– to overcome the difficulties of placing language material on the Internet, it is necessary to build an artificial model of natural language;

– in recent years, a lot of work has been done in Azerbaijan on the development of electronic dictionaries. At the same time, the issues of placing electronic dictionaries, semantic dictionaries of the works of writers and poets representing the literary language - various concordances on the Internet space should also be carried out on a single principle. This may be important in terms of creating concordances that can fully cover a specific language in the future;

– in the future, the development of scientific foundations and unified principles for the placement of language material on the Internet through joint creative cooperation between linguists, computer linguistics specialists, engineers, and programmers is an important issue of the day.

Scientific novelty of the research. The scientific novelty of the dissertation is that it presents a comprehensive and systematic analysis of various forms of language materials placed in electronic space or on the Internet, a pioneering achievement in the field of Azerbaijani linguistics. In the course of the research process, a range of electronic dictionaries, electronic libraries and encyclopedias have been classified, and their potential uses have been investigated. The deficiencies of extant electronic lexicons and translation software designed for the Azerbaijani language have been identified, and recommendations have been given for the development of more optimal structures. Concurrently, the dissertation has delineated the linguistic databases that characterise world languages, their types, their role in linguistic research, as well as described the structure of the extant information base of the Azerbaijani language and outlined proposals for its improvement. In addition to the analysis of the structural features of national language corpora in various languages, proposals were made regarding search and automatic marking systems for the national corpus of the Azerbaijani language.

The theoretical and practical significance of the research. The issue of optimal placement and utilisation of language material in electronic space is of great theoretical importance for modern linguistics, information technologies and pedagogy. Its study allows us to deepen our understanding of how the digital environment affects the functioning of language and its presentation, dissemination, and learning. Firstly, the study of the optimal placement of language material in electronic resources contributes to the development of the theoretical foundations of digital linguistics, in particular, to the issues of classification, structuring and formatting of language data. This enables the establishment of standards and methodologies that guarantee the effective utilisation of linguistic resources in digital environments, including electronic dictionaries,

text corpora, training platforms and linguistic databases. Secondly, the analysis of the possibilities of using language material in electronic space opens up new perspectives for the study of the phenomenon of language dynamics. Digital technologies facilitate the tracking of changes in language, the analysis of speech data in real time, and the study of the relationships between various aspects of the language system. This contributes to the expansion of the theoretical basis of linguistics and its connection with other sciences, such as cognitive science, sociolinguistics and computational linguistics. Thirdly, the study of the use of language material in the digital environment contributes to the development of multimodal communication theories, since electronic space allows for the integration of textual, audiovisual and interactive elements. This theoretical framework provides a foundation for the development of innovative knowledge presentation methods and learning opportunities, which are crucial for the creation of effective digital educational resources.

The theoretical significance of the topic is therefore twofold: firstly, it provides a basis for optimising the presentation of language material in electronic space; secondly, it opens up new horizons for linguistic research by improving the quality of educational resources and contributing to a deeper understanding of language processes in context.

At present, a abundance of written texts, audio fragments, images, other illustrative materials, websites, blogs, online newspapers and magazines, PDF versions of scientific, artistic, journalistic texts, and so forth, which contain a wide variety of information, are placed in the electronic space on a daily basis. Language materials placed on the Internet represent a variety of languages. In recent years, a wide variety of materials related to Turkic languages, including Azerbaijani, have been placed on this platform. In the process of placing language materials on the Internet, it is possible to benefit from dissertation work as a practical tool in the placement of electronic dictionaries, language corpora, educational teaching materials, and in the preparation of blogs, websites, and other mass media for various purposes.

Approbation and application of the research. The main provisions of the research work were reflected in articles and theses published in republican and international indexed scientific journals, and in reports made at various scientific conferences.

7 scientific articles (including 1 abroad), 10 conference materials (including 2 abroad) were published in connection with the dissertation work.

Name of the organization where the dissertation work is carried out. The dissertation work was carried out at the Department of AI and Computational Linguistics of the Linguistics Institute named after İ.Nasimi of the Azerbaijan National Academy of Sciences.

The volume of the structural units of the dissertation separately and the total volume with a sign. The dissertation consists of an introduction of 7 pages, I chapter of 51 pages, II chapter of 40 pages, III chapter of 37 pages, conclusion of 2 pages, list of references of 19 pages, list of abbreviations of 2 pages. The total volume of the dissertation is 160 pages; the number of characters is 237,796 characters.

MAIN CONTENT OF RESEARCH

The Introduction section of the dissertation is where the scientific relevance and degree of development of the research topic is justified. It is also where the goals and tasks set during the research are explained, and the provisions presented for defence. The dissertation provides information regarding its theoretical and practical significance, object and subject, research method, approval and application, structure of the dissertation, and total sign volume.

The I chapter of the dissertation is entitled “**Linguistic material in electronic space as an information source**” and consists of seven paragraphs. The initial paragraph of I chapter explores the notion of “**Electronic space as an information source**”. This paragraph will discuss the information, its characteristics, linguistic information, information space and its elements. It will also discuss the fact that the Internet is a means of

storing and using information, helping people to obtain the necessary information, conduct correspondence, get acquainted with the news, get education, learn languages, etc., regardless of their location.

Electronic space, which is of great importance in the development of society, is defined as an environment consisting of information and communication networks. The advent of electronic space has facilitated the acquisition of information through a range of capabilities, including cloud technologies, artificial intelligence, large-scale databases and internet resources.

The categorisation of information is predicated on its diversity and characteristics. Accordingly, information is categorised into the following types: visual, auditory (sound, phonetic), audiovisual and tactile. This categorisation is based on the perception of the senses. Furthermore, according to the orientation of information, there are special types intended for specialists in various fields of mass information, science, technology, culture and production, intended for all members of the information society. Each distinct category of information is further subdivided into subtypes. To illustrate this point, scientific information is typically categorised into various types, including physical, mathematical, biological, linguistic, and so forth, in accordance with the specific discipline or field in which it is situated.

“Linguistic information refers to the set of information, concepts, and rules about language that are reflected in our consciousness and influence our speech behavior”²⁰. The Internet is a unique place for the placement and use of language materials. There are the following types of language materials placed on the Internet:

1. Various types of electronic dictionaries;
2. Formal text analysis programs;
3. Automatic concordances;
4. Educational teaching aids and informative information portals;

²⁰ Рашидова, З.Д. Курс лекций по дисциплине «Информационные технологии в лингвистике»/ З.Д.Рашидова. – Махачкала: ДГУНХ, – 2018. – с.27.

5. Electronic libraries and bibliographic information systems;
6. Linguistic databases (national language corpora, typological, etymological, linguogeographic), etc.

The II paragraph of I chapter is entitled **“Information Exchange in Electronic Space as a Medium of Communication”**. This chapter employs a comparative analysis of Internet communication and traditional communication.

It is widely acknowledged that communication constitutes one of the most significant components of language. In the field of linguistics, the term “communication” is frequently employed as a synonym for the term “conversation”.

The advent of information technologies has engendered a setting conducive to the proliferation of novel modes of electronic communication, wherein the creation and storage of heterogeneous resources, the evolution of web services, and the dissemination and exchange of information occur. Information and communication technologies (ICT) are currently a key aspect of the modern world, facilitating the exchange of information between individuals residing in geographically distant locations.

In clarifying the term “electronic communication”, Y.A.Zulyar considers this type of communication to be the subsequent stage in the evolution of social communication, whereby information in oral (physiological) and written (virtual) forms is converted into electronic form²¹. A.V.Sokolov considers electronic communication to be a type of social communication, along with oral and written communication. K. Habibova²², in turn, writes about electronic communication: *“Indeed, in the domain of modern linguistics, virtual discourse and electronic communication are regarded from diverse perspectives: as a manifestation of social interaction and the execution of social roles, as a form of oral communication, as well as a distinct novel form of speech or a specific category of written*

²¹ Зуляр, Ю.А. Массовые коммуникации в рекламе: учеб. Пособие / Ю.А.Зуляр. – Иркутск: Ир-кут. гос. ун-т, – 2006. – 488 с.

²² Соколов, А.В. Социальные коммуникации: учебник / А.В.Соколов. – Санкт-Петербург: Профессия, – 2014. – 287 с.

speech”²³. In general, based on the opinion of Y.A. Zulyar, electronic communication can be considered a developed stage of social communication. It is the author's opinion that electronic communication has evolved, resulting in the emergence of a novel form of communication. Consequently, we have observed numerous revolutionary advancements in the domain of communication, including oral, written, printed, and electronic media (e.g., radio broadcasting, cinema, television). Most recently, the advent of the internet has further transformed the landscape of communication.

The III paragraph of I chapter is concerned with the “**Forms and methods of electronic communication, communication in a limited and wide circle**”. This paragraph examines various forms of electronic communication as a new means of communication and their main features.

In the contemporary field of modern linguistics, there is no clear-cut methodology for the classification of acts of communication on the Internet. Consequently, electronic communication is grouped by scientists on various grounds. F.e. E.N.Qalichkina²⁴ and E.A.Biryukova²⁵ distinguish internet communication into short-term (e.g. email viewing) and long-term (e.g. electronic conference or webinar) forms of communication, dependent upon the scale (human-computer, human-human, human-human group) and duration.

D.R.Valiakhmetova gives preference to oral (with a dictaphone, microphone) and written (text and graphic information)

²³ Həbibova, K.Ə. Virtual diskursun linqvistik xüsusiyyətləri / K.Ə.Həbibova. – Bakı: Zərdabi Nəşr MMC, – 2023. – s. 28 .

²⁴ Галичкина, Е.Н. Специфика компьютерного дискурса на английском и русском языках (на материале жанра компьютерных конференций): / на соискание ученой степени кандидата филологических наук / – Волгоград, – 2001. – 212 с.

²⁵ Бирюкова Е.А. Интернет-коммуникация как средство межкультурного общения: жанровые и языковые особенности // – Москва: Язык. Словесность. Культура. – 2014. № 6, с. 75-98.

forms of virtual communication²⁶. In analysing the communication environment, H.H. Huseynov focuses on the distinction between virtual interlocutors, characterised by anonymity or unrecognised identity, and actual interlocutors, who are recognised or known to the speaker.²⁷ In another position, A.L.Kholod proposes a taxonomy of Internet communication according to two physical parameters: the location of the communication participants (physical or virtual) and time. The taxonomy comprises the following categories: at the same time and place (virtual) communication (chat); at the same time but in different places (audio and video conferences, Skype, Zoom, WhatsApp, etc.); communication at the same place but in different times (forums, blogs, social networks, etc.); and communication in different places at different times (e-mail).²⁸ Taking all of this into account, we can classify electronic communication as follows:

- by scale (mass, group, and one-on-one);
- by duration (short and long-term);
- by form of communication (written and oral);
- by type of interlocutor (virtual and actual).

In IV paragraph of I chapter, “**Methods of text processing and information retrieval in electronic space**” are studied. Natural language processing occupies an important place in text retrieval technologies. The process of natural language processing (Natural Language Processing, NLP) includes understanding, analysis, performing various operations on texts, as well as their creation.²⁹

All information retrieval methods are predicated on the processing and analysis of the texts of indexed documents. The

²⁶ Валиахметова, Д.Р. Письменная разговорная речь в контексте особенностей Интернет-дискурса // Бодуэновские чтения: Бодуэн де Куртенэ и современная лингвистика: Междунар. науч. конф., – Казань, – 11–13 дек. 2001, – с. 7–9.

²⁷ Гусейнов, Г.Г. Заметки к антропологии русского Интернета: особенности языка и литературы сетевых людей // – Москва: Новое литературное обозрение. – 2000. № 43, – с. 56 - 59.

²⁸ Холод, А.Л. Понятие интернет-коммуникации // – Минск: ИВЦ Минфина, Карповские научные чтения: сб. науч. ст.– 2017. Вып. 11. – с. 298-302.

²⁹ Когаловский, М.Р. Перспективные технологии информационных систем / М.Р. Когаловский, – Москва: ДМК. Пресс; Компания АйТи, –2003. – 288 с.

majority of information systems are designed to index all documents available within the system.³⁰ A number of problems arise when searching for indexed documents and texts in information retrieval systems, including the following:

- the problem of synonymy;
- the problem of homonymy;
- fixed word combinations;
- morphological variations.

In order to address these issues, current search engines employ text analysis methodologies when processing natural languages. Two principal approaches have been identified for the analysis of these phenomenon: statistical and linguistic.

An effective approach based on statistical analysis is implicit semantic indexing. Implicit semantic indexing represents a search engine algorithm devised for the analysis of web pages. This algorithm takes into account not only the keywords and phrases present on the page, but also the presence of their synonyms, as well as words thematically related to the query. The basis of information research systems that perform automatic processing of text and information research in natural language is linguistic processors. The majority of contemporary linguistic processors are characterised by a modular design, with each stage of text analysis or synthesis being represented by a distinct processor module. Linguistic processors divide the analysis of text into such stages as grapheme, morphological, syntactic and semantic.

In addition, this subsection also discusses semantic and keyword search methods. A semantic search enables you to find documents that are not explicitly included in the search query, but are related to it. In contrast, standard internet searches using keywords are based on the lexical compatibility of the requested terms with those found in internet documents.

V paragraph of I chapter is entitled as **“The role of formal analysis programme in the usage of language materials”**. At the

³⁰ Тихонов, В. Архитектура метапоисковых систем: [Электронный ресурс] / URL: http://www.cmsmagazine.ru/library/items/internet_in_fo/metasearch/

present time, special computer programs developed to solve linguistic problems related to text analysis are performing two functions. Firstly, they are performing large-scale tasks, such as frequency analysis, morphological and syntactic analysis. Secondly, they are performing relatively difficult tasks, such as semantic analysis, automatic determination of the style of the text and even determination of the probable author. This paragraph is concerned with an examination of the working principles of text analysis programs developed for various foreign languages. The primary group of programs for automating text processing is identified, and a comparative analysis is carried out. The programme under discussion in this text analyses the text at different levels (grapheme, morphological, lexical, syntactic and semantic), calculates the frequency of words in the text, and determines the style and originality of the text.

VI paragraph of I chapter is about **“The types of electronic dictionaries and their usage possibilities”**. Among the forms of providing language materials in the electronic space, electronic dictionaries are particularly distinguished due to their importance and relevance. In this study, an investigation is made into the comparative performance of traditional paper dictionaries and electronic lexicographic products with regard to their informative and navigational capabilities. The study goes on to justify the creation of various types of electronic dictionaries, and to provide a brief description of all types of existing electronic dictionaries. Consequently, during the course of the research, it was ascertained that electronic dictionaries have been developed for almost all types of traditional dictionaries. It is acknowledged that traditional dictionaries are categorised as linguistic, encyclopedic, intermediate (linguocultural and terminological) dictionaries. The classification of electronic dictionaries is not carried out according to a single principle; therefore, in this paragraph, a classification is also carried out according to technical and operational characteristics. Concurrently, the developmental stages of electronic dictionaries and databases are documented.

II Chapter of the dissertation is entitled “**Linguistic databases as an important and reliable source**” and consists of four paragraphs. In this paragraph, the typological databases in electronic space as a linguistic information resource is examined. Here the focus is on the analysis of linguistic databases developed worldwide, with a particular emphasis on their characteristics. Additionally, the role of typological databases in comparative-typological studies is examined through a comparative lens.

In addition to general databases, a linguistic database is defined as a linguistic information resource defines different levels of a language. A linguistic database is defined as a component of linguistic information resources, the function of which is to collect information about the world's languages into a single database. A linguistic database is divided into three distinct levels: phonetic, lexicographic and grammatical. One of the most recent and promising areas of linguistics among Internet resources is typological databases (TVB). Typological databases are repositories of information pertaining to the grammatical features and characteristics of various languages, including endangered languages. These databases are regarded as the most efficacious tools for conducting research in the field of comparative linguistics. To illustrate this point, consider the number of grammatical genders observed in different languages.

There are two main types of TVB:

“– *characteristic TVB, covering a wide range of language phenomena;*

– *specialized TVB, designed for the study of individual language phenomena*”³¹.

In the course of the study, a number of characteristic and specialised TVBs representing world languages that have developed in recent years were examined, and a comparative analysis was conducted.

³¹ Кружков, Г. Информационные ресурсы контрастивных лингвистических исследований: типологические базы данных// Системы и средства информ. – Москва: – 2015, том 25, выпуск 1, – с. 200.

The II paragraph of II chapter is entitled **“Issues of using national language corpora in the research process”**. This paragraph discusses the role of national corpora in various fields of linguistics and classifies them in terms of the diversity of issues of research and application of corpora.

In the domain of modern linguistics, national language corpora serve as a primary research source, providing insights into the phonetic, morphological, syntactic and semantic characteristics of the language. A national corpus is defined as an electronic resource that is equipped with a search system for various criteria. This resource contains linguistic data in both oral (speech) and written (text) formats in any given language, and is used for the purpose of conducting research.

It is evident that national corpora boast a number of salient features that render them of significant value for the purposes of typological research. Firstly, the national corpus comprises all types of written and oral texts presented in the language (fiction, journalism, science, business, etc.) along with their characteristics and additional information about them. All of these texts are placed in the appropriate section of the corpus. Given the manner in which these texts have been organised, it is now possible to locate the required examples with ease (for example, advertising texts). The primary objective of the corpus compilers is to ensure the inclusion of all possible genres, styles and topics in the corpus; as such, the corpora are subject to constant updating. Secondly, the national corpus facilitates access to live speech and authentic language material, while simultaneously enabling the statistical acquisition of examples of specific linguistic phenomena within a brief timeframe. It is important to note that language units are presented in the corpus in a natural language environment and not in an artificially created context. This, in turn, facilitates comprehensive and objective research.

In the III paragraph of II chapter, **“The use of text corpora in the scientific research process and the possibilities of their optimal placement in the electronic space”** is analyzed. Modern linguistic research, work on the compilation of dictionaries and

grammars is aimed in one way or another at the use of representative text corpora. In our opinion, the main feature that distinguishes text corpora from other electronic linguistic resources is the representativeness of the corpus and the marking of texts. This paragraph examines the features of various text corpora prepared in the world, compares them, and notes their use for research. For example, the “National Corpus of the Russian Language” has been operating on the Internet since 2003 and is free of charge for any online user. Currently, this “Corpus” includes a total volume of more than two billion words from various Russian texts equipped with linguistic marking. Each of these texts is a large-scale, valuable source of linguistic research in terms of quantity and quality. The following types of markup are used in this corpus: morphological, word-formation structure marking (morphemic composition), syntactic, semantic, lexical-functional, and meta-markup (text parameters).

The National American Corpus (NAC), an analogue of the British National Corpus (BNC), contains 22 million words. As is the case with numerous national corpora, the full version of the National American Corpus is available for a fee. However, 68% of the words are freely available online. In contrast to the BNC and the RNLC, the corpus under discussion does not possess a search interface. Consequently, searching for information in this particular corpus is carried out using universal corpus managers that are not focused on using specific corpora. The corpus has been provided with metatext, part-of-speech, and partly syntactic markup.

IV paragraph of II chapter is entitled “**Phonetic databases as a key component of oral speech corpora**”. This paragraph provides information on the history of the development of phonetic databases, the principles of forming and designing such databases, and examples of phonetic databases.

The sound (speech) corpus of a language constitutes a significant component of linguistic resources, being founded upon a phonetic database. A phonetic database can be defined as a structured collection of speech fragments provided by software. The primary function of a phonetic database is threefold: firstly, to facilitate the

investigation and testing of hypotheses concerning the characteristics of speech sounds; secondly, to assess and refine the parameters of models and methodologies employed for the analysis of speech signals; and thirdly, to develop automated speech recognition and synthesis systems. A phonetic database is utilised in the study of the phonetic properties of speech at the segment and suprasegment levels. This facilitates the development of automatic speech recognition and synthesis systems, as well as the verification and identification of individuals based on their voice, which is one of the biometric parameters.

Chapter III of the dissertation is entitled “**Azerbaijani language in electronic space**” and consists of five paragraphs. I paragraph of III chapter discusses the issues of “**Localisation of computer programs into Azerbaijani language**”. The term “localisation” (lat. “localis” - “local”, “locus” - “place”) was first used towards the end of the last century. It is defined as the process of adapting software to suit the culture, mentality and worldview of each country by translating its user interface, documents and related program documents from one language to another. This encompasses the translation of the date, time, currency symbols, measurement systems, and documents and texts in the program interface into the national language, with consideration for the characteristics of the country's national thinking and tax legislation, thereby supporting the language and national standards. A number of studies have been conducted in Azerbaijan on the subject of localisation. Firstly, in 2010, the translation of Microsoft Corporation products into Azerbaijani was initiated. This corporation has introduced a number of software products into the Azerbaijani language, including the operating systems “Windows XP”, “Windows Vista”, “Windows 7”, “Windows 8”, and the mobile operating system “Windows Phone”. In addition, it has released the software products “Microsoft Office 2003”, “Microsoft Office 2007”, “Microsoft Office 2010” and “Microsoft Office 2013”. In order to facilitate the localisation of “Microsoft” products into the Azerbaijani language, a dictionary of commands and words used in these programs was compiled. This dictionary is referred to as a glossary. In addition, a list of words and

expressions that have not been utilised was collated. This is referred to as the “Azeri Policheck”. Finally, an “Azerbaijani Style Guide” manual was compiled. This manual was compiled in accordance with the phonetic (orthographic), grammatical, lexical (terminological) rules of the Azerbaijani language. The compendium of manuals has been meticulously compiled in a plethora of languages, numbering over a hundred. Utilising these manuals, one can discern the disparities between English-Azerbaijani and other languages utilised in the program texts. The paragraph compares the differences that arise during localisation between English-Russian, English-French, and English-Azerbaijani.

In the II paragraph of the III chapter, **“The possibilities and prospects of representing the Azerbaijani language in electronic Space in the context of artificial intelligence problems”** are examined. This paragraph analyzes the issues of building an artificial model of a natural language.

The Internet has become an integral part of many people's daily lives, offering a valuable resource for addressing language-related challenges and enhancing linguistic proficiency. In this regard, in order to promote, develop, study any language at the global level and solve various application issues in this language, it is necessary to have various linguistic information resources in the electronic space that optimally represent that language. Consequently, the representation of the Azerbaijani language in the electronic domain, along with the creation and utilisation of active linguistic information resources, have become pressing concerns. The deployment and utilisation of language material within the Internet domain is fraught with numerous challenges. Firstly, it is important to acknowledge the challenge of fully encapsulating the complex characteristics of a natural language, including its lexical and grammatical features, phraseology, and other elements, within the confines of a computer. Natural language is characterised by a constant process of self-regulation, whereby new elements are incorporated into the language while outdated elements are eliminated. Language material that is placed in the Internet space for a certain period of time remains there in a static, frozen state. It is an inevitable consequence of the

dynamic nature of the language that this material will become outdated over time. Its status as a representation of contemporary language is therefore no longer valid. Consequently, it is imperative to devise a mechanism that can guarantee the perpetual updating of the language material disseminated on the Internet.

A number of issues have been identified that relate to the widespread use and promotion of the Azerbaijani language. These issues increase the need to create a National Corpus of the Azerbaijani Language (NCAL). Such issue is the creation of an artificial intelligence system for the Azerbaijani language. The primary objective of artificial intelligence in the field of linguistics is the development of large language models (LLM), which represent a category of natural language processing systems (NLP – Natural Language Processing) capable of comprehending and processing natural language, utilising deep learning (DL – Deep Learning) technologies. The working principle of LLMs is to create a new text by collecting texts from various websites, blogs, etc. on the Internet based on a query. In this regard, the collection of Azerbaijani-language texts in a structured manner in a corpus is one of the current issues. In this instance, the challenges inherent in designing the NCAL and the linguistic marking of texts are particularly salient. The Azerbaijani language is an affix language, which gives rise to significant challenges in the context of automatic marking. The multivariateness and homonymy of suffixes in our language thus complicates the resolution of such issues. To illustrate this point, we may consider the suffix *-in*⁴ in Azerbaijani, which is pertinent to both belonging and possessive case. This conflation of cases presents a challenge for computer programs attempting to distinguish between them. Consequently, manual marking should be favoured over automatic marking until such a time as artificial intelligence programs in Azerbaijan reach a certain stage of development.

The III paragraph of III chapter is entitled “**Electronic dictionaries, electronic textbooks and teaching aids of the Azerbaijani language**”. This paragraph conducts an analysis of the work undertaken in the field of computer lexicography for the Azerbaijani language, with a focus on the current state of electronic

dictionaries and teaching aids. The study examines the following resources: Polyglot, Dilmanc, Azleks, Azerdict Obastan, and the Electronic Corpus of Azerbaijani language dictionaries. The analysis identifies the strengths and limitations of these resources.

In the context of a number of ongoing projects in Azerbaijan, electronic dictionaries of the Azerbaijani language have been developed, encompassing translation, explanatory, orthographic, terminological, synonyms, antonyms, and personal names among others. These dictionaries have been compiled on the principle of “word-translation” and “word-explanation”. Finally, the “Concordances” section has been incorporated into the corpus of electronic dictionaries. This development represents a significant innovation in the field of Azerbaijani electronic lexicography and is regarded as the foundation for corpus dictionaries. The concordance was prepared by employees of the AI and Computer Linguistics Department of the Linguistics Institute. The concordance was initially comprised of the works of Huseyn Javid and Mikayil Mushfig, in addition to the critical text of “Kitabi-Dede Korkud”.

The current dictionaries in operation do not include the pronunciation or auditory perception of words, graphic images, video or animation fragments. For instance, a spelling dictionary may include illustrative elements such as images, animations, and the phonetic articulation of words. In contrast, an explanatory dictionary might feature supplementary features including words with semantic relations to the word, meanings of the word in context, use as a homonym, synonyms, and usage in context. This particular type of electronic dictionary is equipped with a “smart search” system that utilises neural networks. In contrast to conventional search engines, which organise words and expressions into sections of the dictionary, this system employs an innovative approach. It comprehensively analyses the context of a word to identify its various meanings and establish its semantic relations, thus, providing a more comprehensive and nuanced understanding of the word's usage.

IV paragraph of III chapter discusses the “**General information base of the Azerbaijani language**”. This paragraph analyzes the current state of the information base of the Azerbaijani

language, provides suggestions and recommendations for improving this base in the future.

The employees of the AI and Computer Linguistics Department of Linguistics Institute have prepared the “Azerbaijani language information base” in accordance with paragraph 6.4.1. “The creation of a general information base of the Azerbaijani language as the state language” was established as a key objective of the “State Program on the use of the Azerbaijani language in the context of globalization and the development of linguistics in the country”. This initiative has been meticulously executed, and the comprehensive information base has been successfully uploaded to the “www.azerbaycandili.az” portal. The “Azerbaijani language information base” portal contains a variety of resources, including essays on the Azerbaijani language, official documents pertaining to the language, and information regarding Azerbaijani linguists.

In the extant corpus of Azerbaijani language studies, certain issues are addressed in a perfunctory manner, lacking comprehensive elaboration. For instance, the oral speech and dialect features of the Azerbaijani language are not reflected in the information base. In order to expand the capabilities of the information base of the Azerbaijani language, it would be advisable to collect and place phonograms of the first sound films shot in the Azerbaijani language in the database, prepare an electronic version of an explanatory dictionary for all levels of the language, organise a database of a dialectological atlas and install interactive maps, etc.

The final paragraph of III chapter is entitled **“Recommendations on the optimal placement and use of language material in electronic space”**. A plethora of information resources, encompassing language corpora, electronic dictionaries, large language models, text analysis programs, and assorted information technologies and tools, are available globally. There is a significant demand for users to access the electronic resources we have catalogued for a variety of purposes. The considerable interest in Internet resources can be characterised as an indicator of their importance and usefulness. In the contemporary era, characterised by the increasing popularity of journalism, a plethora of information

sources and transmitters are engaged in continuous operation on a 24-hour basis. The degree of professionalism exhibited by individuals responsible for the management of individual websites and bloggers varies significantly. Furthermore, some sites have been found to function as sources of disinformation, contributing to societal and interpersonal confusion. It is imperative to identify methodologies for the analysis and regulation of these issues. It is imperative to supplant the dissemination of superficial, insubstantial information with more pertinent and essential knowledge for the betterment of society. It is recommended that this issue may be given due consideration, and the aforementioned points may be regarded as a preliminary recommendation.

Recommendations for the optimal placement and use of language material in electronic space can be grouped as follows:

- Electronic resources provided to users should be scientifically analyzed in advance by experts in this field, and recommendations should be given for this purpose. These recommendations would allow users to use Internet resources more efficiently. The level of professionalism and worldview of the people managing news sites should be checked, at least partially. The problem is that the information disseminated by such sites is also defective in terms of language. The norms of the Azerbaijani language, even spelling rules, are not followed.

- It is important to analyze the importance and usefulness of language materials posted on the Internet and, first of all, to organize and provide users with access opportunities.

- To compare the placement characteristics of language corpora prepared in different countries with language corpora prepared in Azerbaijan and to apply their advantages to future corpora prepared for the Azerbaijani language.

- In addition to determining the possibilities of representation of the Azerbaijani language on the Internet space, it is necessary to classify the language materials in our native language. Furthermore, there is a necessity to promote the wider representation of important electronic materials of the modern era, such as educational teaching materials and electronic dictionaries, on the Internet.

– As the case in all regions of the world, the Internet, electronic communication and exchange have become pervasive in our country. In such conditions, hundreds of millions of written texts, audio fragments, images and other illustrative means are collected in the electronic space every day. The placement of these materials, which are characterised by their expansive nature, vibrant colouration and diverse content, should be adapted to align with the interests of users.

– The preparation of statistical reports on language materials to be placed in electronic space is essential. These reports should serve as a basis for determining the optimal placement of the materials.

– It is imperative that statistics concerning the most frequently used lexicons, textbooks and other widely consulted printed materials are collated. In light of this, particular attention must be directed towards the creation of electronic versions of said materials and their subsequent placement in the designated virtual repository.

– It is crucial to acknowledge the significance of linguistic databases in facilitating linguistic research. These databases, meticulously prepared and stored in electronic formats, serve as invaluable resources, compiling comprehensive information about world languages. Consequently, due consideration should be given to the meticulous preparation of such databases.

– It is essential to accord due significance to the strategic placement of linguistic corpora, or national language corpora, in the electronic domain. These corpora, in conjunction with teaching-related materials, represent a vital instrument for the effective placement and utilisation of language materials on the Internet.

– In relation to the matters of extensive representation and effective utilisation of the Azerbaijani language in the electronic domain, future endeavours should be directed towards the expansion of activities in the sphere of establishing scientific foundations and unified principles for the optimal placement of Azerbaijani language material in the electronic space. This should be achieved through collaborative creative endeavours between linguists and information technology specialists.

CONCLUSION

1. Internet communication, a recent development, can be defined as a new type of communication. It is a corpus of spontaneous and unprepared texts that combines the main features of oral and written speech. With its genres (e.g. chat, social networks, groups, blogs), it allows us to describe the main psycholinguistic, sociological and cultural features of speech. This is in contrast to traditional means of communication.

2. It is estimated that millions of pieces of information are published on the Internet on a daily basis in a variety of languages. The development of semantic web technologies has facilitated the process of finding and analysing the necessary information. In this regard, significant attention is allocated to the indexing of documents and the sorting of results in search engines, thereby ensuring a more precise response to the user's keyword search.

3. In recent years, a significant number of programmes have been developed by major IT companies such as Google, IBM, and Microsoft for the purpose of automatic text analysis. These programmes have been employed extensively in the domain of text processing and corpus compilation across a wide range of world languages. Nonetheless, the endeavours undertaken within our nation in this domain, despite their origins in the Soviet era and the considerable progress achieved, have not yet attained the desired standard.

4. While electronic dictionaries continue to dominate traditional dictionaries in terms of practicality and distribution, there has been a recent decline in the latter's share in distribution areas. However, this decline is expected to continue until all dictionaries acquire the ability to read and write. Nevertheless, the development and refinement of lexicographical databases and electronic lexicons represents a fundamental challenge in the field of computer lexicography.

5. The term “linguistic information resources” is employed to denote materials posted on the Internet pertaining to language. Linguistic databases constitute a subset of linguistic information

resources, the function of which is to collate information pertaining to world languages. In the context of comparative research, the utilisation of national corpora and typological databases is widely regarded as the most efficacious research tool.

6. One of the primary instruments for the placement and utilisation of language materials on the Internet is the linguistic corpus, or the national language corpus, which serves as a dependable source of information regarding the phonetic, morphological, syntactic, and semantic structure of the language. In the context of modern linguistics, the corpus functions as a research source.

7. The primary objective of this study is to examine the phonetic characteristics of speech at both the segmental and suprasegmental levels. The development of automatic speech recognition and synthesis systems is also a key focus, as is the establishment of a comprehensive phonetic database. This database will serve the purpose of verifying and identifying individuals based on their voice, which is considered one of the biometric parameters. The overarching goal of this research is to create a general database that encompasses the properties of speech sounds.

8. The process of adapting computer programs for a specific region or language is termed “localisation”, whereby the program text is translated according to the standards of each language. Furthermore, special elements specific to a particular language are added to it. In order to achieve high quality during the process of localisation, it is essential that the task is carried out by specialists who possess a high level of knowledge of foreign languages and the cultural characteristics of the language in question. In addition, these specialists must also have experience in the technical field. It is submitted that there is considerable value in the intensification of the localisation work that has been undertaken in recent years on the Azerbaijani language.

9. Given the dynamic and evolving nature of language, there are numerous challenges associated with the placement and utilisation of language material on the Internet. It has been posited by researchers that the development of an artificial model of natural

language may provide a solution to the issue. The acceptance of an artificial model prepared for a natural language is contingent upon its capacity to imitate natural language to a certain extent.

The field of artificial intelligence has the potential to make significant contributions to this area, with the capacity to address numerous challenges related to language.

10. In recent years, a significant amount of research has been conducted in Azerbaijan on the development of electronic dictionaries and online information resources. However, there is still a need to create and develop optimally structured electronic resources based on modern technologies and to systematically describe all levels of the language, with a view to improving speaking, writing, reading and listening skills in the Azerbaijani language.

The main provisions of the dissertation are reflected in the following published articles of the author:

1. Azərbaycan dili elektron məkanda // – Bakı: DİÇİLİK İnstitutunun əsərləri, – 2018. № 2, – s.187-200.

2. Elektron lüğətlərin təsnifat prinsipləri // – Bakı: Filologiya məsələləri, – 2022. № 12, – s. 89-100.

3. Elektron məkanda dil materiallarının verilmə üsulları // Lənkəran Dövlət Universitetinin Elmi Xəbərləri. – 2019. №1, – s. 35-44.

4. Linqvistik verilənlər bazalarının quruluş xüsusiyyətləri // “Azərbaycan dilinin İKT problemləri, İKT-nin Azərbaycan dili problemləri” respublika elmi-praktiki konfransı, – Bakı: 21-22 fevral, – 2023, – s. 156-162.

5. Linqvistik verilənlər bazası linqvistik informasiya resurslarının tərkib hissəsi kimi // Gənc alim və mütəxəssislərin ikinci beynəlxalq konfransı, – Bakı, 3-6 mart, – 2020. – s. 418-419.

6. Mətnin avtomatik təhlil proqramları və onların iş prinsipləri // – Bakı: Elm, Dilçilik araşdırmaları, – 2020, №1, – s. 117-122.

7. Semantik axtarış sisteminin özəllikləri // – Bakı: Terminologiya məsələləri, – 2024. №1, – s. 202-207.

8. Kompüter proqramlarının Azərbaycan dilinə lokallaşdırılması // Elmlərarası inteqrasiya: linqvodic, linqvokulturoloji və psixolinqvistik aspektlər beynəlxalq elmi konfransın materialları. – Sumqayıt: 19-20 dekabr, – 2019. – s. 382-384.

9. Azərbaycan dilinin elektron lüğətləri // “Müstəqillik illərində üslubların inkişafı və dilin lüğət tərkibinin zənginləşmə istiqamətləri” mövzusunda respublika elmi konfransı, – Bakı: Elm, – 25-26 dekabr, – 2019, – s. 231-235.

10. Türk dünyasında maşın tərcüməsinin probleminə bir baxış // VI Beynəlxalq Türk Dünyası Araşdırmaları Simpoziumu, – Bakı: Elm və təhsil, – 13-15 iyun, – 2019, – s. 99-103.

11. Роль лингвистических баз данных в исследованиях // Международная научно-практическая конференция «Инновационная траектория лингвистических исследований в современном мире». – Москва: Институт иностранных языков, Российский университет дружбы народов, – Мехико: Национальный автономный университет Мексики UNAM, – 3 декабря, – 2021. – с. 36-44.

12. Compiling of Phonetic Database Structure: [electronic resurs] // Path of Science: International Electronic Scientific Journal, – 2021. Vol. 7, № 4, – s. 4001-4006. URL:

<https://pathofscience.org/index.php/ps/article/view/881/867>

13. Konkordanslar korpus dilçiliyinin mühüm sahəsi kimi // Tanınmış türkoloq alim Məhəbbət Mirzəliyevanın 70 illiyinə həsr olunmuş “Azərbaycan dilçiliyinin aktual problemləri” mövzusunda respublika elmi konfransı, – Bakı: Unicopy çap mərkəzi, – 28 oktyabr, – 2019, – s.188-191.

14. Böyük dil modellərinin Azərbaycan dilinə tətbiqi və problemləri // International scientific and practical conference on “Issues of Turkic Philology” at Fergana State University of the Republic of Uzbekistan, – 21-22 may, – 2025, – s. 608-612. URL: <https://zenodo.org/records/15448552>

15. Yeni ünsiyyət növünün formalaşmasında elektron məkanın rolu // – Bakı: Elm, Dilçilik araşdırmaları, – 2025, № 2, – s. 116-123.

16. Dilçilikdə süni intellekt məsələlərinə dair // AZLİNG
2024 – Azərbaycan dili və müasir dilçilik: problemlər, perspektivlər
və çağırışlar” adlı növbəti beynəlxalq elmi konfransının materialları.
– Bakı: Elm və təhsil, – 2025. – s. 135-143.

17. Milli dil korpusları və böyük dil modelləri: türk dilləri
kontekstində təhlil // Süni intellekt və humanitar elmlər: müasir
vəziyyət və perspektivlər” Beynəlxalq elmi konfransın materialları. –
Bakı: ADU, – 21 noyabr, – 2025. – s. 51-53.



The defense will be held on 3 November 2026 at 13⁰⁰ at the meeting of the Dissertation council FD 1.06 operating under the Linguistics Institute named after Nasimi of the Azerbaijan National Academy of Sciences.

Address: Baku, AZ 1143, The avenue H.Javid 115, V floor, ANAS, The Institute of Linguistics named after Nasimi.

Dissertation is accessible at the Library of Azerbaijan National Academy of Sciences, The Institute of Linguistics named after Nasimi.

Electronic version of the abstract is available on the official website of the Institute of Linguistics named after Nasimi, Azerbaijan National Academy of Sciences.

Abstract was sent to the required addresses on 17 September 2026.

Signed for print: 17.03.2026

Paper format: 60x84 $\frac{1}{16}$

Volume: 47 000

Number of hard copies: 20