

## Specifics and Adapting Big Data Technologies for Modern Translation Studies: Navigating Turbulence, Uncertainty and Systemic Risk (Through the Lens of SHIVA-Linguistics)

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 adapting Big Data technologies to modern translation studies in conditions of turbulence, uncertainty and systemic risks. The potential of Big Data for the analysis of large language arrays, translation corpora and the dynamics of translation practices within the framework of SHIVA-linguistics is substantiated. The key directions, opportunities, limitations and prospects of the integration of Big Data, artificial intelligence and expert translation analytics are identified.*

**Keywords:** *digital philology; machine linguistics; big data, translation studies, SHIVA linguistics, digital transformation, artificial intelligence, machine learning, language corpora, machine translation, translation analytics, turbulence, uncertainty, systemic risks.*

### Introduction.

Modern humanities are developing in complex and complex conditions [1, 2]: turbulence, uncertainty [3], accelerated digitalization and systemic risks. In these conditions, Big Data technologies are becoming an important tool for processing and intelligent analysis of large-scale humanitarian data [4]. Their adaptation requires taking into account contextuality, semantic ambiguity and dynamism of language processes [5], which is especially relevant for SHIVA-linguistics, which considers language as a complex and adaptive system.

Modern translation studies are developing in conditions of rapid complication of the information environment [6], which is characterized by high dynamics of turbulence and uncertainty, acceleration of digital transformation, exponential growth of data volumes and increased systemic risks [7]. For translation activities, this means a transition from a predominantly text-centric research model to a more complex digital paradigm, in which the object of analysis is not only a single text, but also large sets of diverse linguistic data [8], formed in various communicative, social, professional and technological environments [9].

In these conditions, Big Data technologies [10] acquire particular importance, providing the collection, storage, integration, processing and intellectual analysis of large-scale arrays of textual, multimodal and metalinguistic information. Their application allows us to significantly expand the analytical capabilities of translation studies [11]: to identify stable patterns of translation transformations, to investigate the dynamics of lexical-grammatical and stylistic decisions, to compare original and translated texts, to model translation strategies and to predict the characteristics of future translation decisions [12].

That is why the relevance of adapting Big Data to modern translation studies is further enhanced by the formation of a new research environment, which should be

considered within the framework of SHIVA-linguistics - a conceptual direction focused on the study of language and communication in conditions of high variability, uncertainty, nonlinearity, hybridization, and volatility of information. In such an environment, translation appears not as a static act of interlingual transformation, but as a dynamic, probabilistic, and context-dependent process that occurs within a complex digital ecosystem.

Therefore, the purpose of this publication is to systematically identify the features and specifics of adapting Big Data technologies to modern translation studies in conditions of turbulence, uncertainty, and systemic risks, as well as to reveal their methodological potential within the framework of SHIVA-linguistics.

### **Results.**

The digitization of translation activities is transforming translation studies from a predominantly text-centric analysis to the study of large-scale arrays of linguistic and translation data. Big Data allows working with parallel corpora, translation memories, terminological resources, digital archives and automatic translation results, revealing patterns that are difficult to detect in limited samples. Therefore, the most promising is a hybrid methodology that combines Big Data, corpus analysis, machine learning and expert interpretation.

Adapting Big Data requires taking into account the specifics of linguistic data: contextuality, ambiguity, variability and cultural conditioning. The classic characteristics of Big Data - Volume, Velocity, Variety, Veracity, Value - in translation studies should be supplemented with the parameters Context, Volatility and Risk, which reflect the contextuality of linguistic units, the dynamics of the linguistic environment and the consequences of possible errors.

The key directions are: creation of large-scale parallel and comparative corpora; automated detection of translation patterns and transformations; dynamic terminological analysis; assessment of the quality of machine translation; research of changes in language and translation practices.

Thus, Big Data allows us to move from the analysis of individual translation decisions to the identification of stable patterns only at the level of large language arrays.

The integration of Big Data with machine learning and artificial intelligence creates the basis for automatic terminology extraction, detection of semantic correspondences, text classification, analysis of translation errors and prediction of translation decisions. However, the effectiveness of such systems directly depends on the quality and representativeness of the data. Incomplete, erroneous or culturally unbalanced corpora are capable of reproducing algorithmic distortions. Therefore, technological development should be accompanied by linguistic expertise.

In terms of SHIVA-linguistics, language is considered as a dynamic, open and adaptive system that functions in an environment of high variability and uncertainty. Big Data allows you to record the emergence of new terms, hybrid language forms and digital communicative practices, tracking their development practically in real time.

At the same time, turbulence, uncertainty and systemic risks require dynamic adaptability of analytical systems: they must not only process accumulated data, but also constantly take into account changes in the language environment.

The limitations of Big Data in translation studies include low quality or unrepresentativeness of data, noise, duplication, language and genre imbalance, algorithmic bias and limited interpretability of results. Additional risks are associated with confidentiality, copyright, information security and the possibility of uncritical trust in automated solutions.

Therefore, an effective model should be based on the principle of "data + algorithm + expert interpretation", and not on automation as an end in itself.

A promising architecture can be described as a cyclical process: data collection → cleaning and structuring → annotation → storage → intelligence analysis → pattern recognition → expert interpretation → translation decision → feedback and data update.

Unlike a linear model, such a system is constantly updated with new translation data. For SHIVA-linguistics, its adaptability, scalability, contextuality and risk-orientation are especially important. It is the combination of Big Data, artificial intelligence and human expertise that forms the basis of sustainable and intellectually oriented modern translation studies.

### **Conclusions.**

The analysis shows that Big Data technologies are becoming one of the most promising tools for the transformation of modern translation studies. Their use ensures the transition from the analysis of limited text samples to the systematic study of large-scale, dynamic and diverse language arrays.

The specifics of the adaptation of Big Data to translation studies are determined by the need to take into account the contextuality, variability, cultural conditioning and semantic ambiguity of language data. Therefore, the traditional Big Data model should be supplemented with the parameters of contextuality, volatility and risk, which is especially relevant in the conditions of SHIVA-linguistics.

The most promising areas of application of Big Data are the development of parallel corpora, automated analysis of translation transformations, dynamic terminological monitoring, assessment of the quality of machine translation, study of language dynamics and intelligent prediction of translation decisions.

The integration of Big Data with machine learning and artificial intelligence significantly expands the capabilities of translation studies, but at the same time increases the requirements for the quality, representativeness and ethical correctness of data. Automation should not be seen as a replacement for professional translation expertise; on the contrary, a hybrid model combining big data, intelligent algorithms and expert interpretation is the most effective.

Within the framework of modern SHIVA linguistics, Big Data is intended to be considered as the basis of an adaptive research ecosystem capable of functioning in conditions of turbulence, uncertainty and systemic risks. Prospects for further research are related to the development of intelligent Big Data platforms for dynamic translation monitoring, the creation of risk-oriented models of translation data analysis, the development of multimodal corpora and the formation of hybrid Big Data + AI + Human Expertise systems that ensure the stability, adaptability and quality of modern translation.

## References

1. Zhang, Z., Song, W., & Liu, P. (2024). Making and interpreting: Digital humanities as embodied action. *Humanities and Social Sciences Communications*, 11, Article 13. <https://doi.org/10.1057/s41599-023-02548-3>
2. De Luca, E. W., Fallucchi, F., Ghattas, B., & Spielhaus, R. (2024). The digital transformation processes for supporting digital humanities researchers in text analysis. *Journal of Documentation*, 80(2), 378–391. <https://doi.org/10.1108/JD-07-2022-0143>
3. 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://ej.journal.kspu.edu/index.php/ej/article/view/422/418>
4. 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>
5. 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>
6. 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.
7. 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>
8. 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.
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. Krasnyuk, M., Elishis, D. (2024). Perspectives and problems of big data analysis & analytics for effective marketing of tourism industry. *Science and Technology Today*, 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)
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>
12. Baayen, R. H. (2024). The wompom. *Corpus Linguistics and Linguistic Theory*, 20(3), 615–648. <https://doi.org/10.1515/cllt-2024-0053>