

THE ROLE OF MODERN LINGUISTIC RESEARCH AND INNOVATIVE TECHNOLOGIES IN THE DEVELOPMENT OF PHILOLOGICAL SCIENCES

<https://doi.org/10.5281/zenodo.21568606>

Risqulova Surayya Kurganovna

Abstract

This article analyzes the role of modern linguistic research and innovative technologies in the development of philological sciences. It examines the potential of corpus and computational linguistics, digital text analysis, and artificial intelligence technologies to improve the effectiveness of scientific research.

Keywords

philology, modern linguistics, innovative technologies, corpus linguistics, computational linguistics, digital analysis, artificial intelligence.

Аннотация

В статье анализируется роль современных лингвистических исследований и инновационных технологий в развитии филологических наук. Рассматриваются возможности корпусной и компьютерной лингвистики, цифрового анализа текста и технологий искусственного интеллекта в повышении эффективности научных исследований.

Ключевые слова

филология, современная лингвистика, инновационные технологии, корпусная лингвистика, компьютерная лингвистика, цифровой анализ, искусственный интеллект.

Annotatsiya

Ushbu maqolada filologiya fanlarini rivojlantirishda zamonaviy lingvistik tadqiqotlar va innovatsion texnologiyalarning o'рни tahlil qilinadi. Korpus lingvistikasi, kompyuter lingvistikasi, raqamli matn tahlili va sun'iy intellekt texnologiyalarining ilmiy izlanishlar samaradorligini oshirish imkoniyatlari yoritiladi.

Kalit so'zlar

filologiya, zamonaviy lingvistika, innovatsion texnologiyalar, korpus lingvistikasi, kompyuter lingvistikasi, raqamli tahlil, sun'iy intellekt.

INTRODUCTION

In the context of today's globalization and digitalization, the development of philological sciences and the study of linguistic and literary phenomena based on modern scientific approaches have become increasingly important. The expansion of information flows, the intensification of international communication, and the growing volume of digital texts require traditional philological research methods to be enriched with new methodologies and technologies. In this regard, modern linguistic research plays a significant role in improving the theoretical foundations of philological sciences and providing a deeper understanding of the meaning, functions, and usage characteristics of linguistic units.

The development of anthropocentric, cognitive, communicative-pragmatic, linguocultural, and discourse-based approaches in contemporary linguistics makes it possible to study language not merely as a grammatical system, but as a phenomenon closely connected with human thought, culture, social relations, and communicative activity. These approaches contribute to a comprehensive analysis of language and speech phenomena, the identification of the national and cultural characteristics of different languages, and the strengthening of the interdisciplinary nature of philological research.

The integration of innovative technologies into scientific research has also created new opportunities in the field of philology. Corpus linguistics, for instance, enables large volumes of texts to be systematically collected, classified, and statistically analyzed. Computational linguistics facilitates the automatic processing of natural language units, the development of electronic dictionaries, machine translation, and speech recognition. Artificial intelligence, machine learning, and digital text analysis tools allow linguistic data to be processed rapidly, hidden patterns to be identified, and the accuracy and reliability of scientific conclusions to be improved.

Modern technologies are also having a considerable impact on the development of research in literary studies and translation studies. Within the framework of the digital humanities, opportunities are expanding to analyze the linguistic features of literary works, an author's individual style, systems of characters, genre characteristics, and intertextual relationships using electronic tools. Digital libraries, academic databases, and online platforms further enhance philologists' access to research sources, facilitate the exchange of scientific findings, and broaden their participation in international research projects.

At the same time, the use of innovative technologies requires researchers to possess digital literacy, the ability to critically evaluate information, competence in working with electronic sources, and the skills necessary to interpret the obtained results scientifically. Technological tools can accelerate the research process;

however, their outputs must be analyzed on the basis of philological knowledge, theoretical preparation, and scientific reasoning. Therefore, integrating traditional philological methods with modern digital approaches is one of the principal factors ensuring the consistent development of the field.

Thus, the integration of modern linguistic research and innovative technologies expands the scientific potential of philological disciplines and creates favorable conditions for the accurate, systematic, and comprehensive study of linguistic and literary phenomena. Research in this area is particularly important for improving the theoretical and practical effectiveness of philological studies, shaping a modern scientific methodology, and advancing the field in accordance with international scientific achievements.

Degree of Research on the Topic

The use of modern linguistic research and innovative technologies in the development of philological sciences has been studied from various perspectives by foreign, Russian, and Uzbek scholars.

Among foreign scholars, N. Chomsky explored the generative nature of language and issues of linguistic theory; D. Crystal investigated modern trends in linguistics and the language of the Internet; J. Sinclair and T. McEnery developed methods of corpus linguistics and electronic text analysis; while G. Leech examined the principles of computer-assisted linguistic data analysis.

Among Russian scholars, V. A. Zvegintsev studied the theory and methodology of linguistics; Yu. N. Karaulov developed the concept of linguistic personality; V. Z. Demyankov investigated cognitive linguistics and discourse analysis; and A. N. Baranov addressed issues of computational linguistics and automatic text processing.

Among Uzbek scholars, A. Nurmonov examined the theoretical and methodological foundations of Uzbek linguistics; Sh. Safarov studied cognitive linguistics and pragmalinguistics; N. Mahmudov investigated the relationship between language, thought, and speech culture; D. Khudoyberganova explored the anthropocentric analysis of texts; and B. Mengliyev researched modern Uzbek-language corpora and problems of computational linguistics.

These studies have addressed the theoretical foundations of modern linguistics, corpus and computational linguistics, and digital text analysis. Nevertheless, the comprehensive study of these areas through their integration with innovative technologies remains a relevant scientific task.

Statistical Analysis and Results

At present, modern linguistic research aimed at advancing philological sciences is mainly conducted in such areas as corpus linguistics, computational

linguistics, natural language processing, artificial intelligence, digital lexicography, machine translation, discourse analysis, and the digital humanities. Indicators from international linguistic platforms demonstrate that the scope of research in these areas is expanding rapidly.

Corpus linguistics is one of the most dynamically developing fields. Currently, the Sketch Engine system contains more than 800 ready-made corpora representing over 100 languages, with a total volume exceeding one trillion words. The largest English-language corpus available in the system contains more than 80 billion words. Moreover, the TenTen family of multi-billion-word web corpora has been developed for more than 50 languages. These figures indicate that linguistic phenomena can now be examined not merely through small collections of individual texts, but on the basis of datasets containing billions of words.

The use of online texts in digital linguistic research has also increased considerably. The Common Crawl open database contains more than 300 billion web pages collected over 19 years, with approximately 3–5 billion new pages added every month. Its data have been used in more than 12,000 academic studies. Furthermore, at least 64 percent of the 47 major language models developed between 2019 and 2023 were trained using Common Crawl data. This confirms that natural language processing and AI-based linguistic research increasingly rely on large-scale textual datasets.

The study of multilingualism has also become an important area of contemporary linguistics. Common Crawl's language identification tool can distinguish 160 languages. However, these languages are not represented equally in online content. In one of the analyzed archives, English-language materials accounted for 44.42 percent, German-language materials for 5.44 percent, and Polish-language materials for 1.75 percent. Of the 160 languages examined, only 14 accounted for more than one percent of online content. This statistical disparity demonstrates the need to develop national electronic corpora, digital dictionaries, and automated analysis tools for low-resource languages, including Uzbek.

Modern research also devotes considerable attention to the digital analysis of parliamentary speech, political discourse, media language, and social communication. For example, a parliamentary corpus developed within the CLARIN infrastructure covers 29 parliaments, 24,000 speakers, and one billion words. Such corpora make it possible to statistically examine political terminology, speech strategies, gender-related language patterns, social relations, and historical changes in language. The average of 2,909 monthly access sessions recorded through CLARIN's central platform indicates a growing academic demand for digital language resources.

Within computational linguistics, research is expanding in automatic text classification, speech recognition, sentiment analysis, question-answering systems, terminology extraction, and machine translation. Large language models, in particular, are being actively applied to text generation, editing, translation, and semantic analysis. As a result, the object of linguistic research now extends beyond conventional written texts to include spoken language, audio recordings, video materials, online communication, and multimodal discourse.

Significant transformations are also taking place in digital lexicography. Electronic corpora make it possible to automatically identify word frequency, emerging meanings, grammatical characteristics, and collocational patterns. Consequently, the traditional process of dictionary compilation is evolving into an electronic, continuously updated system that can adapt to users' needs. This approach enables researchers and lexicographers to identify new terms, neologisms, and language units specific to online communication more rapidly.

The overall statistical analysis reveals three major trends in contemporary linguistic research: the expansion of research datasets to millions and billions of words, the automation of linguistic analysis, and the growing demand for multilingual digital resources. Therefore, the priority tasks for advancing Uzbek philology include improving a large-scale national corpus of the Uzbek language, expanding spoken and written language databases, creating AI-compatible linguistic datasets, and developing software for the automatic analysis of Uzbek-language texts.

Discussion

Philological sciences constitute a complex of disciplines that study the principles governing language, literature, texts, speech, translation, and intercultural communication. In the current context of globalization and digitalization, the content and methodology of philological research are undergoing significant transformation. Rather than limiting the study of linguistic phenomena to traditional descriptive or comparative-historical methods, there is a growing need to analyze them using large-scale digital data, computer programs, and artificial intelligence tools. As a result, new scientific fields based on the integration of linguistics, information technology, psychology, cultural studies, sociology, and cognitive sciences are emerging in contemporary philology.

An important feature of modern linguistic research is the study of language in close connection with the human factor. The anthropocentric approach interprets language as an expression of human thought, worldview, national values, and communicative activity. This approach has contributed to the development of such fields as cognitive linguistics, pragmalinguistics, linguocultural studies, and

psycholinguistics. Cognitive linguistics examines how knowledge and concepts are represented in linguistic units, while pragmalinguistics investigates the purpose and impact of speech in a particular communicative situation. Linguocultural studies, in turn, reveal the relationship between language and national culture.

Discourse analysis is also one of the leading areas of contemporary philology. Unlike an ordinary text, discourse encompasses the author, listener or reader, social environment, communicative purpose, and cultural context as an integrated whole. The study of political, academic, pedagogical, medical, advertising, online, and media discourses makes it possible to identify communicative relationships within society, methods of influence, and the social functions of linguistic units. In particular, research into the linguistic features of social media communication, blogs, online news, and digital comments is generating new scientific findings.

Corpus linguistics has introduced greater accuracy and evidence-based analysis into modern linguistic research. A language corpus is a collection of texts compiled according to specific principles, classified, and prepared for electronic processing. Corpora make it possible to determine word frequency, the distribution of grammatical forms, fixed expressions, terms, neologisms, and stylistic features. This method reduces the subjectivity associated with selecting isolated examples and enables researchers to substantiate their conclusions using large amounts of empirical material.

Parallel corpora are particularly important for translation studies and comparative linguistics. In such corpora, an original text is aligned with its translation into another language. This makes it possible to examine the equivalence of translation units, grammatical transformations, methods of expressing national and cultural concepts, and the translator's individual approach. Expanding parallel corpora of Uzbek and English can help improve translation quality, enrich electronic dictionaries, and enhance machine translation systems.

Computational linguistics is developing at the intersection of philology and information technology. This field studies the processing of human language through computational methods. Its principal tasks include automatic text editing, spell-checking, part-of-speech tagging, syntactic analysis, speech-to-text and text-to-speech conversion, machine translation, and the development of question-answering systems. These technologies are widely applied not only in scientific research but also in education, translation, the mass media, and publishing.

The development of artificial intelligence and machine-learning technologies has further expanded the possibilities of natural language processing. Modern algorithms can identify recurring linguistic patterns in large text collections, summarize content, classify texts by topic, extract key concepts, and determine

emotional attitudes. These tools save researchers time and increase the efficiency of processing large volumes of material. However, AI-generated results cannot be regarded as absolute scientific truth. They must be verified through philological knowledge, contextual analysis, and critical evaluation.

Digital lexicography is another promising field of philological research. Whereas traditional printed dictionaries reflect the state of a language during a particular period, electronic dictionaries can be updated regularly and can rapidly incorporate new words and meanings. Electronic lexicographic databases may provide a word's definition, pronunciation, grammatical characteristics, examples of use, synonyms, antonyms, and equivalents in other languages. Electronic dictionaries connected to corpus data allow users to observe how words are used in authentic speech and written texts.

The digital humanities are also taking literary research to a new stage of development. The digitization of literary texts enables researchers to examine authorial style, character systems, plot construction, lexical richness, and intertextual relationships using computational tools. Stylometric analysis can identify an author's characteristic word choices, sentence structures, use of grammatical forms, and recurring linguistic units. This method is important for determining the authorship of anonymous works, comparing the styles of different writers, and identifying historical changes within a particular author's creative output.

Opportunities for studying spoken language are also expanding in modern philology. Speech corpora compiled from audio and video materials facilitate the examination of pronunciation, stress, intonation, dialects, sociolects, and communication strategies. Automatic speech-recognition technologies accelerate the conversion of spoken materials into written form. These possibilities have considerable scientific significance for dialectology, phonetics, folklore studies, and sociolinguistics. In particular, digitizing endangered dialects and examples of oral heritage contributes to preserving the richness of the national language.

Innovative technologies are also influencing the content of philological education. Virtual learning platforms, mobile applications, multimedia resources, and interactive tasks create favorable conditions for independent language learning. Using electronic corpora, students can observe how words and grammatical units function in authentic texts, compare lexical meanings through digital dictionaries, and assess their knowledge using automated evaluation systems. Thus, innovative tools applied in research are also being introduced into education, strengthening the relationship between theory and practice.

At the same time, digital linguistic research faces several challenges. These include the quality of electronic texts, copyright protection, personal data security, algorithmic errors, and digital inequality among languages. Online texts do not always conform to standard literary language norms, and conclusions based on such materials may not fully represent all forms of language use. Therefore, electronic materials must be carefully selected, cleaned, classified, and evaluated according to scientific criteria.

In Uzbek philology, it is necessary to develop national language corpora, introduce morphological and syntactic annotation of texts, create electronic terminology databases, and strengthen research on automatic Uzbek speech recognition. Particular attention should be paid to processing Uzbek texts written in both Latin and Cyrillic scripts within a unified system, digitizing dialect materials, creating high-quality parallel corpora, and converting the national literary heritage into electronic formats.

Thus, modern linguistic research and innovative technologies are expanding the scope of philological sciences and making it possible to conduct a comprehensive study of linguistic and literary phenomena based on extensive empirical evidence. Combining the interpretative and semantic potential of traditional philological analysis with the speed and accuracy of digital technologies is one of the principal conditions for the further advancement of philological sciences.

Conclusion

Modern linguistic research and innovative technologies are expanding the theoretical and practical potential of philological sciences. Corpus and computational linguistics, digital lexicography, and artificial intelligence tools make it possible to analyze large volumes of linguistic data rapidly and systematically. The integration of traditional philological methods with digital approaches improves the accuracy, objectivity, and effectiveness of research.

Recommendations

1. The national, parallel, and spoken-language corpora of the Uzbek language should be expanded and regularly updated.
2. Practical training courses on corpus linguistics, artificial intelligence, and digital text analysis should be organized for philologists, teachers, and students.
3. The number of scientific projects aimed at the automatic analysis and translation of Uzbek-language texts and the development of electronic dictionaries using modern technologies should be increased.

REFERENCES:

1. Nurmonov, A. History of Uzbek Linguistics. – Tashkent: Uzbekistan, 2002.
2. Mahmudov, N. In Search of Ways to Conduct a Comprehensive Study of Language. – Tashkent: Mumtoz Soʻz, 2012.
3. Safarov, Sh. Cognitive Linguistics. – Jizzakh: Sangzor, 2006.
4. Safarov, Sh. Pragmalinguistics. – Tashkent, 2008. – 318 p.
5. Khudoyberganova, D. The Anthropocentric Study of Text. – Tashkent: Fan, 2013.
6. Hakimov, M. Fundamentals of Uzbek Pragmalinguistics. – Tashkent: Akademnashr, 2013.
7. Yoʻldoshev, M. Literary Text and the Fundamentals of Its Linguopoetic Analysis. – Tashkent: Fan, 2007.
8. Poʻlatov, A., and S. Muhamedova. Computational Linguistics. – Tashkent, 2007. – 122 p.
9. Rahimov, A. Fundamentals of Computational Linguistics. – Tashkent: Akademnashr, 2011. – 160 p.
10. Xolmonova, Z. Computational Linguistics. – Tashkent: Asian Book House, 2020. – 198 p.
11. Mengliyev, B., Sh. Hamroyeva, and D. Ahmedova. Computational Lexicography. – Tashkent, 2021. – 177 p.
12. Hamroyeva, Sh. A Concise Explanatory Dictionary of Corpus Linguistics Terms. – Tashkent: GlobeEdit, 2020. – 93 p.
13. Zaxarov, V., B. Mengliyev, and Sh. Hamroyeva. Corpus Linguistics: Corpus Compilation and Use. – Tashkent: GlobeEdit, 2021. – 167 p.
14. Dadaboyev, H. Uzbek Terminology. – Tashkent: Nodirabegim, 2020. – 142 p.
15. Sayfullayeva, R., B. Mengliyev, G. Boqiyeva et al. Modern Uzbek Literary Language. – Tashkent: Fan va texnologiya, 2009.