

PHILOLOGY AND LINGUISTICS

Big Data Technologies in Modern Philological Studies within the SHIVA and WARMI Era

Goncharenko Svitlana¹

¹ Senior Lecturer;
Department of philology and translation,
Kyiv National University of Technologies and Design; Ukraine

Abstract. The article explores the possibilities of applying Big Data technologies in modern philology in the conditions of SHIVA and WARMI worlds. The features of the analysis of large-scale text corpora, digital discourse and language data using artificial intelligence are considered. The role of Big Data in the transformation of the methodology of philological research and the development of digital humanities is substantiated.

Keywords: *modern philology; Big Data; digital humanitarianism; artificial intelligence; natural language processing; digital discourse; SHIVA; WARMI.*

Introduction.

The current stage of development of world science is characterized by an unprecedented complication of the information environment, acceleration of the pace of knowledge generation, digitalization of scientific activity and qualitative change of principles of information processing [1]. After all, if a few decades ago philology mainly relied on qualitative analysis of limited arrays of texts, today the object of research is multimillion-dollar collections of digital documents, electronic libraries, archives, media content, social networks, Internet communication and multimodal language data. Such comprehensive changes in information provision necessitate the adaptive implementation of a complex of Big Data technologies, which allow adaptively performing analysis and analytics of processes at a qualitatively higher level of scale, accuracy, efficiency and objectivity [2].

The development of Digital Humanities, computational linguistics, corpus philology, intelligent text analysis and artificial intelligence is gradually transforming traditional

PHILOLOGY AND LINGUISTICS

research approaches. Instead of studying individual works or limited language samples, the researcher gets the opportunity to analyze millions of texts of different genres, eras, languages, and communicative situations, revealing patterns that were previously inaccessible to classical methods of philological analysis. In the conditions of Big Data technology, it becomes not just a tool for automating scientific work, but one of the fundamental factors in the methodological transformation of modern philology. Such processes become particularly relevant in the conditions of the formation of a new global model of social development, which is increasingly characterized through the concepts of SHIVA and WARMI worlds. These concepts reflect qualitatively different, but interconnected characteristics of the modern global space. SHIVA world symbolizes a state of high uncertainty, nonlinearity, rapid technological change, the destruction of established models of the functioning of social systems, the constant emergence of new information flows and digital ecosystems. At the same time, the concept of the WARMI world emphasizes the growing role of sustainability, adaptability, interdisciplinary integration, intellectual collaboration, human-centeredness and harmonization of interaction between technologies and society. It is the combination of high turbulence of the SHIVA environment and the need to form sustainable WARMI research practices that determines the modern requirements for the development of the humanities.

For philology, such a transformation means the need to transition from mainly local research to the analysis of global language processes, which are on the scale of the global information space. Digital traces of human communication daily form colossal volumes of textual information, covering fiction, scientific publications, mass media, Internet communication, social networks, digital archives, multimedia resources and various forms of digital discourse. Processing such data sets is impossible exclusively by traditional methods of philological analysis, which determines the growing importance of Big Data technologies.

Therefore, taking into account the above, the author argues that the main purpose of this article is to study the features of the application of Big Data technologies in modern philology, determine their role in the conditions of the SHIVA and WARMI worlds, and identify the main directions of

PHILOLOGY AND LINGUISTICS

transformation of philological research under the influence of digital technologies.

Main Part.

Big Data technologies are a set of methods, algorithms, software platforms and computing infrastructures designed for storage, processing, integration and intelligent analysis of extremely large arrays of structured and unstructured information [3]. Their fundamental difference lies not only in the volumes, modes and types of processed data [4], but also in the ability to identify complex patterns, hidden relationships and dynamic processes inaccessible to traditional statistical methods [5].

In relation to modern philology, Big Data forms a new research paradigm. If classical philology was based mainly on detailed qualitative analysis of individual literary works, text corpora or linguistic phenomena, then modern digital technologies allow for large-scale quantitative analysis of almost all available textual production of humanity. As a result, the object of research is not individual texts, but complex information ecosystems, including millions of different types of documents.

Big Data technologies are most actively used in the creation and analysis of national and international text corpora. Modern corpora contain billions of word usages, cover different historical periods, genres, functional styles, territorial variants of the language and features of professional communication. The analysis of such arrays allows us to study language dynamics, the processes of the emergence of new words, changes in the semantics of lexical units, the transformation of grammatical constructions, the evolution of syntactic models and the development of communicative practices.

Of particular importance is the analysis of Internet discourse, the volume of which is increasing exponentially every year. Social networks, news resources, blogs, forums, video hosting, user comments and digital platforms form a unique array of linguistic data, reflecting the almost instantaneous reaction of society to political, economic, cultural and technological changes. That is why the study of digital discourse is becoming one of the most promising areas of modern philology.

Big Data technologies allow analyzing not only the content of texts, but also the features of their distribution, citation, information impact, emotional coloring,

PHILOLOGY AND LINGUISTICS

communicative effectiveness and interaction of various discursive communities. The use of Text Mining, Natural Language Processing, machine learning and neural networks methods significantly expands the analytical capabilities of the researcher. Automatic identification of thematic clusters, analysis of message tone, detection of hidden semantic connections, construction of graphs of semantic relations and modeling of language evolution are becoming important elements of modern philological methodology.

In the conditions of the SHIVA-world, such technologies are becoming particularly relevant. The high rate of emergence of new texts, the constant transformation of digital platforms, the rapid change of language norms and the emergence of new communicative forms require almost continuous monitoring of the information space. Traditional research methods are not operational enough to analyze such processes. On the contrary, Big Data technologies provide the ability to process information in a mode close to real time, which allows recording the emergence of new language trends at the early stages of formation.

At the same time, the concept of the WARMI-world assumes that digital technologies should be used not only to increase the efficiency of analysis, but also to ensure the sustainable development of scientific knowledge. This means the need to form open research platforms, develop international digital corpora of texts, standardize methods for processing language data, ensure the reproducibility of scientific results and wide interdisciplinary cooperation of specialists in the fields of philology, computer science, mathematical statistics, artificial intelligence and cognitive sciences.

It should be noted that the implementation of Big Data technologies means abandoning traditional philological methods, and, on the contrary, contributes to the formation of a new integrative methodology for studying language and text. Quantitative analysis, based on the processing of huge amounts of information, allows you to identify statistically significant patterns, but their scientific interpretation, as before, requires the use of classical methods of linguistic, literary, stylistic, discursive and hermeneutic analysis. It is the combination of computing technologies with deep humanitarian interpretation that ensures the most reliable and scientifically sound results.

Big Data technologies are beginning to play a special role in the study of literary heritage. Digital libraries,

PHILOLOGY AND LINGUISTICS

electronic archives and repositories open up the possibility of a comprehensive analysis of works of different eras, national literatures and authorial schools. Automated text processing allows us to study the features of the author's idiosyncrasy, the dynamics of vocabulary use, the structure of the story, the evolution of artistic images, intertextual connections, the features of translations and reception of literary works in different cultural spaces. At the same time, network analysis methods allow us to model complex relationships between authors, works, literary trends and cultural traditions, forming new ideas about the development of world literature.

The study of digital communication is also a promising direction. Modern people create significant amounts of text information every day using social networks, messengers, blogs, forums, news platforms and other digital services. Such data reflect not only linguistic changes, but also the transformation of communicative behavior, changes in speech strategies, the formation of new genres of digital communication, the spread of Internet slang, emojis, memes and other multimodal means of communication. Thanks to big data technologies, it is possible to study such processes at the global level, revealing the patterns of development of digital language practically in real time.

In the conditions of the SHIVA-world, the ability of philological research to respond promptly to rapidly changing information flows is of particular importance. Modern digital platforms generate billions of text messages every day, the content of which reflects the dynamics of public sentiment, cultural processes, political events and social transformations. The use of Big Data technologies allows monitoring such changes, identifying the emergence of new concepts, tracking changes in the semantics of words, the spread of new terms and the formation of relevant discursive practices. Thus, philology becomes not only a science of language and literature, but also an effective tool for studying modern social processes.

At the same time, the principles of the WARMI-world orient the scientific community towards ensuring sustainable, ethically responsible and human-centered development of digital technologies. This involves compliance with the requirements of information security, protection of personal data, transparency of information processing algorithms, reproducibility of scientific results and open access to

PHILOLOGY AND LINGUISTICS

research data. In humanities research, these aspects are of particular importance, since the analysis of linguistic data is directly related to the study of human communication, cultural identity and social consciousness.

The development of artificial intelligence technologies significantly expands the analytical capabilities of Big Data processing in philology. Modern large language models are capable of performing automatic classification of texts, thematic modeling, detection of hidden semantic structures, machine translation, automatic summarization, entity extraction, determination of emotional coloring of messages and prediction of further development of language processes. At the same time, the use of such technologies requires a critical assessment of the results obtained, since artificial intelligence algorithms are able to reproduce statistical patterns, but do not always take into account the cultural, historical, pragmatic and intertextual context of language phenomena.

Thus, the further development of modern philology is associated with the formation of a new adaptive research paradigm based on the synergy [6] of humanitarian knowledge, mathematical methods, Big Data technologies, artificial intelligence and digital infrastructure of science. Such integration allows to significantly expand the scope of philological research, increase the objectivity of scientific conclusions and provide a deeper understanding of the patterns of language functioning in a rapidly changing digital society.

Conclusions.

The conducted study allows us to conclude that Big Data technologies are becoming one of the key factors in the transformation of modern philology. Their use ensures the transition from the analysis of limited text samples to the study of global information arrays, significantly expanding the possibilities of studying linguistic, literary and communicative processes. Unlike traditional methods, Big Data allows us to identify hidden patterns of language development, model complex semantic and discursive relationships, monitor the dynamics of digital communication and analyze language changes in a mode as close to real time as possible.

It has been established that in the conditions of the SHIVA-world, Big Data technologies act as a tool for adapting philological science to high uncertainty, rapid changes in the information environment and continuous growth in the volume of digital texts. At the same time, the WARMI-world

PHILOLOGY AND LINGUISTICS

concept determines the need to use these technologies on the basis of sustainability, openness, interdisciplinary interaction, scientific reproducibility and ethical responsibility. It is the combination of the computing power of modern digital platforms with the humanitarian methodology of interpretation that ensures the receipt of scientifically reliable results. Prospects for further research are related to the development of intelligent systems for natural language analysis, the improvement of methods for processing multilingual text corpora, the integration of artificial intelligence technologies with traditional philological approaches, the creation of new generation digital ecosystems and the development of complex models for studying language dynamics in the context of digital transformation [7]. The implementation of these areas will contribute to the further development of modern philology as a high-tech interdisciplinary field of scientific knowledge, capable of effectively investigating linguistic and cultural processes in the new global information reality.

References:

- [1] I. Hrashchenko, S. Krasniuk, T. Tsalko, S. Nevmerzhytska, N. Liubymova (2024). Effective computer modeling in educational management. *ScientificWorldJournal*. – September 2024. – Issue 27, Part 3. – pp. 3–17. – URL: <https://www.sworldjournal.com/index.php/swj/issue/view/swj27-03>
- [2] Krasnyuk M.T. (2006). Problemy zastosuvannya system upravlinnia korporativnyimi znanniamy ta yikh taksonomiia [Problems of applying corporate knowledge management systems and their taxonomy]. *Modeliuvannya ta informatiini systemy v ekonomitsi* – Modeling and information systems in the economy, vol. 73, p. 256 [in Ukrainian].
- [3] Naumenko, M. (2024). Models of business knowledge in artificial intelligence systems for an effective competitive enterprise. *International scientific journal "Internauka". Series: "Economic Sciences"*. № 6. DOI: <https://doi.org/10.25313/2520-2294-2024-6-10010> [In Ukrainian].
- [4] Naumenko, M. (2024). Analiz ta analityka velikykh danykh v marketynh i torhivli konkurentospromozhnoho pidpriemstva [Analysis and analytics of big data in marketing and trade of a competitive enterprise]. *Grail of Science*, # 40, pp. 117–128. DOI: <https://doi.org/10.36074/grail-of-science.07.06.2024.013> (in Ukrainian).
- [5] M. Krasnyuk, T. Motsyuk, S. Krasniuk (2021). Relevanz des wahlfaches «Wissensdetektion in unstrukturierten daten» bei der ausbildung von mastern der technischen und humanitären fachrichtungen. *Education and science of today: intersectoral issues and development of sciences : collection of scientific papers "ΛΟΓΟΣ"* with proceedings of the I International scientific and practical conference,

PHILOLOGY AND LINGUISTICS

Cambridge, United Kingdom, March 19, 2021. Vol. 3. Cambridge-Vinnytsia : P.C. Publishing House & European Scientific Platform, 2021. pp. 66–67.

- [6] Derbentsev, V. D., Serdiuk, O. A., Soloviov, V. M., & Sharapov, O. D. (2010). *Synergistic and econophysical methods of studying dynamic and structural characteristics of economic systems*. Cherkasy: Brama-Ukraine. – 2010 [in Ukrainian].
- [7] Goncharenko S. (2025). Intelligent information technologies for innovative management of advanced philology projects. *Innovations and New Directions in Scientific Research : Proceedings of the 2nd International Scientific Conference (Manchester, United Kingdom, 20 September 2025)*. – Lulu Press, Inc., 2025. – P. 176–179. URL: <https://er.knutd.edu.ua/handle/123456789/31648>