СЕРІЯ «ФІЛОЛОГІЯ»

UDC 519.6 : 004.912 :: 81`32

https://doi.org/10.52058/2786-6165-2025-6(36)-25-49

Krasniuk Svitlana Oleksandrivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (063) 240-55-85, https://orcid.org/0000-0002-5987-8681

Goncharenko Svitlana Mykolaivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (044) 256-29-49, https://orcid.org/0000-0002-7740-4658

Denysenko Vitalina Mykolaivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (044) 256-29-49, https://orcid.org/0000-0002-4918-5830

Petrenko Viktoriia Yuriivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (044) 256-29-49, https://orcid.org/0009-0001-6403-2082

Redko Svitlana Petrivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (044) 256-29-49, https://orcid.org/0000-0002-2474-5344

Roienko Liudmyla Vitaliivna senior lecturer, Kyiv National University of Technologies and Design, Institute of Law and Modern Technologies, Kyiv, tel.: (044) 256-29-49, https://orcid.org/0000-0002-6794-0051

KNOWLEDGE-BASED VS DATA-DRIVEN PARADIGMS OF TEXT MINING IN MODERN MACHINE LINGUISTICS IN CONTEXT OF GLOBAL CRISES

Abstract. The introduction and the first part of the article are devoted to the analysis of how crisis transformations (both predicted and force majeure) at the global/regional/national levels (which can be caused by: wars, pandemics,

environmental threats and disasters, socio-political revolutions and cataclysms, mass migrations and rapid digitalization (including the avalanche-like introduction of AI in the context of Big Data)) - form new cardinal challenges/risks not only for applied linguistics, but also for philology in general. These challenges/risks are rapidly transforming both the conceptual and methodological foundations of modern philology in general, and the technological and algorithmic foundations of applied machine linguistics in particular. The authors argue that modern philology (which traditionally focused on hermeneutic text analysis) should now be transformed in an accelerated anti-crisis mode in the direction of maximally interdisciplinary, contextual and flexible concepts/paradigms/approaches.

In particular, the applied focus of such an accelerated transformation should be renewal of the concept of Text Mining, in the direction of synergistic processing and analysis of natural language, capable of effectively working with semi-structured, multidimensional Big Data in conditions of information turbulence in a crisis humanitarian context.

Taking into account the above, the second, main part of the article reflects the results of a systematic comparative study of the two main methodological paradigms of modern machine linguistics (knowledge-based and data-driven) taking into account the possible impact of all types and levels of crisis phenomena. Moreover, the knowledge-based paradigm is based on linguistic knowledge explicitly formalized by human experts (dictionaries, grammars, logical rules, other ontologies), and the data-driven paradigm uses statistical algorithms, classical machine learning algorithms, and deep neural network learning to detect hidden patterns in large corpora of text without prior expert linguistic modeling (without prior manual formalization).

Since data-driven methods and algorithms dominate modern machine linguistics, this is why the article pays additional attention to the data-driven paradigm, which is currently the main one in the tasks of machine translation, text generation, syntactic analysis, virtual assistants, and large language models (LLMs) for the most modern linguistic systems - from Google Translate to ChatGPT.

For the data-driven paradigm, advantages, disadvantages, and recommendations are highlighted, in particular, it is noted that data-driven methods demonstrate the highest efficiency in conditions of large amounts of data and, at the same time, in tasks that do not have strict requirements for explainability and interpretability of results.

The article also examines the advantages, disadvantages, and recommended application areas and tasks where knowledge-based technologies remain critically important and EFFECTIVE - in particular, legal, medical, and humanitarian tasks

that require high accuracy and transparent logic of interpretation. Thus, the authors conclude that in conditions of local and global challenges and crisis phenomena, none of the paradigms is universal, because the future of machine linguistics lies in synergistic hybrid systems that provide a balance between adaptability, productivity, interpretability, ethics, and cultural and/or linguistic stability.

That is, in crisis and unstable modern conditions - the most effective are hybrid approaches to Text Mining, which combine the scalability, adaptability and speed of using data-driven methods and models with the interpretability, transparency and semantic depth of the knowledge-based paradigm.

Such a hybrid approach opens up new horizons for philology as a science, which becomes not only an object, but also a subject of the digital transformation of society in the current era of crises.

In detail - modern machine linguistics is no longer an exclusively applied technical or/and humanitarian field, but is turning into an active participant in social transformations, capable of supporting information security, humanitarian interaction, preservation of linguistic heritage and cultural identity in modern global, regional and national crisis conditions.

Keywords: machine linguistics, text mining, knowledge-based paradigm, data-driven paradigm, machine learning, crisis conditions

Краснюк Світлана Олександрівна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (063) 240-55-85, https://orcid.org/0000-0002-5987-8681

Гончаренко Світлана Миколаївна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (044) 256-29-49, https://orcid.org/0000-0002-7740-4658

Денисенко Віталіна Миколаївна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (044) 256-29-49, https://orcid.org/0000-0002-4918-5830

Петренко Вікторія Юріївна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (044) 256-29-49, https://orcid.org/0009-0001-6403-2082

Редько Світлана Петрівна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (044) 256-29-49, https://orcid.org/0000-0002-2474-5344

Роєнко Людмила Віталіївна старший викладач, Київський національний університет технологій та дизайну, Інститут права та сучасних технологій, м. Київ, тел.: (044) 256-29-49, https://orcid.org/0000-0002-6794-0051

ПАРАДИГМИ TEXT MINING, ЩО ЗАСНОВАНІ НА ЗНАННЯХ АБО ДАНИХ В СУЧАСНІЙ МАШИННІЙ ЛІНГВІСТИЦІ У КОНТЕКСТІ ГЛОБАЛЬНИХ КРИЗ

Анотація. Вступ та перша частина статті присвячена аналізу того, як кризові трансформації (і прогнозовані і форс-мажорні) на світовому/регіональному/національному рівнях (що можуть бути спричинені: війнами, пандеміями, екологічними загрозами і катастрофами, соціально-політичними революціями та катаклізмами, масовими міграціями та стрімкою цифровізацією (зокрема і лавиноподібним впровадженням AI в умовах Big Data)) - формують нові кардинальні виклики/ризики не лише перед прикладною лінгвістичною наукою, але і перед філологією загалом. Ці виклики/ризики прискорено трансформують як концептуальні і методологічні засади сучасної філології загалом, так і технологічні та алгоритмічні засади прикладної машинної лінгвістики зокрема. Автори обґрунтовують, що сучасна філологія (що традиційно зосереджувалася на герменевтичному аналізі тексту), нині має в прискореному антикризовому режимі трансформуватися в напрямку максимально міждисциплінарних, контекстних і гнучких концепцій/парадигм/підходів.

Зокрема, в прикладному фокусі такої прискореної трансформації має бути оновлення концепції Text Mining, в напрямку синергетичної обробки та аналізу природної мови, здатної ефективно працювати з напівструктурованими, багатовимірними Big Data в умовах інформаційної турбулентності в кризовому гуманітарному контексті.

Враховуючи вищенаведене, у другій, основній частині статті – відображено результати системного порівняльного дослідження двох основних методологічних парадигм сучасної машинної лінгвістики (knowledge-based та data-driven) з урахуванням можливого впливу всіх типів і рівнів кризових явищ. Причому, knowledge-based парадигма - базується на явно формалізованих людьми-експертами лінгвістичних знаннях (словники,

граматики, логічні правила, інші онтології), а data-driven парадигма - використовує статистичні алгоритми, алгоритми класичного машинного навчання й глибокого нейромережевого навчання для виявлення прихованих закономірностей у великих корпусах тексту без попереднього експертного лінгвістичного моделювання (без попередньої ручної формалізації).

Оскільки data-driven методи та алгоритми домінують у сучасній машинній лінгвістиці, саме тому, в статті додаткову увагу приділено саме data-driven парадигмі, яка на сьогодні є основною у завданнях машинного перекладу, генерації тексту, синтаксичного аналізу, віртуальних асистентів і великих мовних моделей (LLMs) для найсучасніших лінгвістичних систем — від Google Translate до ChatGPT.

Для data-driven парадигми виділені переваги, недоліки, надані рекомендації, зокрема зазначено, що саме data-driven методи демонструють найвищу ефективність в умовах великих обсягах даних та, одночасно, у задачах, які не мають жорстких вимог до пояснюваності та інтерпретованості результатів.

У статті також досліджено переваги, недоліки, рекомендовані прикладні сфери та задачі, де knowledge-based технології залишаються критично важливими і ЕФЕКТИВНИМИ.

Зокрема, юридичні, медичні та гуманітарні задачі, що вимагають високої точності й прозорої логіки інтерпретації.

Таким чином, автори доходять висновку, що в умовах локальних та глобальних викликів і кризових явищ, жодна з парадигм не є універсальною, адже майбутнє машинної лінгвістики — за синергетичними гібридними системами, які забезпечують баланс між адаптивністю, продуктивністю, інтерпретованістю, етичністю та культурною та/або мовною стійкістю. Тобто у кризових та нестабільних сучасних умовах - найбільш ефективними є гібридні підходи до Text Mining, які поєднують масштабованість, адаптивність та швидкодію використання data-driven методів та моделей із інтерпретованістю, транспарентністю та семантичною глибиною knowledge-based парадигми. Такий гібридний підхід відкриває нові горизонти для філології як науки, що стає не лише об’єктом, а й суб’єктом цифрової трансформації суспільства в поточну епоху криз.

Деталізуючи - сучасна машинна лінгвістика більше не є виключно прикладною технічною чи/та гуманітарною галуззю, а перетворюється на активного учасника соціальних трансформацій, здатного підтримувати інформаційну безпеку, гуманітарну взаємодію, збереження мовної спадщини та культурної ідентичності в сучасних глобальних, регіональних та національних кризових умовах.

Ключові слова: машинна лінгвістика, text mining, заснована на знаннях парадигма, керована даними парадигма, машинне навчання, кризові умови.

Introduction

1. Philology in the 21st century is undergoing significant changes under the influence of technological, cultural and epistemological factors. If traditionally it was perceived as a set of disciplines that study language, literature and text in a historical and cultural context, today philology is transforming into an innovative science that integrates digital technologies, multimodal approaches and interdisciplinary practices.

The impact of instability and crises on modern philology is multidimensional and manifests itself both in the content of philological research and in the transformation of methodologies, goals and the role of humanitarian knowledge in society. In an era of global challenges - wars, pandemics, political transformations, environmental threats - philology is experiencing a rethinking of its function, boundaries and responsibility. Philology as a science of word, text and discourse inevitably responds to the challenges of global crises and military conflicts. In the context of social instability, mass migrations, the destruction of cultural codes and the formation of new collective traumas, philological research is gaining new relevance, transforming both research objects and methodologies.

2. The impact of global crises, imbalances and instability on modern linguistics is multi-level and covers both the substantive aspects of linguistic research and methodological, social and ethical dimensions. In the 21st century, linguistics increasingly functions not only as a descriptive science, but also as a tool for responding to transformations in a world marked by wars, migration, digitalization, cultural and environmental crises.

Computational linguistics, as an area at the junction of linguistics and computer science, is gaining particular importance in the context of global crises and military conflicts. Modern challenges radically change the tasks, methods and areas of application of machine language technologies, emphasizing their role in ensuring information security, humanitarian aid, preserving cultural heritage and supporting communication in crisis zones.

3. Modern innovative philology, being at the intersection of humanitarian and technical knowledge, actively integrates methods and technologies characteristic of related disciplines, including mathematical linguistics (studying language using formal models) and machine linguistics (closer to computer implementation). It should be noted that mathematical linguistics is a formal and quantitative study of language using mathematical methods and models. Mathematical linguistics includes: the theory of formal grammars, logical analysis of language, information theory, statistics and probabilistic models.

The goal and objectives of mathematical linguistics are to describe and study natural language as a formal system.

But currently, the more promising direction of this interdisciplinary integration, precisely in today's complex (global and regional, long-term instability, crises and conflicts) and complex (semi-structured multidimensional Big Data) conditions, is mainly machine linguistics, which is aimed at effective formalization and automation of processing, analysis, interpretation (and then for generation) of natural language and/or texts (in batch and streaming modes). In other words, machine linguistics is the synergistic application of mathematical, statistical, algorithmic and intellectual/cognitive technologies to create computer systems that are capable of processing, analyzing, understanding and generating natural language (NLP).

4. Machine linguistics is developing at the intersection of two fundamentally different methodological paradigms - knowledge-based and data-driven. In modern conditions, data-driven approaches dominate machine linguistics, although knowledge-based methods have not lost their significance and remain important in a number of specific tasks.

5. Knowledge-based technologies in machine linguistics are a classical, but still relevant paradigm in some tasks, focused on a "transparent", explicit interpretation (inductive and deductive) of language/speech/text at the level of meanings and conceptual structures (and not only on the basis of statistical patterns, mathematical models learned from data [1]).

Knowledge-based technologies are a class of approaches in machine linguistics that focus on using explicit linguistic and conceptual knowledge for automatic processing, analysis, interpretation, and even generation of natural language and/or texts. This knowledge is formalized in the form of dictionaries, ontologies, grammars, and represented in the form of logical rules, frames, semantic networks, and occasionally predicate logic [2].

Unlike purely data-driven methods (statistical or neural network models), knowledge-based technologies use formalized knowledge [3] (grammatical, semantic, lexical, etc. [4]), which is aimed at semantic and pragmatic understanding of the text, approaching cognitive understanding of the subject domain (which is especially important for intersection tasks).

6. Data-driven approaches in machine linguistics involve automatic processing of natural language (Natural Language Processing) using statistical methods, classical ML algorithms or even trained/learned neural network models (which provide very high-quality results in complex, even interdisciplinary tasks, but require very large and high-quality labeled input data [5]). That is, unlike knowledge-based approaches, where knowledge is formalized manually, in data-driven methods, the key role is played by accumulated corpora of labeled data and

machine learning algorithms, which allow detecting patterns in language WITHOUT the need for a priori linguistic modeling by human experts [6].

Currently, in modern machine linguistics, most applied tasks, including the most advanced systems (such as ChatGPT, Gemini, Copilot etc.), are dominated by data-driven methods, especially using deep learning and large-scale language models (LLMs). *However, the authors put forward the thesis that it is HYBRID technologies that combine both approaches [7], [8] - will be maximally ADAPTIVE, and, therefore, maximally effective (optimal), especially in complex, crisis-like external dynamic conditions/factors.*

Formulation of the problem

Text mining is a composite, complex interdisciplinary applied direction in Data Science, which combines almost ALL machine linguistics technologies: statistical analysis, classical ML and Deep ML = for automatic extraction of knowledge/regularities/patterns from unstructured Big Text Data (for their further effective use in knowledge-based or/and data driven AI).

Thus, within the framework of machine linguistics, text mining plays a key role in processing large corpuses of texts, providing opportunities for automated processing, intellectual analysis and analytics of lexical, syntactic and semantic structures. It should be noted that text mining in machine linguistics can be both data-driven and knowledge-based (and in recent decades, it is the data-driven approach to Text Mining that has gained greater popularity).

The example of data-driven text mining in content-analysis mode is shown in Fig. 1.

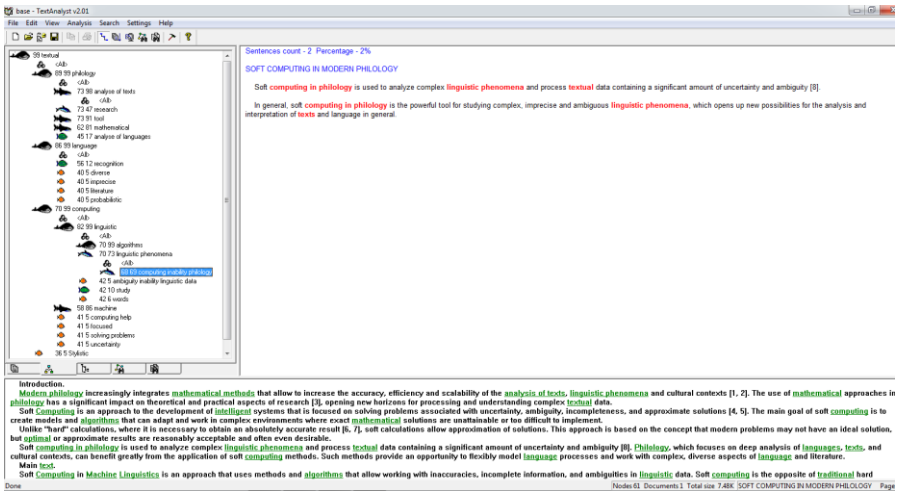

Fig. 1. The example of data-driven paradigm of text-mining in content-analysis mode (TextAnalyst 2.0 software was used and the article https://er.knutd.edu.ua/bitstream/123456789/29518/2/article_Germany_24_pp.%20135-141.pdf was used as the source of text data)

Source: author's modeling results

That is why, taking into account all of the above, the complex task of analyzing and comparing the effectiveness of these two approaches (data-driven and knowledge-based) to Text Mining in complex external crisis conditions/factors (which ultimately have an impact on modern applied linguistics) becomes particularly relevant, BUT also the generation of proposals for improving the accuracy and completeness of Text mining results for both the corporate sector (for example, for Big Tech or Tech Giants), the national security sector (in the field of OSINT intelligence, etc.), and for the tasks of modern applied philological scientific research.

An additional task of this study was not only a critical analysis and systematic comparison of classical ML methods and neural network ML methods within the data driven paradigm for Text Mining, but also the generation of proposals for increasing their effectiveness in the current complex/unstable, often crisis-like external conditions/factors that have both regular and ad-hoc impacts on applied machine linguistics (for example: long-term conflicts and multiple wars in the Middle East, the war in Ukraine that has been ongoing since 2014, etc.).

Analysis of recent research and publications

The main foundations of the modern theory of modern machine linguistics were revealed in recent works by such scientists as: Ahmad, A., Abbasi, I. A., Abbasi, R. H., & Rasheed, B. [9]; Church, K., & Liberman, M. [10]; Gatla, T. R. [11]; Jalilbayli, O. B. [12]; Li, W. [13]; McShane, M., & Nirenburg, S. [14]; Stepanova, I. S., Nykyporets, S. S., & Hadaichuk, N. M. [15]; Tasheva, N. [16].

The modern questions of the classical knowledge-based machine linguistics were considered in recent works by such scientists as: Becker M. [17]; Oussalah M., Mohamed M. [18]; Papegaaij B. C. [19]; Su C., Wu K., Chen Y. [20]; Wątróbski J. [21] and other.

The innovative trends of the data-driven machine linguistics were considered in recent works by such scientists as: Akstinaite V., Garrard P., Sadler-Smith E. [22]; Bambini V. et al. [23]; Boulton A., Vyatkina N. [24]; Gómez-Vilda, P., & Gómez-Rodellar, A. [25]; Kaur K. et al. [26]; Li C. C. et al. [27]; Shawaqfeh A. T. et al. [28]; Vanmassenhove E., Shterionov D., Gwilliam M. [29]; Varda A. G., Marelli M. [30] and other.

Considering the above, currently the task of comprehensive analysis of the effectiveness of these two approaches (data-driven and knowledge-based) in modern Text Mining is not only unresolved, but also the task of developing a set of methodological/technological/algorithmic/parametric recommendations to increase the effectiveness of both paradigms in external, complex, crisis and unpredictable conditions/factors (which currently have a significant impact on both fundamental and applied scientific and practical research in modern machine linguistics).

Purpose of the article

In view of the above, the current goal of the current comprehensive interdisciplinary collective scientific research was not only to comprehensively and critically analyze both the classical knowledge-based paradigm and the more modern data-driven paradigm of Text Mining in modern machine linguistics, but also to offer a set of methodological, technological and parametric practical recommendations for increasing the effectiveness of both paradigms in complex, crisis and unpredictable external conditions/factors/events (which currently have a significant impact on both fundamental and applied scientific and practical research in modern machine linguistics).

Main part of the research

Considering the above main goal of this complex interdisciplinary collective scientific research:

- below in section **A)**, the results of a critical analysis of the classical knowledge-based paradigm of modern machine linguistics are presented;

- and in section **B)**, the results of a critical analysis of the more modern and popular data-driven paradigm of machine linguistics are presented.

- further, in section **C)**, a set of methodological, technological and parametric scientific and practical recommendations are proposed to increase the efficiency of both paradigms through their HYBRID use in order to obtain the SYNERGY effect.

A) So, as it is stated at the beginning of this section, this subsection will present the results of a critical analysis of the classical knowledge-based paradigm of modern machine linguistics.

Knowledge-based approach (based on knowledge) in machine linguistics is one of the key directions in the field of natural language processing, oriented use of formalized linguistic knowledge. In contrast to statistical and neural network methods, which mainly rely on large arrays of training data, knowledge-based systems use pre-developed grammatical rules, semantic networks, thesauri, ontologies and lexical resources, structured manually or with the involvement of experts.

The main advantage of this approach is its ability to provide interpreted and stable analysis of language structures, especially in conditions of limited training corpora, high degree of ambiguity or specificity of the subject area. This approach is widely used in tasks of morphological analysis, syntactic and semantic parsing, automatic translation, as well as in the development of expert systems and dialog agents.

One of the defining features of the knowledge-based approach is the use of linguistic models built with the participation of experts. Such models include sentence construction rules, morphological schemes,2 semantic networks, etc.

At the same time, the system is able to perform deductive reasoning based on the existing knowledge base, which significantly expands its functional capabilities. The key components of knowledge-based systems are:

- ontologies - formalized representations of concepts and relations between them; - grammatical formalisms - rules that define permissible syntactic structures; - lexical bases - dictionaries with morphological, syntactic and semantic information.

One of the characteristic features of knowledge-based linguistics is the possibility of logical inference taking into account knowledge, which allows the system to perform deductive reasoning and process even those cases that were presented in the training data.

However, this approach also has certain limitations. First of all, it is the high complexity of developing and maintaining knowledge bases, the complexity of scaling to new languages or domains, as well as limited flexibility in processing living, non-standardized language. *In this regard, the hybridization of knowledge-based and data-driven approaches is becoming increasingly relevant, allowing for the combination of the interpretability of the former with the learning and scalability of the latter.* Given the above, let us formalize the result of the analysis of the main advantages vs disadvantages of the knowledge-based paradigm of modern machine linguistics:

1) Advantages:

- interpretability: Decisions are made based on clearly defined rules.
- reliability in conditions of limited data: Can work when corpora are small or absent.
- controllability: Well suited for tasks where accuracy is critically important (for example, legal or medical texts).

2). Main disadvantages:

- high consumption of time and resources: Requires a lot of effort from linguists to create rules.
- poor scalability: Difficult to adapt to new languages, dialects or styles.
- limited flexibility: Do not cope with ambiguity as well as statist methods.

Thus, knowledge-based machine linguistics continues to play a significant role in the development of intelligent language technologies, especially in cases where a high level of accuracy, transparency of interpretation, and deep understanding of language are required.

B) Therefore, as it is stated at the beginning of this section, this subsection will present the results of a critical analysis of the more modern and popular data-driven paradigm of machine linguistics.

At the beginning of this section, the authors consider it necessary to conduct a chronological analysis of the evolution of machine linguistics paradigms and

their individual components. As it is mentioned above, early machine linguistics systems were built on a knowledge-based paradigm, in particular, they were based on obvious linguistic rules compiled manually by experts.

These systems required: accurate morphological and syntactic descriptions; grammars and dictionaries; significant efforts of human experts: linguists and programmers. Such approaches had high accuracy in limited domains, but did not scale well and did not cope with linguistic diversity and exceptions.

However, since the 1990s, with the accumulation of electronic text corpora, there has been a transition to a data-driven paradigm, in particular, based on statistical methods. Models began to use word and phrase probabilities based on frequency analysis. The example of such statistical (namely, frequency) text-mining performed in the Summarization mode is shown in Fig. 2.

Since the 2010s, neural network methods have begun to replace classical data-driven algorithms, and data-driven linguistics has become associated with deep neural networks (DNN), which outperform statistical methods in most metrics. These models are trained on huge corpora and are capable of solving a wide range of tasks without the need for manual development of linguistic rules.

Three stages of development of such neural network data-driven architectures should be distinguished: - recurrent neural networks (RNN, LSTM) - for processing sequences; attention mechanisms (Attention) - improved the quality of translation and understanding of the context; transformers (Transformers) - revolutionized the entire field of NLP (for example, BERT, GPT). Models based on transformers are especially important, such as: BERT - a model for understanding text; GPT - a generative model used for dialogue and text generation; T5, XLNet, RoBERTa - advanced architectures for a wide range of linguistic tasks.

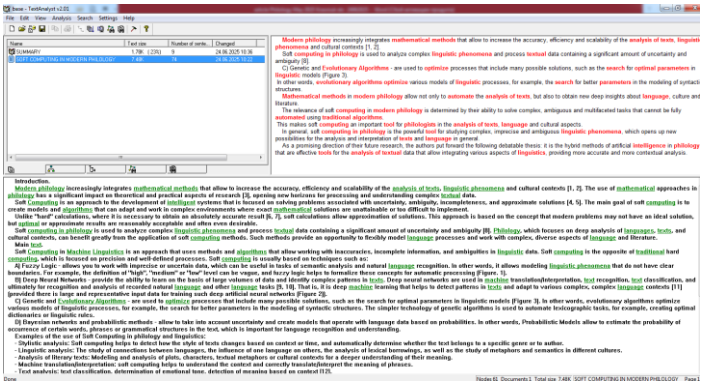

Fig. 2. The example of performed statistical (namely, frequency) text-mining in Summarization mode (TextAnalyst 2.0 software was used and the article https://er.knutd.edu.ua/bitstream/123456789/29518/2/article_Germany_24_pp.%20135-141.pdf was used as the source of text data)

Source: author's modeling results

Summarizing the above brief chronological analysis of the evolution of machine linguistics paradigms and their individual components, it can be argued that over the past three decades, modern machine linguistics has undergone a significant transformation: from the traditional knowledge-based (often rule-based) paradigm = to a data-driven paradigm.

That is, unlike the traditional rule-based approach, where linguists manually developed formal grammars, data-driven methods rely on machine learning and obtaining patterns from language corpora. Modern achievements in the field of machine learning, especially deep learning (deep ANN), allow processing huge arrays of texts and effectively starting/training models (including LLM) on practical heuristics, rather than on formal rules previously set by experts. This allows such data-driven machine linguistics systems to effectively adapt to the diversity and variability of language.

The data-driven approach has ALREADY become the algorithmic basis (informal standard) of such machine linguistics technologies as machine translation, voice assistants, chatbots and tone analysis systems.

This transition marks not only a technological but also a paradigm shift in language science, opening up new horizons in the understanding, modeling and application of human language.

Key Principles and Features of the data-driven approach in machine linguistics:

- Big Text Data as a source of new patterns/patterns/knowledge. The data-driven approach is based on the premise that language information can be extracted from large corpora - collections of texts collected from the Internet, books, news, social networks, etc. Models are "trained" by language patterns, exploring: lexical combinations; morphological forms; syntactic structures; pragmatic and semantic connections.

- the use of initially statistical, later probabilistic, and since the 2010s, neural network models to model the deep and semantic representation of language. Modern models form vector representations of words and phrases (embeddings). This allows: to understand synonyms and homonyms; to find semantic proximity; to recognize hidden dependencies and context, etc.

- different types of machine learning on labeled and/or UNlabeled Big Text Data. That is, using: unsupervised machine learning - patterns are obtained from previously Unlabeled/UNlabeled data (for example, Word2Vec, BERT); using supervised machine learning, previously qualitatively labeled/annotated text corpora are used for training/training (for example, Named Entity Recognition tasks, text classification); and using semi-supervised machine learning, labeled/annotated text corpora and Large Unlabeled/Unannotated text data sets are used AT THE SAME TIME.

- very minimal participation of human linguists in creating formal grammars or other ontologies;

- the possibility of effective scaling to other related languages, dialects, jargons if there is sufficient input data for retraining/retraining existing models;

The main areas of application of the data-driven approach in machine linguistics:

1. Machine translation. The data-driven approach formed the basis of statistical machine translation (SMT), and then neural network (NMT). For example, Google Translate switched from phrase SMT to neural network models (neural network transformers), which significantly improved the quality of translation. They surpass the old rule-based and statistical systems in accuracy and smoothness.

2. Tokenization and morphological analysis. Machine learning systems have learned to automatically split texts into tokens, recognize lemmas and morphological features based on labels and large corpora.

3. Syntactic analysis. Statistical parsers, such as Stanford Parser or spaCy, are trained on annotated corpora and can predict dependency trees without explicit grammars.

4. Semantic analysis. Data-driven methods allow not only name entity recognition (NER), sentiment analysis and even the construction of embeddings (verbal representations), such as Word2Vec, GloVe, BERT, which capture subtle semantic dependencies between words.

5. Speech recognition and synthesis. Modern ASR (automatic speech recognition) and TTS (text-to-speech) systems are also built on data-driven approaches, including neural networks and deep learning. For example, voice assistants (Siri, Alexa, Google Assistant) use data-driven models for: converting speech to text (ASR); understanding meaning; generating natural sounding (TTS)

6. Scientific linguistic research. A data-driven approach allows: studying language change over time (diachrony); analyzing language variations (sociolinguistics); finding patterns that are inaccessible in manual processing, etc.

Advantages of the data-driven paradigm of modern machine linguistics, in particular its most relevant direction - artificial deep neural networks (the current successes of ChatGPT, Google Translate, BERT, GPT, T5 confirm this):

- high scalability and the possibility of implementing distributed learning: this allows you to scale the use of the model in various industries/domains from edge computing to cloud NLP WITHOUT changing/reengineering the architecture and hyperparameters of deep NN.

In addition, it should be noted that it is deep neural networks that can (even should) not only train on very large text corpora, but also have good performance in the mode of their use/operation.

- flexibility for: different languages, domains (subject areas);

tasks (translation, summarization, tone analysis, etc.), styles, jargons, dialects.

- high efficiency, i.e. accuracy and productivity of automated processing, analysis and analytics of "live" language (including in STREAM format), including slang, jargon, pronunciation errors and features, memes, etc.

- the possibility of automatic adaptive learning. That is, constantly repeated relevance due to the possibility of effective additional training (or even retraining) on new data sets.

In machine learning, in particular when working with deep neural networks, additional training strategies (fine-tuning strategies) allow you to adapt a pre-trained model to new tasks, domains or languages.

This significantly reduces training costs and allows you to get high quality even on small datasets.

Disadvantages and limitations of the data-driven paradigm of modern machine linguistics, in particular its most relevant direction - artificial deep neural networks:

- the need for large amounts of data:

without very large data corpora and/or labeled examples, such models are unable to train effectively and qualitatively.

- require large computational resources;

- are usually "black boxes" - it is difficult to interpret the linguistic logic of the model;

- ethical threats and threats to the confidentiality of private data due to significant vulnerability to biased and/or "dirty" (i.e., poor-quality) input data (the possibility of learning during model training and subsequent reproduction of biases from the training corpus of input data);

- significant risks of "absorption" (even accelerated marginalization, up to disappearance) for rare languages and dialects - after all, data-driven methods rely on large volumes of linguistic data, most often available on the Internet. However: large languages (for example, English, Chinese, Spanish) have large digital corpora: Wikipedia, books, social networks, news. However, it is rare languages and dialects that are either poorly represented in the digital space, or do not have digitized corpora at all.

As a result: deep neural networks “learn” mainly on dominant/popular languages, and languages with a small number of speakers and low digital representation can be ignored or incorrectly processed, which leads to their marginalization in the digital environment.

Considering the above shortcomings and limitations of deep neural networks in machine linguistics, researchers are currently actively working on

efficient, compact and adaptive models that retain the advantages of scalable deep NNs, but reduce their cost.

Thus, modern data-driven machine linguistics was a revolution in the methods of analyzing, modeling and understanding human language. Instead of rigid rules, flexible and powerful models capable of automated learning and adaptation were used.

It was the data-driven approach that allowed us to bring language processing to a qualitatively new level, bringing the interaction between man and machine closer to natural dialogue.

However, with the growth of power and scale of data-driven models, questions arise: ethics, transparency, stability and interpretability. That is why the authors argue that the future of modern EFFECTIVE machine linguistics lies in a hybrid combination of the best technological solutions of the data-driven paradigm with theoretical linguistics and expert knowledge through elements of the knowledge-based paradigm.

C) So, as it is stated at the beginning of this section, this subsection will offer a set of methodological, technological and parametric scientific and practical recommendations for increasing the efficiency of both paradigms through their HYBRID use in order to obtain the SYNERGY effect.

Over the past decade, machine linguistics has undergone radical changes caused by the development of deep learning methods.

If earlier the main role in data-driven approaches was played by classical machine learning algorithms (such as support vectors, naive Bayesian classifier, logistic regression, etc.), now the leading position has been taken by neural network models, primarily architectures based on transformers.

Classical machine learning in NLP has demonstrated effectiveness in the tasks of text classification, tone detection, named entity recognition and thematic modeling.

Its advantages lie in the relative simplicity of implementation, unpretentiousness to computing resources, as well as the interpretability of the results. What is particularly valuable is that these methods remain competitive in data-limited or resource-poor domains.

However, with the growth of text corpora and computing power, neural network approaches have come to the fore.

In particular, deep learning models such as recurrent neural networks (RNN), bidirectional LSTMs, and especially transformers (BERT, GPT, T5), have demonstrated significant improvements in accuracy across a wide range of tasks, from machine translation to automatic text generalization and natural language generation. Due to their ability to contextually model language and automatically extract features, neural models significantly outperform classical

algorithms in complex tasks where understanding semantics and ambiguity is important.

In other words, in modern data-driven machine linguistics, neural network methods (specialized neural network architectures and corresponding deep machine learning algorithms)

have almost completely replaced classical machine learning, especially in large, complex problems.

However, classical ML still has its niche. Below are the results of a systematic analysis of these two DIFFERENT variants of data-driven machine linguistics.

Based on artificial neural networks and deep machine learning, data-driven machine linguistics has the following ADVANTAGES:

- high efficiency on complex tasks (machine translation, text generation, speech recognition, dialog systems);
- automatic feature extraction from large semi-structured and large unstructured data;
- contextual understanding (thanks to transformers (BERT, GPT) neuromodels learn context, semantics, ambiguity);
- scalability.

Based on artificial neural networks and deep machine learning, data-driven machine linguistics has the following DISADVANTAGES:

- high requirements for hardware resources (a lot of labeled/labeled QUALITATIVE data is required, powerful specialized GPUs, time to train models based on the architecture of deep neural networks);
- "black box" of the trained neural network (it is difficult (often Impossible/Inexpedient) to interpret how the model makes conclusions);
- error prone (in rare or specialized contexts, on small input data sets, in the presence of distortions/biases in the input data).

Based on CLASSIC machine learning methods/algorithms (SVM, Naive Bayes, Random Forest, Logistic & Parametric Regressions, Apriori etc.) data-driven machine linguistics has the following ADVANTAGES:

- simplicity and speed (learn quickly even on small data sets);
- interpretability (results are often easier to explain);
- resource-friendly (classical machine learning can be performed even without specialized GPUs);
- useful for basic NLP tasks (text classification, spam filtering, sentiment analysis on small corpora).

Based on CLASSIC machine learning methods/algorithms (SVM, Naive Bayes, Random Forest, Logistic & Parametric Regressions, Apriori etc) data-driven machine linguistics has the following DISADVANTAGES:

- limited understanding of context (work with fixed features (n-grams, frequencies, etc.));
- require manual creation of features (linguistic or domain knowledge is required);
- inferior to neural networks in complex tasks (generation, semantic analysis, QA systems, etc.).

Below, in the table, we present in comparison mode the most popular algorithms of classical and neural network ML for basic tasks of machine linguistics.

Table 1

Comparison of the most popular algorithms of classical and neural network ML for basic tasks of machine linguistics.

Main tasks of machine linguistics	Classical ML	Deep ML
Classification of texts	Naive Bayes, RF	RoBERTa, DistilBERT
Tonality analysis of corpus of text	SVM, Logistic Regression	BERT, LSTM
Machine translation	Phrase-Based SMT	Transformer (Google NMT)
Text generation	---	GPT-3, GPT-4, LLaMA
Named Entities recognition	CRF (earlier), SVM	BERT+CRF (better)

As an interim conclusion to the above, it is worth noting that in modern data-driven linguistics, neural network models dominate, especially transformers (BERT, GPT, T5, RoBERTa, mBERT, XLM-R, etc.).

However, 1). in low-resource conditions, in the absence of large corpora or for rapid prototyping, classical ML is still actively used; 2). in some industrial tasks, where speed and stability are important, old proven ML methods remain relevant.

Conclusions

1. Modern innovative philology is not only an update of methods, but also a change in the paradigm of humanitarian knowledge: from traditional hermeneutic analysis to hybrid, digital and creative ways of interpreting texts. It combines analytical and creative thinking, technology and humanitarian reflection, opening up new horizons for understanding language, culture and text in the era of digitalization.

Modern philology in conditions of instability ceases to be a "quiet humanitarian science", it becomes a tool for understanding, interpretation and reaction

to the crisis. Its task is not only to analyze texts, but also to form cultural stability, ethical sensitivity and linguistic memory of society.

Philology in the 21st century, faced with global crises and military conflicts, goes beyond classical textocentrism and becomes an active participant in the public understanding of traumas, historical memory and cultural transformations. It integrates humanities knowledge with digital and critical methods, striving to preserve the values of multilingualism.

2. Modern linguistics in the context of global crises ceases to be just an academic science: it becomes a relevant tool for analyzing social changes, responding to challenges, protecting linguistic rights and identity. Global imbalances push linguistics to openness, flexibility and ethical responsibility.

In times of global upheaval, machine linguistics becomes not only a technological tool for language processing, but also an active participant in social processes: from protecting the information space to assisting in humanitarian missions. Its development requires the combination of technical innovation, ethical responsibility and a deep understanding of cultural contexts.

3. Machine linguistics in modern philology acts not so much as an auxiliary technical discipline, but as a full-fledged element of the scientific paradigm, contributing to new forms of describing language, text and discourse. Its development paves the way for the creation of more accurate, representative and reproducible models of language and cultural processes, which makes it an integral part of the future of philological sciences.

Modern machine linguistics in philology: expands traditional methods of language analysis; allows working with larger volumes of texts; provides new perspectives for stylistics, poetics, historical linguistics; makes philology a more accurate and quantitative science.

4. Knowledge-based technologies in machine linguistics allow the creation of more accurate, deep, flexible, interpretable, semantically based systems for processing, analysis and further generation of a text.

Knowledge-based methods/algorithms are indispensable in situations where high accuracy, transparency of text interpretation, contextual analysis, translation, semantic parsing and intellectual analysis of conceptually complex texts are required - from legal and scientific to artistic and philosophical.

Knowledge-based technologies - are used both in technical tasks (translation, chatbots, summarization), and in humanitarian research (text analysis, stylistics, semantics, discourse).

Modern machine linguistics, based on these knowledge-based methods, is becoming a powerful tool for digital philology, cognitive science and the study of language as a system of knowledge in the modern dynamic conditions of global digital communications.

5. Data-driven approach in machine linguistics is a concept when natural language processing (NLP) is based on empirical data, mainly large corpora of texts, without explicit manual coding/formalization of linguistic rules/regularities/patterns. This approach is opposed to rule-based or knowledge-based strategies and the Data-driven approach is becoming dominant in modern applied linguistics and computational philology, especially given the aforementioned complex and complex external conditions of modern philology (and machine linguistics in particular).

Thus, Data-driven methods are the basis of modern machine linguistics and lie at the heart of the development of artificial intelligence technologies. Due to their scalability, automaticity and high performance, they are actively used both in academic research and in applied tasks - from machine translation to intelligent assis
tants.

In other words, it is the Data-driven approach in machine linguistics that has become the basis for modern text processing, where the model learns to understand language based on large arrays of examples. This has paved the way for scalable and powerful NLP systems - from Google Translate to ChatGPT - but has also required new and more complex approaches to ethics, interpretability, and linguistic accuracy. It is these limitations in interpretability and the need to make the most of big data that are driving the development of hybrid systems that combine data-driven and knowledge-based approaches.

6. Technologies based on Deep ANN are the basis of modern data-driven linguistics, they provide the best results, especially in complex and multi-context tasks.

Despite the obvious advantages of neural network approaches, classical machine learning has not lost its relevance.

Classical ML is still appropriate and useful in limited conditions or as a basic approach. It is still widely used in cases where fast, resource-saving processing is required or when the most TRANSPARENT and UNDERSTANDABLE interpretability/explanability of the model is required, for example, in legal or medical systems.

In other words, although modern data-driven machine linguistics is mainly based on neural network technologies, the final choice of methodology depends on a number of factors - in particular, the nature of the task, data availability, computing resources and requirements for model transparency.

The authors emphasize that the OPTIMAL approach is often a combination in modern machine linguistics of both these strategies (classical ML and neural network) within the framework of appropriate hybrid data-driven technologies and systems.

7. As noted above, Data-driven approaches dominate modern machine linguistics, but knowledge-based methods remain useful as support — especially for increasing accuracy, controllability, and interpretability.

In the era of global crises and military conflicts, none of the approaches — neither knowledge-based nor data-driven in modern machine linguistics — is universal. In the context of global crises and military conflicts, the difference, the differences & divergences between these approaches become especially significant, determining the effectiveness, adaptability, and ethical capacity of linguistic technologies.

The future of modern machine linguistics in conditions of crises/instabilities/cataclysms/wars lies precisely in hybrid technologies that SYNERGETICALLY combine both the potential power of accumulated data and the depth of expert tacit knowledge/experience. After all, effective linguistic solutions require a BALANCE of interpretability and scalability, ethical responsibility and technological adaptability. For example: - post-processing of DeepNN model results using linguistic rules; - use of dictionaries, morphological analyzers or semantic ontologies (such as WordNet) to improve the quality of machine translation, etc.

8. Moreover, the authors argue that under the influence of Big Data, Machine Learning and AI technologies, modern linguistics in the 21st century requires a transformation not only of methods, but also of values - towards their greater openness, interdisciplinarity and digital humanitarian ethics.

References:

1. Tuhaienko, V., & Krasniuk, S. (2022). Effective application of knowledge management in current crisis conditions. *Grail of Science*, (16), pp. 348–358

2. Naumenko, M. (2024). Modeli biznesovykh znan v systemakh shtuchnoho intelektu dlia efektyvnoho konkurentnoho pidpryiemstva [Models of business knowledge in artificial intelligence systems for an effective competitive enterprise]. Mizhnarodnyi naukovyi zhurnal "Internauka". Seriia: "Ekonomichni nauky". – *International scientific journal "Internauka". Series:"Economic Sciences"*. № 6.DOI: https://doi.org/10.25313/2520-2294-2024-6-10010

3. Sytnyk V.F., Krasniuk M.T. (2002) Polityka upravlinia znanniamy naftohazovoi kompanii yak kliuchovyi faktor pidvyshchennia yii efektyvnosti [Oil and gas company's knowledge management policy as a key factor in increasing its efficiency]. *Problemy formuvannia rynkovoi ekonomiky - Problems of the formation of a market economy*, K.:KNEU, vol. 10, 2002. [in Ukrainian].

4. Krasnyuk M.T. (2006) Problemy zastosuvannia system upravlinnia korporatyvnymy znanniamy ta yikh taksonomiia [Problems of applying corporate knowledge management systems and their taxonomy]. *Modeliuvannia ta informatsiini systemy v ekonomitsi - Modeling and information systems in the economy*, vol. 73, p. 256 [in Ukrainian].

5. Krasnyuk, M., & Krasniuk, S. (2020). Application of artificial neural networks for reducing dimensions of geological-geophysical data set's for the identification of perspective oil and gas deposits. *Scientific bulletin ΛΟΓΟΣ*, 18-19. https://doi.org/10.36074/24.04.2020.v2.05.

6. Kulynych Y., Krasnyuk M. (2022). Efektyvna intehratsiia oriientovanoi na znannia kontseptsii biznes-pravyl v ramkakh upravliaiuchoi intelektualnoi korporatyvnoi informatsiinoi systemy [Effective integration of a knowledge-based business rules concept within a managed intelligent corporate information system]. *Current state and prospects for the development of enterprises in Ukraine : theory, methodology, practice* : a collective monograph. Poltava: PC «Astraya», 2022. – C. 73-79 [in Ukrainian].

7. M. Krasnyuk, S. Goncharenko, S. Krasniuk (2022) Intelektualni tekhnolohii v hibrydnii korporatyvnii SPPR (na prykladi Ukrainskoi naftohazovydobuvnoi kompanii) [Intelligent technologies in hybrid corporate DSS (on the example of Ukraine oil&gas production company)] *Innovatsiino-investytsiinyi mekhanizm zabezpechennia konkurentospromozhnosti krainy:* kolektyvna monohrafiia / za zah. red. O. L. Haltsovoi - *Innovation and investment mechanism for ensuring the country's competitiveness:* collective monograph / by general ed. O. L. Khultsova. – Lviv-Torun: League-Pres, 2022. – pp. 194-211 [in Ukrainian].

8. Krasnyuk, M. (2014). Hibrydyzatsiia intelektualnykh metodiv analizu biznesovykh danykh (rezhym vyiavlennia anomalii) yak skfladovyi instrument korporatyvnoho audytu [Hybridization of intelligent methods of business data analysis (anomaly detection mode) as a standard tool of corporate audit]. *Stan i perspektyvy rozvytku oblikovo-informatsiinoi systemy v Ukraini - The state and prospects of the development of the accounting and information system in Ukraine: materials of the III International science and practice conf*. [m. Ternopil, October 10-11. 2014]. TNEU, 2014. pp. 211-212 [in Ukrainian].

9. Ahmad, A., Abbasi, I. A., Abbasi, R. H., & Rasheed, B. (2025). Exploring the Intricate Relationship between Semantics and Computational Linguistics. *Liberal Journal of Language & Literature Review*, *3*(1), 164-181.

10. Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. *Frontiers in Artificial Intelligence*, *4*, 625341.

11. Gatla, T. R. (2024). A Groundbreaking Research in Breaking Language Barriers: NLP And Linguistics Development. *International Journal of Advanced Research and Interdisciplinary Scientific Endeavours*, *1*(1), 1-7.

12. Jalilbayli, O. B. (2022). Forecasting the prospects for innovative changes in the development of future linguistic education for the XXI century: the choice of optimal strategies. *Futurity Education*, *2*(4), 36-43.

13. Li, W. (2021). Role of machine learning and artificial intelligence algorithms for teaching reform of linguistics. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 3251-3262.

14. McShane, M., & Nirenburg, S. (2021). *Linguistics for the Age of AI*. Mit Press.

15. Stepanova, I. S., Nykyporets, S. S., & Hadaichuk, N. M. (2024). Exploring the evolving dynamics of axiological concepts in the modern linguistic space: a comprehensive scientific analysis. *Publishing House "Baltija Publishing"*.

16. Tasheva, N. (2024). The Evolution of Modern Linguistics: Key Concepts and Trends. *Medicine, pedagogy and technology: theory and practice*, 2(11), 31-39.

17. Becker, M., Hulpuş, I., Opitz, J., Paul, D., Kobbe, J., Stuckenschmidt, H., & Frank, A. (2020). Explaining arguments with background knowledge: Towards knowledge-based argumentation analysis. *Datenbank-Spektrum*, *20*, 131-141.

18. Oussalah, M., & Mohamed, M. (2022). Knowledge-based sentence semantic similarity: algebraical properties. *Progress in Artificial Intelligence*, *11*(1), 43-63.

19. Papegaaij, B. C. (2019). *Word expert semantics: an interlingual knowledge-based approach* (Vol. 1). Walter de Gruyter GmbH & Co KG.

20. Su, C., Wu, K., & Chen, Y. (2021, August). Enhanced metaphor detection via incorporation of external knowledge based on linguistic theories. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 1280-1287).

21. Wątróbski, J. (2020). Ontology learning methods from text-an extensive knowledge-based approach. *Procedia Computer Science*, *176*, 3356-3368.

22. Akstinaite, V., Garrard, P., & Sadler-Smith, E. (2022). Identifying linguistic markers of CEO hubris: A machine learning approach. *British Journal of Management*, *33*(3), 1163-1178.

23. Bambini, V., Frau, F., Bischetti, L., Cuoco, F., Bechi, M., Buonocore, M., ... & Bosia, M. (2022). Deconstructing heterogeneity in schizophrenia through language: a semi-automated linguistic analysis and data-driven clustering approach. *Schizophrenia*, *8*(1), 102.

24. Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time.

25. Gómez-Vilda, P., & Gómez-Rodellar, A. (2024). Data-Driven Vs Model-Driven Approaches in Cognitive Speech Processing. *In The Theory of Mind Under Scrutiny: Psychopathology, Neuroscience, Philosophy of Mind and Artificial Intelligence* (pp. 649-680). Cham: Springer Nature Switzerland.

26. Kaur, K., Chahar, S., Manwal, M., Kaswan, K. S., & Dhatterwal, J. S. (2024, July). Bridging Linguistics and Machine Learning: A New Human-Machine Communication. In *2024 1st International Conference on Sustainable Computing and Integrated Communication in Changing Landscape of AI (ICSCAI)* (pp. 1-6). IEEE.

27. Li, C. C., Dong, Y., Liang, H., Pedrycz, W., & Herrera, F. (2022). Data-driven method to learning personalized individual semantics to support linguistic multi-attribute decision making. *Omega*, *111*, 102642.

28. Shawaqfeh, A. T., Jadallah Abed Khasawneh, Y., Khasawneh, A. J., & Khasawneh, M. A. S. (2024). Data-Driven Language Assessment in Multilingual Educational Settings: Tools and Techniques for Proficiency Evaluation. *Migration Letters*, *21*(2), 701-711.

29. Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. *arXiv preprint arXiv:2102.00287*.

30. Varda, A. G. D., & Marelli, M.(2023).Data-driven cross-lingual syntax: An agreement study withmassively multilingual models.*Computational Linguistics*,*49*(2),261-299.

Література:

1. Tuhaienko, V., & Krasniuk, S. Effective application of knowledge management in current crisis conditions //Міжнародний науковий журнал *Grail of Science*, (16), С. 348–358

2. Науменко М.А. Моделі бізнесових знань в системах штучного інтелекту для ефективного конкурентного підприємства // *Міжнародний науковий журнал «Інтернаука». Серія: «Економічні науки».* 2024. №6. DOI: https://doi.org/10.25313/2520-2294-2024-6-10010

3. Ситник В.Ф., Краснюк М.Т. Політика управління знаннями нафтогазової компанії як ключовий фактор підвищення її ефективності // *Проблеми формування ринкової економіки*: Міжвідомчий науковий збірник. Заснов. у 1992 р. Вип. 10 / Відп. ред. О.О. Беляєв. – К.: КНЕУ, 2002.

4. Краснюк М.Т. Проблеми застосування систем управління корпоративними знаннями та їх таксономія // *Моделювання та інформаційні системи в економіці*: міжвід. наук. зб. Відп. ред. В.К. Галіцин. Київ : КНЕУ, 2006. Вип. 73. С. 34–44.

5. Krasnyuk, M., & Krasniuk, S. Application of artificial neural networks for reducing dimensions of geological-geophysical data set's for the identification of perspective oil and gas deposits // *Збірник наукових праць ΛΟГОΣ*, 18-19. https://doi.org/10.36074/24.04.2020.v2.05

6. Кулинич Ю. М., Краснюк М. Т. Ефективна інтеграція орієнтованої на знання концепції бізнес-правил в рамках управляючої інтелектуальної корпоративної інформаційної системи // *Current state and prospects for the development of enterprises in Ukraine*: theory, methodology, practice. – 2022. – С. 73-79.

7. Krasnyuk M., Goncharenko S., Krasniuk S. Intelligent technologies in hybrid corporate DSS (on the example of Ukraine oil&gas production company) // *Інноваційно-інвестиційний механізм забезпечення конкурентоспроможності країни*: колективна монографія / за заг. ред. О. Л. Гальцової. Львів-Торунь : Ліга-Прес, 2022. С. 194–211.

8. Краснюк М.Т. Гібридизація інтелектуальних методів аналізу бізнесових даних (режим виявлення аномалій) як складовий інструмент корпоративного аудиту // *Стан і перспективи розвитку обліково-інформаційної системи в Україні* : матеріали III Міжнар. наук.-практ. конф. [м. Тернопіль, 10-11 жовт. 2014 р.] / редкол. : З. В. Задорожний, В. А. Дерій, М. Р. Лучко [та ін.] ; гол. ред. З. В. Задорожний. Тернопіль : ТНЕУ, 2014. С. 211–212.

9. Ahmad A. et al. Exploring the Intricate Relationship between Semantics and Computational Linguistics // *Liberal Journal of Language & Literature Review*. – 2025. – Т. 3. – №. 1. – С. 164-181.

10. Church K., Liberman M. The future of computational linguistics: On beyond alchemy // *Frontiers in Artificial Intelligence*. – 2021. – Т. 4. – С. 625341.

11. Gatla T. R. et al. A Groundbreaking Research in Breaking Language Barriers: NLP And Linguistics Development // *International Journal of Advanced Research and Interdisciplinary Scientific Endeavours*. – 2024. – Т. 1. – №. 1. – С. 1-7.

12. Jalilbayli O. B. Forecasting the prospects for innovative changes in the development of future linguistic education for the XXI century: the choice of optimal strategies // *Futurity Education*. – 2022. – Т. 2. – №. 4. – С. 36-43.

13. Li W. Role of machine learning and artificial intelligence algorithms for teaching reform of linguistics // *Journal of Intelligent & Fuzzy Systems*. – 2021. – Т. 40. – №. 2. – С. 3251-3262.

14. McShane M., Nirenburg S. Linguistics for the Age of AI. – *Mit Press*, 2021.

15. Stepanova I. S., Nykyporets S. S., Hadaichuk N. M. Exploring the evolving dynamics of axiological concepts in the modern linguistic space: a comprehensive scientific analysis // *Publishing House "Baltija Publishing"*. – 2024.

16. Tasheva N. The Evolution of Modern Linguistics: Key Concepts and Trends // *Medicine, pedagogy and technology: theory and practice*. – 2024. – Т. 2. – №. 11. – С. 31-39.

17. Becker M. et al. Explaining arguments with background knowledge: Towards knowledge-based argumentation analysis //Datenbank-Spektrum. – 2020. – Т. 20. – С. 131-141.

18. Oussalah M., Mohamed M. Knowledge-based sentence semantic similarity: algebraical properties //Progress in Artificial Intelligence. – 2022. – Т. 11. – №. 1. – С. 43-63.

19. Papegaaij B. C. Word expert semantics: an interlingual knowledge-based approach. – Walter de Gruyter GmbH & Co KG, 2019. – Т. 1.

20. Su C., Wu K., Chen Y. Enhanced metaphor detection via incorporation of external knowledge based on linguistic theories //Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. – 2021. – С. 1280-1287.

21. Wątróbski J. Ontology learning methods from text-an extensive knowledge-based approach //Procedia Computer Science. – 2020. – Т. 176. – С. 3356-3368.

22. Akstinaite V., Garrard P., Sadler-Smith E. Identifying linguistic markers of CEO hubris: A machine learning approach //British Journal of Management. – 2022. – Т. 33. – №. 3. – С. 1163-1178.

23. Bambini V. et al. Deconstructing heterogeneity in schizophrenia through language: a semi-automated linguistic analysis and data-driven clustering approach //Schizophrenia. – 2022. – Т. 8. – №. 1. – С. 102.

24. Boulton A., Vyatkina N. Thirty years of data-driven learning: Taking stock and charting new directions over time. – 2021.

25. Gómez-Vilda P., Gómez-Rodellar A. Data-Driven Vs Model-Driven Approaches in Cognitive Speech Processing // The Theory of Mind Under Scrutiny: Psychopathology, Neuroscience, Philosophy of Mind and Artificial Intelligence. – Cham : Springer Nature Switzerland, 2024. – С. 649-680.

26. Kaur K. et al. Bridging Linguistics and Machine Learning: A New Human-Machine Communication //2024 1st International Conference on Sustainable Computing and Integrated Communication in Changing Landscape of AI (ICSCAI). – IEEE, 2024. – С. 1-6.

27. Li C. C. et al. Data-driven method to learning personalized individual semantics to support linguistic multi-attribute decision making //Omega. – 2022. – Т. 111. – С. 102642.

28. Shawaqfeh A. T. et al. Data-Driven Language Assessment in Multilingual Educational Settings: Tools and Techniques for Proficiency Evaluation //Migration Letters. – 2024. – Т. 21. – №. 2. – С. 701-711.

29. Vanmassenhove E., Shterionov D., Gwilliam M. Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation //arXiv preprint arXiv:2102.00287. – 2021.

30. Varda A. G., Marelli M. Data-driven cross-lingual syntax: An agreement study with massively multilingual models //Computational Linguistics. – 2023. – Т. 49. – №. 2. – С. 261-299.