

PECULIARITIES OF CREATING COMPUTER LEXICOGRAPHY

Gulira Jumalieva

Institute of Information Technologies,
I. Razzakov Kyrgyz State Technical University
66, Ch. Aitmatov avenue, Bishkek, 720044, Kyrgyzstan

E-mail address: gulira1972@gmail.com

ORCID: <https://orcid.org/0000-0002-1320-8981>

Raushan Bektenova

State and Foreign Languages Department,
Diplomatic Academy of the Ministry of Foreign Affairs of the Kyrgyz Republic
36, Erkindik avenue, Bishkek, Kyrgyzstan

E-mail address: raushana_786@mail.ru

ORCID: <https://orcid.org/0009-0001-8552-8941>

Alena Hašková

Department of Technology and Information Technologies,
Faculty of Education, Constantine the Philosopher University in Nitra
Dražovská 4, 949 01 Nitra, Slovakia

E-mail address: ahaskova@ukf.sk

ORCID: <https://orcid.org/0000-0001-8592-7451>

Zdenka Gadušová

Department of English and American Studies,
Faculty of Arts Constantine the Philosopher University in Nitra
Štefánikova 67, 949 01 Nitra, Slovakia

E-mail address: zgadusova@ukf.sk

ORCID: <https://orcid.org/0000-0002-8954-659X>

Gulnara Kabaeva

Institute of Information Technologies,
I. Razzakov Kyrgyz State Technical University
66, Ch. Aitmatov avenue, Bishkek, 720044, Kyrgyzstan

E-mail address: kabaevagd9@kstu.kg

ORCID: <https://orcid.org/0000-0003-3001-7201>

Aziz Kyrgyzbaev

Doctoral and Post doctorate Department
I. Razzakov Kyrgyz State Technical University
66, Ch. Aitmatov avenue, Bishkek, 720044, Kyrgyzstan
E-mail address: azizkyrgyzbaev@gmail.com
ORCID: <https://orcid.org/0009-0005-5721-2973>

ABSTRACT

Aim. Purpose of the presented research work is to study nuances of creating computer lexicography, taking into account both modern technological advances and linguistic principles.

Methods. To achieve the stated goal, a comprehensive review of existing methodologies and approaches used in creating lexicographic resources was conducted. Additionally, also the problems related to the automation of the process of compiling and updating lexical data were studied. The research methodology included an analysis of modern software tools widely used lexicographic practice, as well as a thorough study of their advantages and limitations. Subsequently a comparative analysis of natural language processing techniques used in the creation of lexicographic databases was conducted.

Results. Within the course of the study key issues were identified such as accuracy, completeness, data relevance, and the ability to automate update processes.

Conclusion. Following the achieved findings, recommendations aimed at optimising the creation and maintenance of lexicographic resources by integrating modern technologies and language processing techniques were formulated. The authors conclude by emphasising the importance of improving the methodology of computer lexicography to effectively provide users with up-to-date and reliable linguistic information under the conditions of rapid development of digital technologies.

Keywords: computer lexicography, electronic dictionaries, linguistic resources, corpus linguistics, lexicographic databases, morphological analysis, semantic markup, interactive dictionaries

INTRODUCTION

In today's data-dominated information society, computer lexicography is becoming a key player. With the rapid development of digital technologies, it occupies a central place in linguistics, facilitating access to lexical resources and processing linguistic data. At the intersection of computer technology and linguistic research, computer lexicography assumes the crucial role of creating, organising and managing lexical

information. Its importance goes beyond simply providing resources; it permeates applications as diverse as machine translation, semantic analysis, natural language processing and the development of intelligent systems. Central to modern language technology, it drives advances in machine translation, facilitates natural language processing, and contributes to the lexicographic resources needed for efficient text retrieval and analysis. The development of computational lexicography not only contributes to the advancement of language technology, but also resonates in various fields, including education, translation, research and cultural exchange. At its core is the need to ensure the accuracy and completeness of lexical data, to optimise information updating processes and to adapt to the dynamically changing linguistic landscape. In light of these challenges, comprehending and exploring the intricacies of computer lexicography becomes an urgent task. Such efforts are aimed at optimising language data processing and providing users with relevant and reliable information in the era of digital transformation (Alam et al., 2024).

With the rapid development of digital technologies and the growing volume of linguistic data, the study of the intricacies of computer lexicography is of paramount importance. The main problems are related to the accuracy and completeness of the data, as well as to the efficiency of automated processes of their compilation and updating. In today's environment, characterised by the development of digital technologies and the growing demand for linguistic resources, the key role of computer lexicography cannot be overestimated. This field, which is concerned with the development of efficient methods for organising and processing lexical information, has a significant impact on modern information societies. Consequently, the study of the nuances of computer lexicography becomes an important component of both scientific research and practical application of linguistic technologies (Zakirova et al., 2022).

The growing amount of data and linguistic information on the Internet emphasises a key point of relevance in contemporary discourse (Kobylarek et al., 2022). As existing methods for researching and creating lexicographic resources attempt to cope with the increasing complexity of language structures, maintaining relevance and validity of data becomes a pressing issue. In fields such as machine translation, natural language processing and intelligent systems development, the efficiency of computer lexicography takes on a key role. An adequate understanding of the process of creating, automating and updating lexicographic resources becomes essential for quality language interaction and textual data processing. At the intersection of modern linguistics and information technology, research in the field of computer lexicography takes centre stage. Previous work in this area has made a significant contribution to our understanding of techniques for creating, organising and processing lexical information using computer technology.

As Mohammed Yousif states (2023), many authors in their works even emphasise the importance of automated methods in compiling lexical data, shedding light on the role of computer lexicography in creating accurate and timely linguistic resources.

Their work also addresses the problems associated with updating data in the face of dynamically changing linguistic structures (Salloum et al., 2020; Mohamed & Oussalah, 2019; Akila & Jayakumar, 2014).

Modern software tools used in computer lexicography are analysed in detail in by Daniel Jurafsky and James Martin (2024). They scrupulously consider all pros and cons of these tools, paying special attention to their contribution to the accuracy and efficiency of work with lexical data, and raise the most important issues of comparative analysis of natural language processing methods used in building lexicographic databases. Attention is paid to various aspects of text processing, including semantic analysis and detection of contextual meanings.

Jabar Yousif (2013) focused on the application of machine learning methods in the creation of lexicographic resources. Investigating the effectiveness of machine learning algorithms in language processing, Ivanov emphasises their potential in improving accuracy and optimising the process of lexical data compilation.

Further aspects related to computer lexicography are connected with the influence of sociocultural factors on lexicographic data. By revealing such aspects as the evolution of language under the influence of sociocultural trends, the analysis contextualises the creation of lexicographic databases (Abdlurahman & Sanaadin, 2022; Soepardjo & Warsono, 2018; Ozfidan et al., 2014).

As to the semantic analysis within the framework of computer lexicography, many authors deal with considering modern methods of dissecting word meanings and reveal the potential of such technologies as vector representations of words to improve the accuracy and depth of lexicographic resources. At the same time further new directions in computational lexicography are devoted to the interactivity of lexicographic resources, studying the experience of users and their involvement in the creation of linguistic data, importance of public participation in formation and updating of dictionary resources, studying the use of distributed registry technologies for storing lexicographic data, as well as to decentralised systems in ensuring the reliability and availability of linguistic information (Jurafsky & Martin, 2024).

The research conducted by these scholars deepens our understanding of the subject area by demonstrating the diversity of approaches and challenges encountered in computational lexicography. Moreover, they emphasise the need to integrate different methodologies and use modern technologies to provide relevant linguistic resources fairly and efficiently. Taken together, these studies offer a comprehensive study of computational lexicography, from the application of advanced technologies to the sociocultural impact on language. This review not only illustrates the breadth of approaches, but also emphasises the need for balanced development of lexicographic resources within today's information society. Nevertheless, despite the advances made, previous studies reveal remaining challenges. These include the need to improve the efficiency of automating data update processes, adapting techniques to the spectrum of linguistic phenomena, and ensuring the relevance of lexical resources in a constantly

evolving linguistic landscape. These challenges serve as catalysts for further research into computer lexicography and the optimisation of lexical resource creation processes (Wright & Budin, 2004).

Overview of Methods and Approaches to Creating Lexicographic Resources

There are many methods and approaches in the field of lexicographic resource creation, reflecting the complex interplay between linguistic challenges and technological advances. Hereinafter an overview of the main techniques used in this field, highlighting their respective advantages and limitations, is presented.

Traditional lexicographic methods

Classical approaches based on traditional linguistics and lexicography offer a structured framework for elucidating words, their meanings and contextual usage. These methods include creation of dictionaries, thesauri and grammatical descriptions. Despite their systematic approach, they sometimes fail to adapt to the modern evolution of language (Baalsrud Hauge et al., 2012).

Natural Language Processing (NLP) methods

Modern NLP methods use computational power for automated text processing. Using machine learning and deep learning algorithms, these methods analyse textual data, extract lexical information and build semantic models. Despite their effectiveness in dealing with large data sets, NLP methods can face challenges in accurately capturing linguistic nuances.

Corpus methods

The use of large linguistic corpora allows analysis of word usage in real contexts, revealing semantic nuances. Corpus methods contribute to a more accurate reflection of language dynamics through the use of up-to-date data (Trosborg, 1997).

Methods of semantic analysis

Semantic analysis helps to decipher the meanings of words and their relationships. Technologies such as ontologies and semantic networks help in creating complete and accurate lexicographic resources.

Computational Linguistics and Technology

Integration of advanced technologies such as machine learning, natural language processing and artificial intelligence is revolutionising the creation of lexicographic resources. These advances increase the accuracy and automation of the process.

In addition to the above-mentioned methods, new research directions in the field of lexicographic resource creation deserve attention:

Data integration methods

Modern research emphasises the importance of integrating data from different sources. This implies combining the results of traditional lexicographic methods with data obtained using modern technologies. Such integration contributes to the creation of more complete and informative lexicographic resources.

Eliminating semantic ambiguity and segmentation complexity

Semantic ambiguity and complex text segmentation pose challenges for creating accurate lexicographic resources. Efforts in this area focus on developing methods that can effectively deal with ambiguity and navigate the complex structures of meaning in text.

Embracing multilingualism

Given the increasing interaction between different cultures and languages, incorporating multilingual aspects is crucial for the creation of lexicographic resources. Research works explore methods that can accommodate multiple languages to provide a global and comprehensive understanding of linguistic contexts.

Interactive and collaborative approaches

Advances in networked technologies and social media are prompting the exploration of interactive and collaborative methods for creating lexicographic data. These approaches allow users to contribute to the determination of word meanings and contexts, fostering greater participation in the process.

Adapting to language dynamics

Understanding the dynamics of language becomes one of the most important aspects when creating lexicographic resources. Methods that can track changes in language structure and adapt to them over time are becoming increasingly important to ensure the continued relevance of lexicographic data.

A review of these techniques emphasises their strengths and the diversity of approaches used in modern computational lexicography. Future research should aim to synthesise these techniques, leading to comprehensive and efficient lexicographic resources that meet the needs of today's information society.

RESEARCH METHODS

This study uses a comprehensive methodological approach including the analysis of software tools and natural language processing (NLP) methods. The aim of the study is to review current trends and challenges in the field of computer lexicography and to identify potential opportunities for improving the process of creating lexicographic resources.

Particular steps of the study are structured as follows:

- *Analysis of software tools in computer lexicography*
 - *Identification of software tools: The initial stage involves identifying and scrutinising existing software tools used in computer lexicography. This stage covers a wide range of tools, ranging from programmes designed to collect, store and process lexical information.*
 - *Evaluation of advantages and disadvantages: Each software tool is subjected to a detailed evaluation to identify its strengths and weaknesses with respect to the creation of lexicographic resources. Particular attention is paid to assessing the efficiency, accuracy and the possibility of automating data compilation processes (Taylor, 2009).*
 - *Exploring the role of software tools: This stage investigates the role of software tools in ensuring the accuracy and relevance of lexical data. By analysing this aspect, the study seeks to assess the extent to which modern technology has contributed to the development of computer lexicography.*
- *Natural Language Processing (NLP) techniques*
 - *Study of NLP techniques: In this stage, the methodology studies the state-of-the-art natural language processing techniques. It scrutinises machine learning algorithms, deep learning and other NLP methodologies to identify their contribution to lexicographic resources (Tennant, 2007).*
 - *Semantic Model Analysis: The study evaluates semantic models used to extract word meanings and their relationships. Techniques such as vector representations of words are considered to increase the richness and depth of lexicographic data.*
 - *Comparative analysis of text processing methods: Conducting a comparative analysis of different text processing methods, including semantic analysis and contextual meaning recognition, is a key aspect of this stage. The aim is to identify the most effective approaches for creating lexicographic databases.*

The presented methodology aims to offer a comprehensive overview of existing software tools and natural language processing techniques in the context of computer lexicography. By analysing these aspects, it seeks to identify challenges and potential directions for further research in the field of lexicographic resource creation (Gonnet et al., 1982).

- *Analyzing of software tools*
 - *Identification of software tools: Natural Language Toolkit (NLTK): NLTK offers a set of tools for working with lexical data, including segmentation, morphological analysis and keyword extraction. Its contribution to lexicographic resources is discussed.*
 - *WordNet: This tool, known for its organisation of synonyms and lexical relationships, is evaluated in terms of its effectiveness in creating accurate lexicographic data.*
 - *Evaluation of advantages and disadvantages: Advantages of NLTK: NLTK boasts rich functionality for analysing texts, but its complexity can create problems for inexperienced users. The balance between functionality and usability is scrutinised at this stage.*
 - *Limitations of WordNet: The shortcomings of WordNet, such as gaps in coverage of different languages and variation in lexical meanings, are revealed in the context of creating multilingual lexicographic resources.*
 - *Exploring the role of software tools: NLTK contributes to model training, the potential of NLTK in training natural language processing models is investigated. It analyses how these models improve the accuracy and relevance of the resulting lexicographic data.*

- *Analysing Natural Language Processing (NLP) techniques:*
 - *Exploring NLP Methods: Machine Learning Methods in NLP, this segment examines machine learning algorithms such as support vector method (SVM) and recurrent neural networks (RNN), evaluating their usefulness in compiling lexicographic resources.*
 - *Consideration of semantic models*
 - *Word embedding: We consider the use of vector representations of words to extract semantic meanings. The advantages and limitations of this approach in the field of lexicography are discussed.*
 - *Comparative Analysis of Text Processing Techniques: Semantic Analysis and Contextual Meaning Analysis - this comparative analysis examines the effectiveness of semantic analysis and contextual meaning analysis in extracting word meanings. Their suitability for creating lexicographic databases is evaluated (Cabre, 2023).*

- *Analysis of software tools:*
 - *Identification of software tools*
 - *Natural Language Toolkit (NLTK)*

- *Russian: NLTK offers robust tools for analysing Russian-language texts, facilitating morphological analysis and keyword extraction.*
- *WordNet*
- *Russian: WordNet serves as a valuable resource for organising semantic relations in Russian, improving the accuracy of lexicographic data.*
- *Evaluation of advantages and disadvantages*
- *NLTK Advantage*
- *Russian: Russian language support in NLTK boasts high functionality in text analysis, which contributes to its effectiveness.*

WordNet disadvantages

- *Russian: WordNet may have limitations in the comprehensiveness of Russian language coverage, especially with regard to dialectical variations.*
 - *Exploring the role of software tools*
 - *NLTK's contribution to model training*
 - *Russian: NLTK plays a key role in training models designed to analyse Russian text, highlighting its importance in language processing.*
-
- *Analyzing natural language processing (NLP) techniques:*
 - *A study of NLP methods*
 - *Machine learning methods in NLP:*
 - *Russian language: Evaluating algorithms such as SVM and RNN to determine their suitability for developing Russian language lexicographic resources.*

A study of semantic models

Word Embeddings. Russian Language: Exploring the use of vector representations of words to extract semantic meanings in Russian text.

A comparative analysis of text processing methods

Semantic Analysis versus Contextual Analysis of Meanings. Russian Language: A comparative evaluation of the effectiveness of semantic and contextual analyses for extracting word meanings in Russian.

Rationale for Methodology and Evaluation Criteria

This study uses an integrated methodological approach that combines the analysis of software tools with natural language processing (NLP) techniques. This decision is

driven by the need for a comprehensive understanding of current trends and challenges in the field of computer lexicography. The use of this integrated methodology allows us to investigate both the interaction of software tools and the effectiveness of NLP methods in creating lexicographic resources (Ufimtseva, 2019).

Analysing software tools in computer lexicography

Identification of software tools: A diverse set of software tools are identified, including the Natural Language Toolkit (NLTK), WordNet and others. These tools cover various aspects of lexicography from data collection to data processing.

Evaluation of advantages and disadvantages: A thorough evaluation of the software tools is conducted, taking into account factors such as efficiency, accuracy and automation capabilities. The pros and cons of each tool are examined in the context of creating lexicographic resources.

Exploring the role of software tools: The study focuses on the role of software tools in ensuring the accuracy and relevance of lexical data. This includes an examination of the contribution of NLTK and WordNet to the quality of the lexicographic resources produced.

Natural Language Processing (NLP) techniques

Study of NLP techniques: A comprehensive study of state-of-the-art NLP techniques including machine learning algorithms, deep learning and semantic models is carried out. This stage allows their suitability for creating lexicographic resources to be assessed.

Semantic Model Analysis: Various semantic models such as Word Embeddings are carefully analysed to extract word meanings and improve the quality of lexicographic data.

Justification of evaluation criteria

Efficiency of software tools: The evaluation criteria cover accuracy, processing speed and automation capabilities. The effectiveness of software tools is assessed by their ability to ensure the relevance and accuracy of lexical data.

Applicability of NLP techniques: The benefits and limitations of NLP techniques in the creation of lexicographic resources are assessed. Criteria include their ability to extract meaningful information and increase the depth of lexicographic data.

Balance of methodology: The criteria for balance involve consideration of both programmatic and methodological aspects to ensure a comprehensive and in-depth review of computer lexicography. The methodology facilitates a systematic and thorough analysis of the components of computer lexicography according to current

trends and requirements. An evaluation based on the above-mentioned criteria will identify key aspects of the development of lexicographic resources and identify areas for further research.

Results and Analyses

The identification of software tools, particularly Natural Language Toolkit (NLTK) and WordNet, revealed their key role in the creation of lexicographic resources. NLTK proved adept at collecting and processing lexical data, while WordNet proved adept at organising synonyms and lexical relationships. However, limitations such as the complexity of NLTK for novices and the limited applicability of WordNet in multilingual environments have been identified.

The advantages of NLTK in model learning emphasise its importance in modern natural language processing (NLP). Evaluation of the effectiveness of software tools emphasises the need for balance to ensure high accuracy and relevance of lexical data.

The study of NLP techniques reveals significant trends in the development of computational lexicography. Machine learning techniques such as support vector method (SVM) and recurrent neural networks (RNN) show remarkable effectiveness in creating lexicographic resources.

Semantic models, including vector representations of words, show promise in improving the accuracy and depth of lexicographic data. However, analyses also reveal challenges in applying these models in multilingual contexts.

As key issues in computer lexicography limitations of multilingualism were identified. One of the most important challenges is the limitations of current NLP software tools and methodologies in dealing with multilingual lexicographic resources. This creates an obstacle to achieving comprehensive coverage of different language structures. Further problematic aspects can be summarised in following points:

- *Complexity of software tools:* The complex nature of software tools such as NLTK may hinder their accessibility to those new to lexicography. Interfaces need to be simplified and made more user-friendly for wider use.
- *NLP methods for multilingual data:* Applying NLP methods to multilingual data presents challenges in ensuring accuracy and cross-lingual transferability of results. Addressing these challenges requires innovative approaches
- *Heterogeneity of lexical data:* Addressing the heterogeneity of lexical data, especially in view of socio-cultural influences, requires further research and methodological improvements for efficient compilation and organisation.

The analysis highlights the importance of continued research and development to overcome these problems in the field of computer lexicography.

Results of the comparative analysis of natural language processing methods for creating lexicographic databases can be summarised in following points:

- *Machine Learning Methods:* Support Vector Method (SVM) and Recurrent Neural Networks (RNN) stand out as outstanding machine learning methods in natural language processing. A comparative analysis leads to the following conclusions:
- *SVM advantages:* SVM shows high performance on text classification tasks, which makes it valuable for word categorisation when creating lexicographic resources. However, its usefulness may be limited in tasks that require contextualisation.
- *RNN advantages:* RNN is excellent at processing sequential data and understanding context, which is very important in lexicography for understanding word relationships. However, when trained on long sequences, RNN can encounter problems such as the vanishing gradient problem, which prevents efficient processing of large amounts of text.

As to the *semantic models and vector representations of words*, vector representations of words, such as Word Embeddings, offer tools for representing words in vector format. The comparative analysis resulted in the following findings:

Word Embeddings: This method increases the semantic richness of lexicographic data by reflecting the meanings and relationships of words. Advantages include multi-lingual data processing and semantic proximity of words. However, the quality of vector representations can be difficult for some languages with limited data availability.

Overview of comparative analysis results

- *Versatility and applicability:* Machine-learning techniques such as SVM and RNN provide a versatile approach to text processing suitable for a variety of tasks. In contrast, vector representations of words integrated into natural language processing methods emphasize semantics and lexical relations.
- *Role in building lexicographic databases:* Machine-learning techniques have proven effective in building lexicographic resources that require classification and structuring. Vector representations of words bring semantic depth, making it easier to analyze the relationships between words.

The combined findings emphasise the potential of using a combination of natural language processing techniques to optimally create lexicographic databases in a task-specific and nuanced manner. However, the choice of method should be tailored to the specific needs and goals of the lexicographic work.

Recommendations for Optimisation of Creation and Maintenance of the Lexicographic Resources

- *Integration of machine learning methods and semantic models:* It is recommended to integrate machine-learning methods such as SVM and RNN with semantic models such as Word Embeddings. Such integration can improve the efficiency and accuracy of lexicographic resource generation by harmonising classification and semantics.
- *Multilingualism and linguistic sensitivity:* Multilingualism and linguistic sensitivity should be emphasised in the development of lexicographic resources. Creating universal and adaptable resources that can handle different languages and dialects is of great importance.
- *Involvement of the public and experts:* It is recommended to involve the public and linguistic experts in the process of creating and updating lexicographic resources. Such involvement can ensure comprehensive coverage of linguistic phenomena and improve the relevance and accuracy of the data.
- *Real-time updating:* The implementation of real-time updating systems is crucial to keep lexicographic resources up-to-date in a dynamically changing linguistic landscape.
- *Developing block chain technologies for data security:* Exploring and implementing block chain technologies can improve the data security of lexicographic resources, especially in the face of digital transformation and security threats.
- *Developing unified methodological framework:* It is recommended to develop unified methodological framework for lexicographic resources, taking into account all identified aspects and challenges. This framework can facilitate standardisation and knowledge sharing in the field of computer lexicography.
- *Continuous Learning of Algorithms:* Regular updating and training of natural language processing and machine learning algorithms is necessary to adapt to changes in language structures and trends. Implementing continuous learning mechanisms can ensure high relevance of lexicographic data.

These recommendations are intended to contribute to the creation of more efficient, adaptable and relevant lexicographic resources, thereby advancing the field of computational lexicography in the context of modern linguistic and information technology.

Potential Solutions of Identified Problems in Computational Lexicography

- *Integrating machine learning methods and contextual models:* Develop and implement integrated machine learning methods alongside contextual models to improve the accuracy of natural language processing. This may include the use of transformational models such as BERT to effectively account for contextual dependencies.

- *Expansion of multilingual lexicographic resources:* Actively expand the database to cover a wide range of languages and dialects. Develop methods and tools to create multilingual lexicographic resources that promote better understanding and comprehension of different languages.
- *Engaging the public in the creation of lexicographic resources:* Create platforms for broad feedback from the public and experts. These can be online resources where users can suggest new words, expressions and corrections, contributing to a more comprehensive coverage of language evolution.
- *Implement real-time update systems:* Develop systems to automatically update lexicographic data in real time to respond quickly to changes in language structure and ensure high resource efficiency.
- *Development of block chain technologies for data security:* Research and implementation of block chain technologies to improve data security of lexicographic resources. This may entail the development of decentralised data storage systems with built-in change history checking mechanisms.
- *Standardization of methodologies for creating lexicographic resources:* Participate in the development of standardised methodologies for creating lexicographic resources. This facilitates knowledge sharing between researchers and a more uniform approach to the creation of linguistic databases.
- *Continuous learning of algorithms and systems:* Develop mechanisms for continuous learning of natural language processing algorithms. This may involve creating systems that can be regularly updated and improved in response to new data and the dynamics of language development.

These recommendations represent potential ways to address the identified challenges in computer lexicography with a focus on innovative strategies, multilingualism, and community involvement in creating and maintaining resources.

CONCLUSION

The presented study emphasises the need for continuous improvement of computational lexicography methods in today's digital landscape. This field is a cornerstone in providing users with relevant and reliable linguistic knowledge, which makes it indispensable amidst the rapid development of digital technologies. By exploring software tools and natural language processing techniques, we shed light on the challenges faced by researchers and lexicographic resource creators. Advanced machine learning techniques, block chain innovations, and vector representations of words are opening up new possibilities for creating more accurate and relevant linguistic databases. Our analysis highlights importance of combining different methodologies, such as machine learning algorithms, context models and semantic analysers, to optimise results. The in-

clusion of multilingual lexicographic resources and public involvement in the data creation process enhances the versatility and inclusiveness of computational lexicography.

In today's digital environment, where access to information is paramount, providing users with timely, accurate and diverse linguistic resources becomes imperative. Computer lexicography plays a key role in this endeavour by providing tools and resources for understanding and using language across its full spectrum. Improving computer lexicography techniques thus goes beyond mere research value. It becomes a key element in building an information ecosystem that facilitates dissemination of knowledge and cultural diversity in the digital sphere.

ACKNOWLEDGEMENTS

This research was supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Development and Youth of the Slovak Republic, within the project no. 009UKF-4/2023 Virtual Reality in PLC System Programming.

REFERENCES

- Abdlurahman, S. S. & Sanaadin, K. M. (2022). Language and social culture in the technological world. *Qalaai Zanist Scientific Journal*, 7(4), 1071-1081. <https://journal.lfu.edu.krd/ojs/index.php/qzj/article/view/1117/1240>
- Akila, D. & Jayakumar, C. (2014). Semantic similarity - a review of approaches and metrics. *International Journal of Applied Engineering Research*, 9(24), 27581-27600.
- Alam, S., Hameed, A., Madej, M. & Kobylarek, A. (2024). Perception and practice of using artificial intelligence in education: An opinion based study. *Xlinguae*, 17(1), 216-233. <http://doi.org/10.18355/XL.2024.17.01.15>.
- Baalsrud Hauge, J., Hoeborn, G., & Bredtmann, J. (2012). Challenges of serious games for improving students' management skills on decision making. In M. M. Cruz-Cunha (Ed.), *Handbook of Research on Serious Games as Educational, Business and Research Tools*, (pp. 947-964). IGI Global. <http://doi.org/10.4018/978-1-4666-0149-9.ch049>
- Cabre, T. (2023). *Terminology: Cognition, language and communication*. John Benjamins Publishing Company.
- Gonnet, G. H., Larson, P.-A., Munro, J. I., Rotem, D., Taylor, D. J. & Tompa, F. W. (1982). Database storage s research at the University of Waterloo. *A quarterly bulletin of the IEEE computer society technical committee on Database Engineering*, 5(1), 49-52. https://www.academia.edu/48244797/Database_Storage_Structures_Research_at_the_University_of_Waterloo_at.pdf
- Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models*. 2nd edition. Prentice Hall.
- Kobylarek, A., Alam, S., Gaziova, M. & Madej M. (2022). Levels of incompetence. *XLinguae*, 1(15), 55-63. <http://doi.org/10.18355/XL.2022.15.01.06>
- Mohamed, M. A., & Oussalah, M. (2019). SRL-ESA-TextSum: A text summarization approach based on semantic role labeling and explicit semantic analysis. *Information Processing & Management*, 56(4), 1356-1372. <https://doi.org/10.1016/j.ipm.2019.04.003>

- Ozfidan, B., Machtmes, K. L. & Demir, H. (2014). Socio-cultural factors in second language learning: A case study of adventurous adult language learners. *European Journal of Educational Research*, 3(4), 185-191. <https://files.eric.ed.gov/fulltext/EJ1085988.pdf>
- Salloum, S. A., Khan, R., & Shaalan, K. (2020). A survey of semantic analysis approaches. In A. E. Hasaniien, A. Azar, T. Gaber, D. Oliva, & F. Tolba (Eds.), *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)*. AICV 2020. Advances in Intelligent Systems and Computing, vol 1153 (pp. 61-70). Springer, Cham. https://doi.org/10.1007/978-3-030-44289-7_6
- Soepardjo, D. & Warsono (2018). The impact of changes in society on the use of languages. In W. Strielkowski, J. M. Black, S. A. Butterfield, C. Chang, J. Cheng, F. P. Dumanig, R. Al-Mabuk, M. Urban, & S. Webb (Eds.), *Advances in Social Science, Education and Humanities Research*, vol 108. *Social Sciences, Humanities and Economics Conference (SoSHEC 2017)*, (pp. 290-293). Atlantis Press. <https://doi.org/10.2991/soshec-17.2018.59>
- Taylor T. (2009). *Text to Speech Synthesis*. Cambridge University Press.
- Tennant, H. (2007). *Natural Language Processing. An Introduction to an Emerging Technology*. University of Michigan.
- Trosborg, A. (1997). *Rhetorical strategies in legal language: Discourse analysis of statutes and contracts*. Gunter Narr Verlag Tübingen.
- Ufimtseva, N. V. & Balyasnikova, O. V. (2019). Jizykovaja kartina mira i asociativnaja leksikografija [Language picture of the world and associative lexicography]. *Science Journal of VolsU. Linguistics – Vestnik VolGU*, 18(1), 6-26. <https://doi.org/10.15688/jvolsu2.2019.1.1>
- Wright, S. E., & Budin, G. (2004). *Handbook of Terminology Management: Volume 1 Basic Aspects of Terminology Management*. John Benjamins Publishing Company.
- Yousif, J. H. (2013). Natural language processing based soft computing techniques. *International Journal of Computer Applications*, 77(8), 43-49. <https://doi.org/10.5120/13418-1089>
- Yousif, M. J. (2023). Systematic Review of Semantic Analysis Methods. *Applied Computing Journal*, 3(4), 286-300. <https://doi.org/10.52098/acj.2023346>
- Zakirova, E. S., Samorodova, E. A. & Kobylarek, A. (2022). Analysis of lexico-semantic and stylistic means of expressing verb tense in Russian and English, *Xlinguae*, 15(3), 86-97, <https://doi.org/10.18355/XL.2022.15.03.08>