Acta et Commentationes, Exact and Natural Sciences
Volume 20(2), 2025, Pages 116–127
CZU: 517.925
ISSN: 2537-6284
E-ISSN: 2587-3644
DOI: 10.36120/2587-3644.v20i2.116-127

Semantic non-compositional exploration of Bessarabian idioms by LLMs

Inga Titchiev, Olesea Caftanatov, and Crinu Dimitriu

Abstract. The study explores the potential of LLMs in interpreting and translating Bessarabian idioms. The central problem addressed is the semantic non-compositionality of idiomatic expressions, which poses a significant challenge for Natural Language Processing since their figurative meaning cannot be derived from literal components. As part of the CI ARiA project, 1000 proverbs were digitized, using a corpus of 400 of them to evaluate the performance of 10 AI models (such as ChatGPT, Gemini, Grok). The methodology is multi-algorithmic, combining textual distance metrics (Levenshtein, Jaccard) with semantic similarity analysis via Sentence Transformers. The results indicate that while models demonstrate a solid capacity to grasp metaphorical meanings, significant differences exist regarding consistency and explanatory style.
2020 Mathematics Subject Classification: 68U15, 68T30.
Keywords: Large Language Models (LLMs), idioms, semantic non-compositionality.

Explorarea non-compoziţională semantică a idiomurilor basarabene de către LLM-uri

Rezumat. Lucrarea explorează potențialul LLM-urilor în interpretarea și traducerea idiomurilor basarabene. Problema centrală abordată este non-compoziționalitatea semantică a expresiilor idiomatice, care reprezintă o provocare majoră pentru procesarea limbajului natural deoarece sensul lor figurat nu poate fi dedus din componentele literale. În cadrul proiectului CI ARiA, au fost digitalizate 1000 de proverbe, utilizând un corpus de 400 dintre ele pentru a evalua performanța a 10 modele AI (precum ChatGPT, Gemini, Grok). Metodologia utilizată este multi-algoritmică, combinând metrici de distanță textuală (Levenshtein, Jaccard) cu analiza similitudinii semantice prin Sentence Transformers. Rezultatele indică faptul că, deși modelele demonstrează o capacitate solidă de a înțelege sensurile metaforice, există diferențe semnificative în ceea ce privește consistența și stilul explicativ.
Cuvinte-cheie: Modele de Limbaj Mari (LLM), idiomuri, non-compoziționalitate semantică.

1. Introduction

In recent years, the accelerated development of Generative Artificial Intelligence (GenAI) has fundamentally shifted the paradigm of how cultural, linguistic, and visual information is created, interpreted, and analyzed. Large Language Models (LLMs) and deep learning-based image generation systems have evolved from experimental tools into core technologies applied in education, research, and the creative industries. In this context, the intersection of traditional cultural heritage and modern AI technologies has emerged as a critical area of exploration.

The central problem addressed in this study is the complexity of idiomatic expressions within computational linguistics. Unlike standard sentence structures, idioms represent systematic deviations from the literal meaning of language. Most automated models, whether for machine translation, semantic analysis, or text generation, rely heavily on the principle of semantic compositionality, assuming that the meaning of a sentence can be derived from the meanings of its constituent words and their grammatical combination. However, this principle fails in the context of idiomatic expressions. As Multiword Expressions (MWEs), idioms possess figurative meanings that are stable and culturally recognized but cannot be deduced from the literal definitions of their components (e.g., the Bessarabian idiom ”a da ortul popii” literally means ”to give the coin to the priest,” but figuratively signifies ”to die”).

As noted by Iordache Golescu [1], “Vorba fără glume pare ca bucatele fără sare, iar pildele la vorbă, tocma ca nişte glume (Speech without humor is like food without salt, and idioms in conversation serve the very same purpose as jokes)”, idioms function like ”jokes that carry wisdom,” which are essential for the flavor and nuance of natural communication. Without them, language risks becoming ”unsalted” or flat. Consequently, identifying and accurately interpreting these expressions presents a significant challenge for Natural Language Processing (NLP). This challenge is particularly acute for under-resourced or regionally specific dialects, such as the Moldavian idioms characteristic of the Bessarabian region.

This research is part of the national project CI ARiA that aims to bridge the gap between cultural heritage and technological advancement. The project focuses on the digitization and augmentation of regional idiomatic expressions from the Republic of Moldova, intending to facilitate deep language learning and promote indigenous linguistic heritage. The project involves collecting and classifying approximately 10,000 tokens and creating interactive educational experiences using Artificial Intelligence and Augmented Reality (AR).

Specifically, this paper analyzes the ”state of the art” of current LLMs, examining their capacity to comprehend, translate, and visualize the semantic depth of Bessarabian idioms. By evaluating how these models process the figurative meanings of idioms, we provide insights into the limitations and strengths of current AI systems in handling deep semantic and cultural concepts.

2. The nature and taxonomy of idioms

In computational linguistics and lexicography, idiomatic expressions are generally categorized under the broader umbrella of multiword expressions. According to the Cambridge Dictionary [2] an **idiom** is a group of words in a fixed order that has a particular meaning different from the meanings of each word on its own. An idiom is traditionally defined as a lexical unit consisting of two or more words, characterized by semantic non-compositionality. This means that the overarching sense of the expression cannot be derived simply by aggregating the literal meanings of its constituent parts.

From a structural and semantic perspective, idioms are defined by three primary characteristics. First, they possess **figurative meaning**, where the relationship between the literal components and the actual significance is metaphorical or metonymic. For example, the Bessarabian expression “*a sparge gheața*” (to break the ice) refers to initiating social interaction, without involving any physical act of breaking frozen water. Second, idioms exhibit **syntactic frozenness** (fixity), which means that they resist grammatical alteration or lexical substitution. One cannot replace components with synonyms without losing the idiomatic sense; for example, modifying “*Dintr-un bob răsar o mie*” (From a grain, a thousand sprouts) to “*Dintr-un bob, o grămadă*” (From a grain, a heap) destroys the established resonance of the phrase. Finally, **institutionalization** requires that the expression be recognized by the linguistic community as a single semantic unit. A clear example is “*a tăia frunză la câini*” (literally: cutting leaves for dogs), which is universally understood as ”wasting time.” In contrast, a constructed phrase like “*a tăia iarbă la pisici*” (to cut grass for cats)—though grammatically valid and semantically parallel—carries - does not have meaning because the community has not encoded it. Similarly, altering “*a fi de râsul găinilor*” (to be the laughingstock of hens) to “*gâștelor*” (geese) breaks the specific cultural code, rendering the phrase without its idiomatic impact.

Folklore researchers have proposed various taxonomies for idioms based on their degree of opacity and grammatical structure. A widely accepted classification distinguishes between four primary categories, ranging from completely opaque to statistically probable. At one end of the spectrum are **pure or opaque idioms**, expressions in which the link between the literal and figurative meaning is either severed or historically obscured.

In the Moldavian linguistic context, a prime example is "*la paștele cailor*" (literally: at the Easter of Horses); thus the constituent words offer no logical clue to the figurative meaning of "*never*" or "*an impossible time*." Similarly, "*a spăla putina*" (to wash the barrel) is semantically opaque, as the literal action of cleaning a vessel has no intrinsic connection to the act of fleeing, except through a lost historical context.

Moving toward greater transparency, **semi-idioms** (or metaphorical idioms) allow the figurative meaning to be partially inferred through cognitive mapping. For example, the Bessarabian expression "*a pune paie pe foc*" (to put straw on the fire) operates on a clear metaphorical logic where the physical act of intensifying a flame maps directly onto the abstract concept of aggravating a conflict.

Distinct from these phrase-level structures are **proverbs and sayings**, which function as complete sentences conveying general truths, moral lessons, or advice. This category is particularly relevant to the Bessarabian corpus, as proverbs often encapsulate the collective experience and archaic wisdom of the community. An illustrative example is "*apa trece, pietrele rămân*" (water passes, stones remain), which conveys the permanence of fundamental values amidst transient events, or "*nu da vrabia din mână pe cioara de pe gard*" (do not trade the sparrow in your hand for the crow on the fence), which advises against risking a certainty for a larger but uncertain gain.

Finally, the taxonomy includes **collocations**, defined as words that statistically co-occur more often than chance. While these exist on the boundary of idiomaticity, they are crucial for natural fluency. In the Moldovan linguistic context, one observes specific pairings such as "*ploaie torențială*" (torrential rain) rather than "*ploaie puternică*" (strong rain), demonstrating a fixed lexical preference despite the semantic interchangeability of the adjectives.

A distinct and fascinating category of idiomatic expressions is one that encodes encodes empirical knowledge of the physical world. Before the formalization of the scientific method, communities relied on acute observation and pattern recognition to survive, eventually crystallizing these insights regarding gravity, thermodynamics, optics, and causality into idioms and proverbs to ensure their transmission across generations. In cognitive linguistics, this phenomenon is framed within Conceptual Metaphor Theory [5], where a source domain rooted in physical reality is mapped onto a target domain representing abstract social or psychological situations. Consequently, the source domain often reveals a surprisingly accurate intuitive understanding of natural laws, frequently referred to as "**Naïve Physics**" or "**Folk Physics**."

This encoding of natural laws is evident across various scientific domains. Principles of thermodynamics and material science, for instance, are preserved in the proverb "*bate*

fierul cât e cald" (strike while the iron is hot) a direct instruction derived from metallurgy regarding the increased malleability of metals at high temperatures, applied metaphorically to the seizing of fleeting opportunities. Similarly, the Moldavian proverb "*aurul nu se unește nici cu fierul, nici cu oțelul* " (gold unites with neither iron nor steel) reflects an intuitive understanding of chemical properties and the distinct nature of noble metals; it uses the physical resistance of gold to alloy with base metals to metaphorically describe the fundamental incompatibility between moral nobility and common baseness.

Furthermore, the inescapability of gravitational force is captured in the idiom "*what goes up must come down*" which applies the laws of physics to the rise and fall of power or fortune, while "*Mărul nu cade departe de pom*" (The apple doesn't fall far from the tree) or "*Așchia nu sare departe de trunchi*" (The woodchip does not jump far from the trunk) relies on the observation of vertical trajectories to explain hereditary traits.

Concepts of causality and fluid dynamics are equally prevalent. The idiom "*Fără foc nu iese fum*" (There is no smoke without fire) inextricably links the observable effect of smoke to its necessary cause, combustion. Finally, the behavior of water provides a compelling case of cross-cultural divergence in physical observation. Although the English proverb "*Still waters run deep*" correctly contrasts laminar flow with turbulence to suggest hidden complexity (a state of being), its Moldavian equivalent, "*Apa lină sapă adânc*" (Smooth water digs deep), emphasizes the erosive power and persistence of the quiet force (an action), perfectly aligning with the geological principles of erosion.

Therefore, idioms are not merely linguistic ornaments, but serve as a repository of a community's empirical interaction with the physical universe, preserving observation-based truths that predate formal scientific classification. Despite substantial progress, LLMs continue to face notable challenges in processing idiomatic language. In particular, they often struggle not only to recognize the empirical natural laws and phenomena encoded in such expressions but, more importantly, to map these elements onto their intended figurative meanings. In light of these challenges, we aim to assess the proficiency of state-of-the-art LLMs in interpreting Bessarabian idioms, testing whether current model architectures can successfully navigate this semantic complexity.

3. Methodology

To assess the semantic alignment between linguistic comprehension and the visual representation of idioms, this study employed a cross-modal experimental design. The methodology combines quantitative textual analysis with qualitative visual evaluation to determine how AI systems navigate the non-compositional nature of Moldavian idioms.

3.1. Data collection and corpus

The primary source for this study was the volume of G. Botezatu et al[3]. From this volume, a dataset of 1000 representative Bessarabian idioms was manually digitized and curated [4]. Within the selected dataset, the corpus includes both idioms characteristic of the Bessarabia region and a significant number of common proverbs also used in Romania, thus this broader linguistic space provides a diverse framework for exploring the semantic non-compositionality of cultural heritage.

These idioms were selected based on their cultural relevance, linguistic complexity, and distinct figurative nature. From these 1000 proverbs, 400 were used to analyze 10 LLMs such as: ChatGPT, Grok, Gemini, DeepSeek, Mistral, Qwen, Alibaba Cloud, Copilot, Perplexity, Claude.

The research was structured into two distinct but interconnected experiments:

(1) **Textual analysis (LLM evaluation)**: Measuring the capacity of generative models to complete, explain, and translate idioms.
(2) **Multimodal visualization:** Investigating how text-to-image models interpret figurative prompts derived from the idioms.

To rigorously assess the performance of the selected models, the experimental workflow was divided into four distinct phases: generation, structuring, algorithmic evaluation, and aggregation.

(1) **Multi-model generation** Each proverb in the dataset was processed individually using ten distinct Large Language Models ChatGPT, Grok, Gemini, DeepSeek, Mistral, Qwen, Alibaba Cloud, Copilot, Perplexity, Claude to generate a comprehensive set of linguistic outputs. For every entry, the models were prompted to provide:
    - **Idiom completion**: Generating the missing segment of the proverb (either the beginning or the ending) to test knowledge recall.
    - **Generate explanation**: When LLMs generate the missing segment of a proverb, we also request an explanation in order to understand the rationale behind the generated continuation and how the figurative meaning is interpreted within the native cultural context.
    - **English interpretations**: Providing both the literal translation and a semantic equivalent (or similar English idiom).
    - **Generate explanation of original idiom**: Explaining the meaning of the original idiom allows us to verify whether the idiom has been generated in its correct form; a strong alignment between this explanation and the explanation

provided during the generation phase indicates deep comprehension. All steps are presented in Figure 1.

	A	B	C	D	E	F	
1	Proverbul original	ChatGPT continuare	ChatGPT - explicatie	ChatGPT EN - sens literar	ChatGPT EN - sens similar	ChatGPT - Explicatie original	Grok - contin
2	Turcul te bate, turcul te judecă.	Turcul te bate, dar cu sabia ta.	Se spune despre situațiile în care cineva e învins sau înșelat cu cruzime mărace.	The Turk beats you, the Turk judges you.	The judge and the executioner should never be the same man.	Se referă la nedreptatea sistemului: cei care îți fac răul este tot el și cel care te „judecă". F un proverb despre abuzul de putere și	Turcul te bate, ceată.
3	Țara moare de tătari, și el bea cu lăutari.	Țara moare de tătari, iar domnii se cinstesc.	Reflectă ipocrizia și nepăsarea celor aflați la putere: poporul suferă și piere, în timp ce conducătorii se destată și petrec.	The country dies from the Tartars, and he drinks with the fidders.	While Rome burns, Nero plays the fiddle. The rich feast while the poor starve. When the ship sinks, the	Descrie indiferența și nepăsarea celor care petrec și se bucură în timp ce țara sau comunitatea suferă.	Țara moare de certăm pe cioc
4	Boieri țin fumul de pe urma săracului.	Boierii țin fumul, iar săracii cenușa.	Proverbul vorbește despre nedreptatea socială — bogații se bucură de partea frumoasă și ușoară, iar săracii rămân cu resturile, cu greul.	The boyars keep the smoke from the poor man's fire.	The rich get richer, the poor get poorer. Where there's smoke, there's a landlord. The master eats the meat, the servant gnaws the bone.	O ironie amară despre exploatarea celor de jos: boierii (cei bogați) iau până și fumul, adică nimicul, de la cel sărac.	Boieri țin fumu scrumul.

Figure 1. Sample of the Collected Data

(2) **Structured data organization** The raw outputs generated by each AI model were systematically organized into a structured CSV format. This step ensured precise alignment between the original proverb (ground truth) and the respective model responses, establishing a clean dataset for comparative analysis.

(3) **Automated comparative analysis (Python implementation)** To evaluate the accuracy and fidelity of the generated texts, a custom Python script was implemented. This script compared the original proverbs against the AI-generated content using an extensive suite of textual and semantic similarity algorithms, specifically:

- **Normalized Levenshtein distance**: To measure character-level edit distance.
- **Jaccard similarity**: To assess the overlap of token sets between texts.
- **SequenceMatcher**: To identify the longest contiguous matching sequences.
- **Jaro-Winkler similarity**: To evaluate string similarity with a bias toward matching prefixes
- **TF-IDF + Cosine similarity**: To measure the angle between vector representations of the texts based on term frequency.
- **Longest common subsequence (LCS)**: To verify the structural integrity of the recalled idioms.
- **Semantic similarity (Neural embeddings)**: Utilizing a multilingual **Sentence transformers** model to evaluate meaning preservation beyond mere lexical overlap.

(4) **Data aggregation and scoring** The resulting scores for each algorithm were collected and aggregated into a unified data structure (CSV/JSON). This consolidation facilitated a robust quantitative analysis, allowing for a statistical comparison of how each model navigates the complexities of Bessarabian idiomatic language.

4. Results analysis using text comparison algorithms

Following the collection and structuring of the Bessarabian idioms corpus, the next stage of the research focused on a comparative analysis of the results generated by various Artificial Intelligence models. The primary objective of this stage was to evaluate the degree of fidelity, consistency, and semantic understanding of the proverbs in relation to their traditional forms and established meanings in Bessarabian popular culture.

Given that idioms are short, semantically dense, and highly figurative linguistic constructions, a simple visual or subjective comparison would have been insufficient. Therefore, the analysis was conducted using text comparison algorithms capable of capturing both formal similarity (at the character and word level) and semantic proximity (meaning, intention, message).

The chosen approach is multi-algorithmic, as no single algorithm can capture all dimensions of a proverb: the exact form, syntactic structure, vocabulary used, and, especially, the figurative meaning. By combining several methods, the results obtained become more robust and their interpretation more credible.

4.1. Algorithms used

Several established algorithms in the field of Natural Language Processing (NLP) were used, each with a well-defined role in the comparative analysis:

String-based algorithms (such as Levenshtein, Jaro-Winkler, or SequenceMatcher) were used to evaluate similarity at the character level and textual order. These are particularly useful for identifying minor variations, spelling differences, or reformulations very close to the original.

Word-based algorithms (Jaccard, Word Overlap) allowed measuring vocabulary overlap while ignoring word order. This approach is relevant for proverbs, where meaning can be preserved even if the phrase structure is modified.

Vector and statistical methods, such as Cosine Similarity using TF-IDF, were used to evaluate the relative importance of terms in each idiom, reducing the influence of very frequent words, and highlighting key concepts.

Deep semantic similarity algorithms, based on pre-trained neural models (Sentence Transformers), completed the analysis. These allow for comparing the actual meaning

of texts, even when there is no obvious lexical overlap, being essential for interpreting figurative meanings of idioms and paraphrases.

By using these algorithms in combination, the analysis is not limited to ”how similar the texts are,” but responds more deeply to the question: do they convey the same meaning?

4.2. Results comparison and metrics interpretation

To evaluate the performance of AI models in completing, explaining and translating proverbs, the results were compared using an extensive set of quantitative metrics, grouped into thematic categories:

Proverb completion accuracy (Figure 2): It analyzes how correctly the original proverb is reproduced. The exact match indicator represents cases where the model generated the established form of the proverb identically. The fact that ChatGPT, Grok, and Gemini obtain the same number of exact matches suggests a similar level of familiarity with the canonical forms.

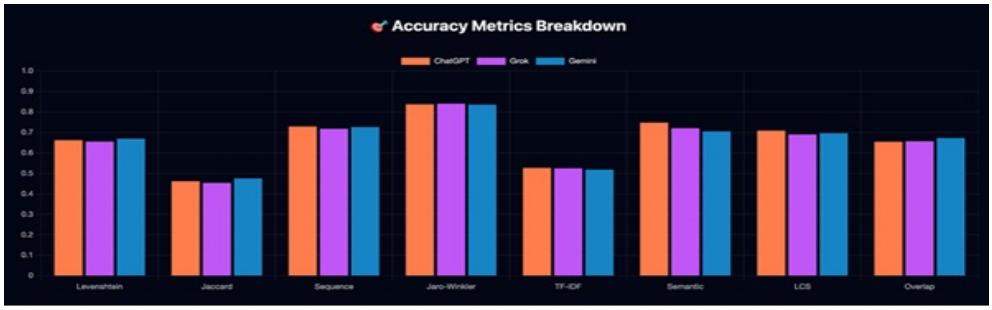

Figure 2. Accuracy Metrics

Textual similarity metrics: For non-identical reproductions, Normalized Levenshtein measures necessary modifications, Jaccard calculates word set overlap (reflecting synonym usage), and Sequence Matcher/LCS ratio analyzes the continuity of sequences to indicate how much of the original structure is preserved. Jaro-Winkler is more lenient with minor differences and favors prefix matches.

Semantic similarity: Essential for idioms, cosine similarity with TF-IDF evaluates the statistical proximity of important terms, while the semantic score (SentenceTransformer) measures the actual proximity of meaning. High values here indicate that models largely succeed in preserving the central idea of the proverb.

Length and style statistics: Metrics such as average word count and character count show the stylistic consistency of the completions.

Explanation quality (Figure 3): The consistency score reflects how well the explanation aligns with the idiom's meaning. Complexity metrics (length, number of sentences, readability) show the explanatory style: some models are concise, others more elaborate. A low word overlap score is a positive result here, as a good explanation interprets rather than repeats the idiom.

Figure 3. Explanation Quality

Inter-LLM and cross-comparisons (Figure 4): It analyzes convergence between models (e.g., ChatGPT vs. Gemini). High semantic scores versus lower lexical scores indicate agreement on meaning rather than wording

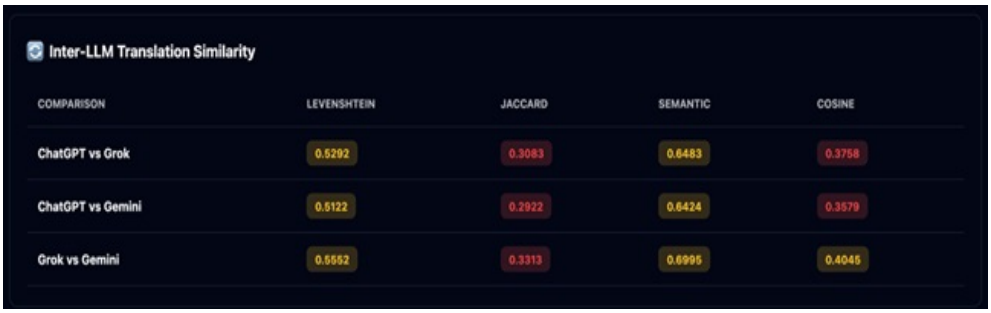

Figure 4. Cross-Comparisons

The comparative analysis of the results generated by the ChatGPT, Grok and Gemini models highlights that all three tools demonstrate a solid capacity to work with Bessarabian idioms, both at the formal and semantic level.

However, the differences between them (Figure 5) become visible when the results are viewed in detail, through the prism of the used metrics.

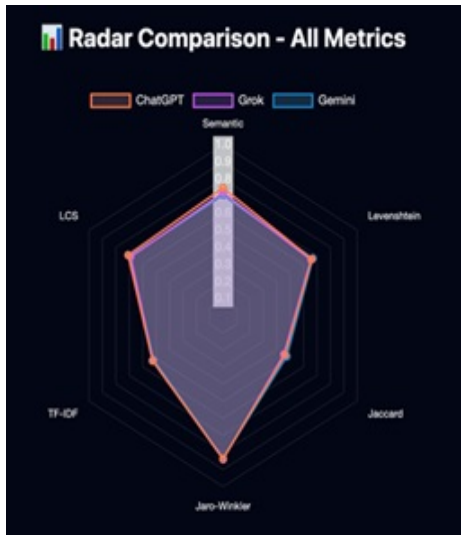

Figure 5. Radar Comparison

5. Conclusions

The exploration of Bessarabian idioms through Large Language Models (LLMs) confirms that current AI architectures have reached a threshold where they can effectively bridge the gap between traditional cultural heritage and digital innovation. This research demonstrates that while semantic non-compositionality remains a fundamental challenge in computational linguistics, modern models possess the necessary depth to navigate figurative meanings that cannot be derived from literal components.

Ultimately, the results indicate that LLMs are no longer merely experimental tools but have become core technologies capable of preserving and interpreting the ”unsalted” nuances of natural communication. This research lays the groundwork for future interactive educational experiences that leverage the synergy between Bessarabian folklore and Generative AI.

Acknowledgements. *This research has been supported by the project 25.80012.5007.79SE (CI ARiA),“Expressing cultural identity through the digital transformation of native idiomatic expressions using AR and AI technologies, with implementation in education”.*

References

[1] Golescu, I. *Proverbe comentate.* Editie ingrijita de dr. Gh. Paschita. Colectia Cognita, Editura Albatros, 1073, pp. 103.

[2] Cambridge dictionary, `https://dictionary.cambridge.org/dictionary/english/idiom`

[3] Botezatu, Gh. et al. *Folk Creation. Proverbs and Sayings*. Editura Stiintifica si Enciclopedica, Bucharest, Romania. (In Romanian: Creatie populara. Proverbe si zicatori), 1981, p. 309.

[4] Țițchiev, I. and Caftanatov, O.Exploring the Bessarabian idiom understanding by LLMs and AI tools. *The 20th edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing*, ConsiLR-2025, 8-10 octombrie 2025, Bucuresti, Romania.

[5] Lakoff, G. and Johnson, M. *Metaphors We Live By*. The University of Chicago Press, Chicago, USA, 1980, 276 p.

Received: September 30, 2025 *Accepted: December 29, 2025*

(Inga Titchiev) Moldova State University, "Vladimir Andrunachievici" Institute of Mathematics and Computer Science, 5 Academiei st., Chişinău, Republic of Moldova
E-mail address: inga.titchiev@math.usm.md

(Olesea Caftanatov) Moldova State University, "Vladimir Andrunachievici" Institute of Mathematics and Computer Science, 5 Academiei st., Chişinău, Republic of Moldova
E-mail address: olesea.caftanatov@math.usm.md

(Crinu Dimitriu) Tehnical University of Moldova, 7 Studentilor st., Chişinău, Republic of Moldova
E-mail address: crinu.dimitriu@gmail.com