

eLex 2025: Intelligent Lexicography Between Automation and Linguistic Responsibility

ABSTRACT: This paper provides an overview of the international conference *Electronic Lexicography in the 21st Century (eLex 2025)*, whose main theme was Intelligent Lexicography. The conference was held from November 18 to 20, 2025, in Bled, Slovenia, and was dedicated to contemporary trends in electronic lexicography, with particular attention to the application of artificial intelligence and large language models in the creation, enrichment, and evaluation of lexicographic resources. The paper examines the dominant research trends, methodological challenges, and open questions that marked the conference, as well as the positioning of Serbian lexicography in that international context. A special section is devoted to the contributions of researchers from Serbia and their experiences in working on the *Dictionary of Contemporary Serbian Language* and in experiments with large language models in the field of phraseology and multiword lexical units.

KEYWORDS: electronic lexicography, intelligent lexicography, digital humanities, large language models, multiword lexical units, eLex 2025, Serbian language

PAPER SUBMITTED: 18 December 2025

PAPER ACCEPTED: 30 December 2025

Ranka Stanković

ranka@rgf.rs

ORCID: 0000-0001-5123-6273

*University of Belgrade
Faculty of Mining and Geology
Belgrade, Serbia*

Aleksandra Marković

malexa39@gmail.com

ORCID: 0000-0003-2985-3042

*Serbian Academy of Sciences and
Art
Institute for the Serbian Language
Belgrade, Serbia*

Translated from Serbian:
Mina Dizdar

1 eLex as a Reference Point of Modern Lexicography

The *Electronic Lexicography in the 21st Century (eLex)* conference series has for more than a decade represented one of the key international venues for the exchange of ideas in the fields of lexicography, corpus linguistics, language technologies, and digital humanities. The *eLex 2025* conference, held

in Bled, is thematically explicitly focused on the concept of intelligent lexicography, thereby emphasizing the need to view contemporary dictionaries not only as digitized products of traditional lexicography but as dynamic, technologically supported knowledge systems.

The Conference proceeding (Kosem et al. 2025), published in open access, testifies to the exceptional thematic diversity and methodological breadth of the contemporary lexicographic community. It includes papers dealing with theoretical questions of sense disambiguation and dictionary entry structure, as well as research oriented toward practical solutions, automation, and evaluation of tools based on artificial intelligence methods.

2 Program Structure of the Conference

The program of the *eLex 2025* conference was carefully structured to enable a balanced dialogue between guest lectures, thematic workshops, poster and demo sessions, as well as working discussions. During the three conference days, presentations were organized at the same time across multiple rooms, covering a wide spectrum of topics from modern electronic and intelligent lexicography. In addition, the conference was preceded by the *CLASSLA-express 2.0 workshop*, held on November 17, and after the closing, on November 20, the post-conference *GWLN-7 Workshop* was held.

The first workshop, with the subtitle *Corpora vs. LLMs: a hands-on workshop on using CLASSLA and AI tools for phraseology and lexicography in South Slavic languages*, was conceived as one in a series of workshops aimed at combining the capabilities provided by corpus search tools with those offered by artificial intelligence tools, as well as examining the capabilities of corpora and large language models (hereinafter LLMs) in performing various tasks relevant to lexicography and linguistics. The workshop examined this capability in relation to multiword lexical units (hereinafter MLUs) in South Slavic languages, mostly with insufficiently developed language resources and technologies.

The second workshop, with the subtitle *Globalex Workshop on Lexicography and Neology*, is part of the GWLN workshop series and presents methods and techniques for finding, researching, classifying, and describing neologisms in the context of their lexicographic processing and approaches in different languages.

A special place in the program was occupied by three guest lectures, which provided a theoretical and methodological framework for further discussion. The guest speakers addressed the role of LLMs in lexicography, in-

stitutional experiences in their application, as well as fundamental questions of data structure in contemporary dictionaries. These lectures clearly indicated a shift in focus from mere digitization toward a deeper understanding of the relationship between linguistic knowledge, data, and algorithms.

The first guest lecture was delivered by Carole Tiberius & Jesse de Does, on the topic: *LLMs and lexicography at the Dutch language Institute* (both authors list this institute as their affiliation). The authors presented the work at the institute, which has a long tradition of creating dictionaries, as well as dictionary databases for the Dutch language (and some other related languages). Lexicographic work is computer-supported, but a significant part of the work is still done manually. Currently, research is being conducted on the ways in which new technologies, including LLMs, can be used to improve certain phases of lexicographic work, but in such a way that data quality and reliability are not compromised. The presentation mentioned in which way the possibility of implementing the Pattern Analysis procedure, devised by the renowned linguist Patrick Hanks, is being examined.

Marko Robnik-Šikonja (from the University of Ljubljana) spoke on the topic of *Large Language Models for Lexicography*. The author emphasized that LLMs redefine methodological approaches in many fields, including linguistics and lexicography. The models are pre-trained on large textual corpora using next-token prediction and adapted for interaction with humans. They are not immune to hallucinations, which is why human involvement is still necessary in their application. In the context of lexicography, models can support several tasks. The author indicated how information from a language database can be used to improve LLMs in performing lexicographic tasks. The methodology he presented was based on knowledge graph extraction, continual pre-training of LLMs, prompt design, and semi-automatic evaluation.

The last guest lecture was delivered by Michal Měchura (Lexical Computing & Dublin City University), the author of the Lexonomy dictionary writing system and DMLex, a modern data model for lexicography. In the lecture on the topic: *We need to talk about data structures in Lexicography*, the author said that it was time for a critical review of the data models used to represent dictionary structure in dictionary writing systems and other softwares. He briefly presented the history of lexicographic data modeling and emphasized that regardless of the notation used (XML, JSON, etc.), the underlying pattern is fundamentally a tree structure, in which different structural units (entries, meanings, definitions) are organized in a parent-child hierarchy. The author believes that the tree pattern is not sufficiently

expressive for all phenomena in dictionaries (cross-referencing, placement of multiword sub-entries, complex hierarchy of sub-meanings) and that this could be more easily represented using a graph-based data structure (such as relational databases or knowledge graphs based on the semantic web style). Měchura believes that dictionary projects that insist only on a tree-based data model do not fully utilize the possibilities of the digital medium, and that the transition to a graph-structured model is not simple because thinking based on tree hierarchy is rooted in the minds of lexicographers and dictionary users. At the end, the author presented the recently standardized data model for lexicography, DMLex, which, being a hybrid model, aims to facilitate the aforementioned transition by combining tree structures with graph structure when necessary.

Thematic sections were organized around key areas: automation of lexicographic processes, application of artificial intelligence and corpus methods, processing of MLUs and phraseologisms, neologisms, as well as user aspects of electronic dictionaries. Such organization enabled related papers to be viewed within a joint methodological framework, while also clearly highlighting differences in approaches and research strategies.

A significant part of the program consisted of poster and demo sessions, which provided insight into specific tools, resources, and experimental systems. They enabled direct exchange of experiences and discussion of practical problems in implementing intelligent lexicographic solutions. In addition, group discussions and workshops further contributed to the interactivity of the conference and the exchange of views on open questions and future development directions.

The overall structure of the *eLex 2025* program reflected a deliberate effort to prevent the conference from becoming merely a sequence of individual presentations, instead positioning it as an integrated forum where theoretical discussions, empirical research, and practical demonstrations mutually reinforced each other.

3 Intelligent Lexicography as a Research Framework

One of the central questions that appears in a large number of papers is: what actually makes lexicography “intelligent”? In discussions at the conference, this concept is not reduced to mere use of artificial intelligence, but implies a thoughtful integration of algorithmic methods into lexicographic processes, while preserving the theoretical foundation and linguistic responsibility.

LLMs appear as a powerful but ambivalent tool. On the one hand, they enable automated generation of definitions, usage examples, and even proposals for dictionary structure. On the other hand, numerous papers show that such models often produce content that is formally acceptable but semantically imprecise, insufficiently differentiated, or insensitive to cultural differences. It is precisely this line between efficiency and reliability that constitutes one of the main thematic threads of the *eLex 2025* conference.

3.1 Multiword Lexical Units as a “Critical Point” of Automation

A special place in the Conference proceeding and conference discussions is dedicated to papers devoted to MLUs (polylexemic units), idioms, and fixed expressions. These linguistic units consistently appear as one of the greatest challenges for automated processing because they require simultaneous understanding of the syntactic structure, semantic non-compositionality, and pragmatic context.

In this sense, *eLex 2025* clearly shows that MLUs are a kind of “litmus test” for intelligent lexicography: it is precisely on them that the limits of statistical and generative approaches are most clearly observed. Several papers emphasize that LLMs can be useful in the phases of pre-selection or assistance, but that their results require careful expert evaluation and correction.

3.2 Automation, Evaluation, and the Question of Control

Another important thematic direction concerns the evaluation of lexicographic tools based on artificial intelligence. Several papers emphasize the need for clear evaluation metrics, gold standards, and reproducibility of results. The variability of LLM outputs, dependence on prompt formulation, and lack of transparency regarding knowledge sources represent a serious challenge for their application in scientifically grounded lexicographic projects.

In this context, the *eLex 2025* conference clearly suggests a shift in focus from the question “Should we use artificial intelligence?” to the question of how to use it responsibly, in a controlled manner, and in accordance with lexicographic principles.

4 COST Actions

Paper (Makhachashvili et al. 2025), within the COST action CA21167 *Universality, diversity and idiosyncrasy in language technology (UniDive)*¹ examine key problems related to the representation of morphosyntactic properties of MLUs in modern lexicographic resources intended for application in natural language processing. Starting from a review analysis of existing lexicons containing syntactic descriptions of MLUs, the paper systematizes current approaches to encoding their internal structure, argument structure, word order, discontinuity, and verb alternations. Special emphasis is placed on the importance of precise syntactic description for various NLP tasks, thereby pointing to the gap between lexicographic theory and the needs of contemporary language technologies. The authors conclude that the improvement and standardization of syntactic description of MLUs represent a key prerequisite for the more efficient integration of lexicographic resources into future NLP systems.

Paper (Kallas et al. 2025) present the results of extensive research conducted within the COST action *European Network on Lexical Innovation (ENEOLI)*², which deals with contemporary practices in studying and documenting lexical innovations. Based on a survey conducted at the end of 2024, in which 249 respondents from 50 countries participated, the paper analyzes approaches to identifying and processing neologisms and new meanings from the perspective of lexicography and terminology. Special emphasis is placed on the composition of research teams, the use of corpora and digital tools, as well as educational and professional needs in this area. The results indicate a change in the profile of knowledge and skills needed to work with lexical innovations, emphasizing the importance of interdisciplinarity and continuous training in modern lexicographic and terminological practice.

Paper (Lindemann and Salgado 2025) present the conceptual foundations and practical implementation of the Wikibase platform developed within the COST action ENEOLI, intended for collaborative work on resources in the field of neology and lexical innovations. The platform integrates three interconnected resources into a unified knowledge graph: a bibliographic and textual repository for research in neology (NeoCorpus), a multilingual metaterminological dictionary for modeling and harmonizing the conceptual apparatus (NeoVoc), and a collection of neologisms documenting new lexical units in various European languages. The paper shows how Wikibase enables in-

1. COST Action CA21167 UNIDIVE

2. COST Action CA22126 ENEOLI

teroperable linking of these resources and their publication as FAIR linked data, with a detailed description of the process and initial results. The authors conclude that such an approach can serve as a reference model for similar collaborative projects based on knowledge graphs in the field of digital lexicography.

Paper (Beliga and Filipović Petrović 2025) present a hybrid methodological framework for identifying phraseological constructions, based on combining LLMs, corpus methods, and expert linguistic evaluation. Starting from a pilot study on Croatian repetitive constructions, the research shows how artificial intelligence can be used to generate candidates, paraphrases, and queries for corpus search, while corpus data are used for empirical confirmation of frequency and usage, with constant supervision by a human expert. The methodology is illustrated by comparing the German construction *X über X* and its Croatian equivalent *X za X-om*, highlighting the differences in the sensitivity of different language models to idiomaticity and formal structure. The paper emphasizes the potential of such layered approach for semi-automated processing of phraseology and the development of multilingual phraseological resources, but also clearly indicates limitations and the need for careful prompt design and expert control of results. The research is related to the results of the COST action *PhraConRep: A Multilingual Repository of Phraseme Constructions in Central and Eastern European Languages* (CA22115)³.

5 Serbian Lexicography in the International Context

Part of the conference program was marked by the contributions from researchers from Serbia, who presented the results of work on contemporary lexicographic resources and experiments with the application of LLMs. Work on the *Dictionary of Contemporary Serbian Language* (RSSJ) was presented as an example of a systematic lexicographic project facing the challenges of automation and digital transformation (Stanković et al. 2025). The digital lexicographic resource RSSJ is intended both for human users, via web and mobile apps, and for machines, through programming interfaces (APIs). The project, coordinated by the diaspora association “Okupljeni oko jezika” (Gathered around the language) and the Society for Language Resources and Technologies (JeRTeh), aims to create a dictionary of about 50,000 frequent lexical units that reflect language use over the past fifty years in various

3. COST Action CA22115, PhraConRep

functional styles. The list of entries (lemmas) was automatically extracted from the Serbian language corpus SrpKor and then manually corrected. The skeleton of dictionary entries was automatically created and enriched with data from the LeXimirka lexical database. Special attention was paid to the automation of the lexicographic process, including the application of static vector representations and large language models for identifying synonyms and creating initial versions of definitions, as well as automatic extraction of frequent word combinations, syntactic patterns, and short illustrative examples. The paper shows that such approaches can accelerate certain phases of dictionary creation, but simultaneously impose the need to adapt lexicographic workflows and redefine traditional lexicographic practices.

Paper (Marković and Stanković 2025) examine the possibilities and limitations of applying LLMs in modern descriptive lexicography, starting from the concept of post-editing lexicography, in which the lexicographer assumes the role of editor of automatically generated dictionary drafts. Based on the experiment conducted on a set of nearly 400 MLUs with idiomatized meaning, the paper examines the usability of various LLMs in tasks of idiom identification, meaning disambiguation, definitions writing, and generating illustrative examples for the Serbian language. Special emphasis is placed on the difference between examples that illustrate the idiomatized meaning of an expression and those in which the given meaning is paraphrased, as well as on comparing the performance of several contemporary models depending on prompt formulation and access method (interactive chat versus API). The results indicate that, despite significant progress, the application of large language models in lexicography still requires careful evaluation and strong expert intervention, especially in processing semantically and pragmatically complex linguistic units.

The presented papers showed that Serbian lexicography is actively involved in contemporary international trends, but also maintains a critical attitude toward the uncritical application of new technologies.

6 Concluding Remarks

The *eLex 2025* conference clearly shows that modern lexicography is at a turning point. Artificial intelligence and large language models offer powerful tools for accelerating and enriching lexicographic work, but simultaneously open a series of theoretical, methodological, and ethical questions. “Smart dictionaries” do not emerge automatically through the application of intel-

ligent algorithms; they require thoughtful design, clear standards, and the constant presence of human expertise.

In this sense, intelligent lexicography, as presented at the *eLex 2025* conference, does not mark the end of traditional lexicography, but rather its transformation within the framework of digital humanities, a discipline that connects technology, language, and humanistic knowledge in a new way. In this process, the role of the lexicographer remains crucial and irreplaceable: without their theoretical knowledge, interpretive knowledge, and responsibility in decision-making, lexicography is reduced to technical language processing, rather than its systematic and reliable description. It is precisely human expertise that ensures that automation serves lexicography, rather than replacing it.

Acknowledgment

This work was supported by the Ministry of Science, Republic of Serbia #GRANT 451 03-136/2025-03/200174, and the Science Fund of the Republic of Serbia (#7276, Text Embeddings– Serbian Language Applications, TESLA); COST ACTION CA21167 Universality, Diversity, and Idiosyncrasy in Language Technology (UniDive)

References

- Beliga, Slobodan, and Ivana Filipović Petrović. 2025. “AI-and Corpus-Based Strategies for Identifying Phraseme Constructions: A Pilot Study on Croatian Repetitive Constructions.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 95–115. Brno: Lexical Computing CZ sro.
- Kallas, Jelena, Kristina Koppel, Kris Heylen, Ilan Kernerman, Ana Ostroški Anić, Federica Vezzani, and Špela Arhar Holdt. 2025. “Neology in Practice: Lexicographic and Terminological Approaches to Lexical Innovation.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 652–672. Brno: Lexical Computing CZ sro.
- Kosem, Iztok, Miloš Jakubiček, Marek Medved', Karolina Zgaga, Špela Arhar Holdt, Tina Munda, and Ana Salgado, eds. 2025. *Electronic lexicography in the 21st century (eLex 2025): Intelligent lexicography. Proceedings of the eLex 2025 conference. Bled, 18–20 November 2025*. Bled: Lexical Computing CZ s.r.o. <https://elex.link/elex2025/proceedings/>.

- Lindemann, David, and Ana Salgado. 2025. “ENEOLI Wikibase: A collaborative working platform for the European Network on Lexical Innovation.” In *Electronic lexicography in the 21st century (eLex 2023): Intelligent lexicography. Proceedings of the eLex 2025 conference*, 285–296. Bled: Lexical Computing CZ s.r.o. <https://doi.org/10.5281/zenodo.17583262>.
- Makhachashvili, Rusudan, Verginica Mititelu, Voula Giouli, Gražina Korvel, Chaya Liebeskind, Irina Lobzhanidze, Stella Markantonatou, Alexandra Markovic, and Ivelina Stoyanova. 2025. “The Challenges of Syntactic Descriptions of Multiword Expressions in Electronic Lexicography.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 252–269. Brno: Lexical Computing CZ sro.
- Marković, Aleksandra, and Ranka Stanković. 2025. “So Close but Still Far: Case Study on Application of LLMs in Idioms Identification, Definition, and Generation of Illustrative Examples.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 79–94. Brno: Lexical Computing CZ sro.
- Stanković, Ranka, Rada Stijović, Mihailo Škorić, and Cvetana Krstev. 2025. “The Dictionary of Contemporary Serbian Language (RSSJ): Advanced Automation and Other Challenges.” In *Electronic lexicography in the 21st century (eLex 2025) Intelligent Lexicography*, 62–78. Brno: Lexical Computing CZ sro.