

PHILOLOGY AND LINGUISTICS

Peculiarities of Big Data Application in Contemporary Linguistics: Navigating Turbulence, Uncertainty and Systemic Risks Through Shiva and Warmi Frameworks

Krasniuk Svitlana¹

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

Abstract. The article explores the features and specifics of Big Data technologies in modern linguistics in the context of the dynamics of uncertainty, turbulence and systemic risks. The potential of Big Data for the analysis of large-scale and dynamic language data, the development of digital philology, the detection of language trends and weak signals is revealed. The need for the integration of Big Data, artificial intelligence and traditional philological methods within the framework of the SHIVA and WARMI concepts of the modern external environment is substantiated.

Keywords: *modern linguistics; computational linguistics; Big Data; digital philology; artificial intelligence; machine learning; natural language processing; language data; uncertainty dynamics; turbulence; systemic risks; SHIVA concept; WARMI concept; digital transformation.*

INTRODUCTION.

Modern socio-economic systems (in particular, the sphere of science and education) are currently developing in conditions of qualitatively more complicated external environment, characterized by the dynamics of uncertainty [1], high turbulence of information processes and increased systemic risks. The transformation of the digital environment [2] significantly changes not only the methods of creating, storing and distributing linguistic information, but also the methodology of its scientific research itself [3]. Modern linguistics gains access to unprecedented in volume, speed of formation, diversity and dynamism of arrays of text and multimodal data [4], which creates an objective need for new tools for their intellectual processing [5]. In these

PHILOLOGY AND LINGUISTICS

conditions, Big Data technologies become not just a technical means of analyzing large arrays of information [6], but an important element of modern scientific methodology, ensuring the transition from local and mainly qualitative analysis of linguistic phenomena to large-scale, dynamic and multidimensional study of language [7].

The use of Big Data is becoming particularly relevant in the context of the conceptual transition from relatively stable models of VUCA and BANI environments to more complex representations of the modern world, which within the framework of SHIVA and WARMI concepts emphasize the interconnectedness of instability, high speed of change, hyper-information saturation, multiplicity of risks, multiplicity of risks. For modern linguistics, this means that language must be considered not as a system of relatively stable structures, but also as a dynamic information system, constantly changing under the influence of social, technological, economic, cultural and communication factors [8]. Traditional methods of linguistic analysis retain their fundamental importance, but their capabilities are insufficient for processing digital corpora, social media, news streams, electronic communications, multimedia content and other sources, the volume of which is constantly growing. Big Data allows you to analyze such arrays [9] using automated collection, integration, storage, cleaning, structuring and intelligent processing of data. This creates conditions for detecting hidden patterns, stable correlations, language trends and weak signals that may be invisible when using traditional sampling methods [10].

The relevance of the study is determined by the need for a systematic understanding of the features and specifics of Big Data technologies in relation to modern linguistics, as well as determining their capabilities and limitations in conditions of high uncertainty, turbulence and systemic risks of the external environment [11]. That is why the current goal of this article is to identify the key features of the use of Big Data technologies in modern linguistics and determine their methodological potential for the development of digital philology in the conditions of SHIVA and WARMI environments.

MAIN PART.

Big Data as a technological basis for the transformation of modern linguistics.

PHILOLOGY AND LINGUISTICS

Big Data is a set of technologies and methods for working with data characterized by large volume, high rate of receipt, diversity, heterogeneity and dynamism. This is especially important for linguistics, since digital communication constantly forms huge arrays of text, speech and multimodal information. The objects of research are digital corpora, social networks, blogs, news streams, electronic documents, speech transcripts, messages and other sources. Their analysis allows us to move from limited samples to large-scale research of speech processes, to identify frequency changes, new constructions, semantic shifts and features of communication behavior. At the same time, Big Data does not replace traditional linguistic interpretation, but expands its empirical capabilities. The most productive is a hybrid model that combines automated quantitative analysis with qualitative interpretation of the results by a philologist.

Specificity of Big Data speech languages.

The specificity of Big Data in linguistics is determined by the complexity of language, its contextuality, polysemy, variability and cultural conditioning. Linguistic data are characterized by heterogeneity, the presence of noise, incompleteness, multilingualism, multimodality and constant variability. Temporal dynamics is of particular importance. Linguistic patterns identified in one time interval may change in the future, therefore modern research should take into account not only the state of the language, but also the direction of its transformation. In the conditions of SHIVA and WARMI concepts, language can be considered as a kind of indicator of environmental changes. The dynamics of vocabulary, discourses, word frequency and semantic connections allows us to detect weak signals of social and informational transformations.

Basic Big Data technologies in modern linguistics.

The technological basis of modern digital linguistics includes the collection and integration of large arrays of linguistic data, their distributed storage and processing, intellectual analysis and visualization of results. Of particular importance are natural language processing, machine learning, and artificial intelligence technologies used for text classification, clustering, entity extraction, topic modeling, sentiment analysis, and semantic relationship discovery. Visual analytics allows us to represent language processes using graphs, networks, and semantic models. Thus, Big Data forms a complex technological ecosystem that combines

PHILOLOGY AND LINGUISTICS

the collection, storage, processing, modeling, visualization, and interpretation of language information.

Big Data and the development of digital philology.

The integration of Big Data contributes to the transition of digital philology to a large-scale and technologically advanced study of language. Their application is promising for corpus and computational linguistics, sociolinguistics, discourse analysis, stylistics, terminology, lexicography and translation studies. Big data allows us to study the dynamics of vocabulary, the spread of neologisms, the change in word meanings, the transformation of discourses and the peculiarities of the linguistic behavior of various social groups. Dynamic language analysis is especially promising, which allows not only to record changes, but also to identify emerging trends. In the future, this creates opportunities for building intelligent language platforms that combine corpora, artificial intelligence, machine learning, semantic analysis and visualization.

Big Data in the conditions of the SHIVA-concept of the external environment.

The SHIVA-concept focuses on the high speed of change, uncertainty, interconnectedness of processes and the need for constant adaptation. For linguistics, this means the need to work with constantly updated flows of linguistic information. Big Data creates technological prerequisites for the formation of dynamic corpora and practically continuous monitoring of linguistic processes. The detection of weak signals becomes especially important: changes in word frequency, the emergence of new concepts or transformations of discursive connections can indicate larger-scale changes in the social environment. Therefore, Big Data in the SHIVA environment performs not only a descriptive, but also a diagnostic and prognostic function.

Big Data in the context of the WARMI concept and systemic risks.

The WARMI concept increases attention to the complexity of the environment, the interconnectedness of processes and the multiplicity of risks. Linguistic information under such conditions becomes part of a broader system of socio-informational interactions and can reflect social tensions, value transformations and changes in communication models. Big Data allows you to compare linguistic data with social, economic, cultural and technological factors, which creates the basis for interdisciplinary analysis. However, digital

PHILOLOGY AND LINGUISTICS

data is not completely neutral: it depends on the characteristics of platforms, algorithms, user behavior, and the structure of information flows. Therefore, a statistically revealed pattern requires mandatory philological and contextual interpretation.

Methodological limitations and risks of using Big Data.

One of the main problems is the quality of the source data. Unrepresentative, incomplete or biased arrays can lead to erroneous conclusions even when using perfect algorithms. Another significant limitation is the algorithmic dependence of the results, especially when analyzing ambiguity, irony, metaphors and culturally specific expressions. Issues of confidentiality, ethics and protection of personal data are important, since the analysis of digital communication can affect the private lives of users. In addition, there is a risk of technological reductionism, in which complex humanitarian phenomena are reduced to quantitative indicators. Therefore, Big Data should be used as a tool to expand philological analysis, and not as its complete replacement.

Prospects for the development of Big Data technologies in modern linguistics.

The promising development of Big Data is associated with the formation of intelligent self-learning systems that combine big data, machine learning, artificial intelligence, generative language models, graph technologies and semantic networks. Dynamic digital corpora are especially promising, capable of automatically updating and recording language changes in almost real time. Multimodal analysis, combining text, speech, images and video, also has significant potential. In the conditions of SHIVA and WARMI environments, the importance of adaptive analytical systems that can change analysis parameters following data transformation is increasing. Such systems can be used for early detection of language trends, discursive changes and information risks.

CONCLUSIONS.

Big Data technologies are becoming one of the key factors in the transformation of modern linguistics and digital philology, as they provide the opportunity to analyze large-scale, dynamic, heterogeneous and multimodal arrays of linguistic information. Their use makes it possible to significantly expand the empirical base of philological research, to move from limited samples to a large-scale

PHILOLOGY AND LINGUISTICS

analysis of the linguistic environment, and to reveal regularities that are difficult to detect using traditional methods.

The specificity of linguistic Big Data is determined by the complexity of the language itself: its contextuality, ambiguity, variability, cultural conditioning and constant variability. Therefore, the application of Big Data requires a combination of computational methods with traditional philological approaches, which provide meaningful interpretation of the obtained results.

In terms of SHIVA and WARMI concepts, Big Data acquires additional methodological significance. They make it possible to consider language as a dynamic indicator of environmental changes, detect weak signals of the transformation of social discourse, track accelerating linguistic processes, and form prerequisites for forecasting new trends.

At the same time, the expansion of technological capabilities is accompanied by the growth of methodological, ethical and informational risks. Unrepresentativeness of data, algorithmic distortions, privacy issues and the risk of excessive quantitative reduction of humanitarian knowledge require a critical and systematic approach to the use of Big Data.

Thus, the most promising is the hybrid paradigm of modern linguistics, in which Big Data, artificial intelligence, machine learning and intelligent analytics are integrated with fundamental methods of philological research. In the future, this may lead to the formation of an adaptive digital philology capable not only of analyzing already accumulated language information, but also of promptly monitoring, explaining and forecasting language transformation in conditions of high uncertainty, turbulence and systemic risks of the modern external environment.

References:

- [1] Nevmerzhytska S. M. (2018). Formation of a strategy for the innovative development of enterprises in conditions of uncertainty. *Scientific Bulletin of the Kherson State University. Series: Economic Sciences*. 2018. Vol. 32. pp. 99–103. URL: <https://ejournal.kspu.edu/index.php/ej/article/view/422/418>.
- [2] Krasnyuk, M., Kulynych, Y., Tuhaienko, V., & Krasniuk, S. (2022). E-business and ecommerce technologies as an important factor for economic efficiency and stability in the modern conditions of the digital economy. *Grail of Science*, 2022, №. 17, pp. 69–81. <https://archive.journal-grail.science/index.php/2710-3056/article/view/412>
- [3] Derbentsev, V. D., B. O. Tishkov, O. D. Sharapov. Systematic methodology for studying the dynamics of the current information economy in the minds of increasing instability. *Modeling and information systems in economics*. – 2013. – Vol. 89. – pp.

PHILOLOGY AND LINGUISTICS

- 47–62 [In Ukrainian].
- [4] Baayen, R. H. (2024). The wompom. *Corpus Linguistics and Linguistic Theory*, 20(3), 615–648. <https://doi.org/10.1515/cllt-2024-0053>
 - [5] Blasi, D. E., & Roberts, S. G. (2018). Studying language evolution in the age of big data. *Journal of Language Evolution*, 3(2), 94–129. <https://doi.org/10.1093/jole/lzy004>
 - [6] Krasnyuk, M., Nevmerzhytska, S., & Tsalko, T. (2024). Processing, analysis & analytics of big data for the innovative management. *Grail of Science*, (38), 75–83. <https://doi.org/10.36074/grail-of-science.12.04.2024.011>
 - [7] Goncharenko S. (2025). BIG Data in modern literary studies. Science: Development and Factors its Influence : Proceedings of the 6th International Scientific and Practical Conference (October 6–8, 2025; Amsterdam, Netherlands). – Amsterdam: Scientific Collection «InterConf», 2025. – P. 82–85.
 - [8] Lee, H.-Y., Li, S.-W., & Vu, N. T. (2022). Meta Learning for Natural Language Processing: A Survey. arXiv preprint arXiv:2205.01500. <https://arxiv.org/abs/2205.01500>
 - [9] Goncharenko S. (2025). Big semi-structured data & deep ANN in computational linguistics. Science and Global Challenges in the Modern World : Proceedings of the 2nd International Scientific Conference (Leicester, United Kingdom, 5 October 2025). – Lulu Press, Inc., 2025. – pp. 133–136.
 - [10] Kaplan, F. (2015). A map for Big Data research in digital humanities. *Frontiers in Digital Humanities*, 2, Article 1. <https://doi.org/10.3389/fdigh.2015.00001>
 - [11] Ganascia, J.-G. (2015). The logic of the Big Data turn in digital literary studies. *Frontiers in Digital Humanities*, 2, Article 7. <https://doi.org/10.3389/fdigh.2015.00007>