

Big Data Technologies in Modern Linguistic Tectology: Managing Systemic Risks and Environmental Turbulence

Svitlana Goncharenko

Kyiv National University of Technologies and Design, Kyiv

<https://orcid.org/0000-0002-7740-4658>

Abstract. *The article explores the specifics, features and prospects for the development of Big Data technologies in modern linguistics in the context of the dynamics of uncertainty, turbulence and systemic risks. The possibilities of big data for corpus, sociolinguistic, discursive and intellectual analysis of linguistic information are revealed. The key problems of representativeness, data quality, algorithmic distortions and ethical limitations are identified. The prospects for integrating Big Data with artificial intelligence, machine learning and predictive analytics are substantiated.*

Keywords: *modern linguistics, digital linguistics, Big Data, language data analysis, natural language processing, machine learning, artificial intelligence, data mining, digital communication, language technologies, uncertainty, turbulence, systemic risks.*

INTRODUCTION.

Modern linguistics is developing in conditions of rapid complication of the information environment, growth of volumes of digital text data and acceleration of processes of their generation, distribution and transformation [1]. Language communication is increasingly moving into digital space, forming huge arrays of texts, language data, multimodal messages, metadata and communication traces of users [2]. At the same time, the external environment of the functioning of science is characterized by dynamics of uncertainty, turbulence and systemic risks [3], which significantly increases the requirements for speed, scalability, adaptability and reliability of linguistic analysis methods. In these conditions, Big Data technologies become not only a tool for processing large arrays of language information, but also an important technological basis for the transformation of modern linguistics [4]. Traditional methods of philological and linguistic research, focused mainly on limited in volume and relatively stable samples, are insufficient for the analysis of modern digital language reality [5]. In such complex external conditions (in any field of science), the following always become of great importance: automated collection of streaming and batch big data, their preprocessing, distributed storage, exploratory visualization (EDA), statistical analysis, descriptive data mining [6], predicative modeling [7], and, finally, prescriptive modeling. But, it is their synergistic integration that allows us to study the object of research (in our case, language) as a dynamic, multidimensional system that is constantly changing, functioning in conditions of high variability and information uncertainty [8]. Big Data becomes especially relevant in

conditions of complex dynamics of the modern communication environment [9]. Language data is characterized by a high update rate, heterogeneity of sources, differences in formats, significant volumes, semantic variability, and the presence of noise information [10]. In addition, digital communication generates new types of language objects: social media messages, comments, digital discussions, user content, electronic documents, automatic speech transcriptions and multimodal communication complexes. This requires the formation of new methodological approaches to their collection, cleaning, structuring, interpretation and verification.

That is why the main goal of this article is to systematically determine the specifics, features and directions of development of Big Data technologies in modern linguistics, taking into account the complex conditions of the dynamics of uncertainty, turbulence and systemic risks. To achieve this goal, it is necessary to reveal the role of Big Data in the digital transformation of linguistics; determine the main properties and features of big language data; characterize the areas of their application; identify key problems and risks of Big Data linguistics; determine the prospects for the development of relevant technologies

RESULTS.

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

Big Data technologies represent a set of methods and tools for collecting, storing, processing and intelligently analyzing large-scale, high-speed and diverse data sets. In modern linguistics, their importance is determined by the rapid growth of digital text, language and multimodal data.

The use of Big Data allows us to move from the analysis of limited samples to the study of large-scale and dynamic language sets. This contributes to the formation of a computational, corpus and intellectual-analytical model of linguistic research, expanding the capabilities of traditional linguistic interpretation.

The specificity of big language data.

Big language data is characterized by its scale, high update rate, variety of formats and sources, contextuality and dynamism. Digital sets simultaneously contain scientific and literary texts, news, social networks, comments, advertising messages, dialogues and other forms of communication.

The specificity of Big Data linguistics also lies in the need to take into account the context and temporal variability of language. Therefore, quantitative data processing should be combined with their linguistic interpretation, verification and assessment of representativeness.

Main areas of application of Big Data in modern linguistics.

One of the key areas is automated corpus analysis, which allows studying the frequency, connectivity and contextual characteristics of language units. Significant opportunities are provided by the use of machine learning and natural language

processing technologies for text classification, thematic modeling, tone analysis, information retrieval and recognition of named entities.

Big Data expands the capabilities of sociolinguistics, allowing to study the linguistic variability of various social and communicative groups, as well as the dynamics of discourses and the spread of linguistic innovations. Multimodal analysis, combining text, language, visual and metadata, becomes promising.

Big Data in conditions of uncertainty and turbulence.

In the unstable digital environment, language data is constantly changing, new sources appear and communicative practices are transformed. Therefore, Big Data technologies must ensure not only scalability, but also adaptability.

Dynamic corpora, streaming processing and operational monitoring of language processes are of particular importance. Integration of Big Data with machine learning allows us to move from retrospective analysis to identifying current trends and predicting possible directions of language changes.

Systemic risks and limitations of Big Data linguistics.

A large amount of data does not guarantee the reliability of the results. Digital arrays may contain errors, duplicates, spam, automatic content and statistical outliers. Representativeness also remains a significant limitation, since digital sources reflect only a certain part of the language community.

Additional risks are associated with algorithmic distortions that arise due to the peculiarities of the source data and machine learning models. It is also necessary to take into account the issues of confidentiality, copyright and ethics of the use of user content. Therefore, the key principles of Big Data linguistics should be the quality, representativeness, verifiability and interpretability of data.

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

The further development of Big Data linguistics is associated with the integration of big data with artificial intelligence, machine learning, natural language processing, multimodal analytics and modern language models.

Hybrid technologies that combine automated processing of large arrays with expert linguistic interpretation, as well as adaptive and self-learning systems that can take into account language changes, are especially promising. Predictive linguistic analytics has significant potential, allowing not only to record, but also to predict language transformations.

As a result, Big Data is gradually transforming from a text processing tool into a systemic technological basis of modern digital linguistics.

CONCLUSIONS.

Big Data technologies form fundamentally new opportunities for the development of modern linguistics, ensuring the transition from the analysis of limited and stable samples to the study of large-scale, dynamic and multidimensional arrays of linguistic

information. Their specificity is determined by the combination of large volume, high update rate, diversity, contextuality and dynamism of linguistic data.

It has been established that in the conditions of uncertainty, turbulence and systemic risks, scalability, adaptability, efficiency and intelligence of Big Data technologies acquire special importance. They allow monitoring of linguistic processes, studying the dynamics of discourses, identifying linguistic innovations, conducting sociolinguistic analysis and forming predictive models.

At the same time, the use of Big Data requires a critical attitude to the quality and representativeness of data. The main system limitations are algorithmic distortions, incompleteness of the digital representation of the linguistic community, data noise, problems of interpretability, confidentiality and ethics.

The prospects of Big Data linguistics are associated with the integration of big data with artificial intelligence, machine learning, natural language processing, multimodal analytics and intelligent modeling. The most promising is the hybrid model, in which the computing power of Big Data is combined with the theoretical and expert interpretation of a linguist.

Thus, the development of Big Data technologies contributes to the formation of a new digital research paradigm in linguistics, focused not only on large-scale analysis of existing language material, but also on the operational detection, modeling and prediction of language transformations in a constantly changing and high-risk information and communication environment.

References

1. 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>
2. 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.
3. Краснюк Світлана (2024). Data Science у освітньому менеджменті. *Діалог культур у Європейському освітньому просторі: Матеріали IV Міжнародної конференції*, м. Київ, 10 травня 2024р. Київський національний університет технологій та дизайну. – К. : КНУТД, 2024. – С. 119-124.
4. 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>
5. Baayen, R. H. (2024). The wompom. *Corpus Linguistics and Linguistic Theory*, 20(3), 615–648. <https://doi.org/10.1515/cllt-2024-0053>
6. Krasnyuk M, Elishis D (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. *Наука і Техніка Сьогодні*, 4(32). [https://doi.org/10.52058/2786-6025-2024-4\(32\)-833-857](https://doi.org/10.52058/2786-6025-2024-4(32)-833-857)

7. 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>
8. 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].
9. 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>
10. 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>

UDC 81'322:004.8

Features of Hybrid Machine Learning in Modern Computational Linguistics: Navigating Systemic Risks, Turbulence and Uncertainty Through the Lens of SHIVA and WARMI Paradigms

Svitlana Krasniuk

Kyiv National University of Technologies and Design, Kyiv

<https://orcid.org/0000-0002-5987-8681>

Abstract. *The article explores the features and specifics of hybrid machine learning in modern computational linguistics in conditions of dynamic uncertainty, turbulence and systemic risks. The main directions of hybridization of neural, symbolic, statistical and linguistic methods are considered, as well as the possibilities of integrating large language models, knowledge graphs and adaptive learning mechanisms. The prospects of hybrid architectures for increasing the accuracy, stability, adaptability and comprehensibility of intelligent linguistic systems in conditions of SHIVA and WARMI paradigms of the external environment are substantiated.*

Keywords: *Digital Philology; Machine Linguistics; Hybrid machine learning; Computational linguistics; Artificial intelligence; Natural language processing; Large language models; Neurosymbolic technologies; Knowledge graphs; Adaptive learning; Uncertainty; Turbulence; Systemic risks; SHIVA; WARMI.*

INTRODUCTION. Modern philology is developing in conditions of qualitatively more complicated external environment [1], which is characterized by acceleration of changes, growth of uncertainty [2], turbulence of information flows, increase in volume and heterogeneity of language data [3], as well as strengthening of