AI-ASSISTED TRANSLATION OF IDIOMATIC LANGUAGE IN NEWS HEADLINES

Magdalena Duvnjak,
[0000-0001-9989-7919]

Vladimir Fedorov*,
[0009-0005-2235-8876]

Miloš Pupavac
[0000-0002-6179-3900]

Singidunum University,
Belgrade, Serbia

Abstract:

This paper examines the adequacy of translating idiomatic language in Serbian media headlines into German and Russian, with particular emphasis on set expressions using artificial intelligence tools (ChatGPT, Gemini, and Google Translate). The corpus comprises twenty-one headlines collected from online news portals, with a focus on the potential and limitations of these three tools, as well as on the relationship between literalness and expressiveness in translation.

The results indicate that these AI tools are highly effective at the lexical and syntactic levels, but also reveal limitations in the domain of phraseology, that is, in the translation of stylistically marked linguistic units.

Keywords:

News Headlines, Phraseologisms, AI-Assisted Translation, Idiomatic Language, Comparative Analysis.

INTRODUCTION

The translation of highly idiomatic language in general, including strongly colloquial expressions, represents one of the major challenges in translation studies. Within this broader category, set phrases naturally come into the foreground, as they typically constitute the most significant obstacle to successful translation. This is due to the fact that set expressions are often culturally bound, and their translation requires not only linguistic but also pragmatic and cultural competence.

Set expressions are characterised by fixedness, idiomaticity, and reproducibility [1], as well as a specific semantic structure that exceeds the sum of the meanings of their individual components [2]. Their strong cultural embeddedness further complicates the translation process [3]. The translation of set expressions is a central issue in translation studies, since their meaning is non-compositional, direct equivalents are often absent, and cultural factors play a decisive role.

Correspondence:
Vladimir Fedorov

e-mail:
vfedorov@singidunum.ac.rs

DOI: 10.15308/Sinteza-2026-491-495

As Nida points out, translation involves achieving dynamic, or functional, equivalence rather than formal correspondence [4]. Similarly, Baker emphasises that idiomatic expressions often require adaptation or paraphrase [5], while Nord highlights the importance of a function-oriented approach to translation in relation to purpose and audience [6].

A particular phenomenon is the use of set expressions in journalistic discourse, especially in news headlines. Here, they perform a specific function: attracting attention, condensing information, and enhancing textual expressiveness.

In news headlines, set expressions enable concise and expressive presentation of complex social issues, while also exerting a strong impact on the reader [7]. It is precisely this combination of linguistic economy and stylistic markedness that makes their translation particularly demanding.

The development of neural machine translation has significantly improved the processing of lexical and syntactic structures, as evidenced in tools such as ChatGPT, Gemini, DeepL, and Google Translate. However, according to Dankers, Lucas and Titov as well as Baziotis, Mathur and Hasler, set expressions remain a critical point: while some systems demonstrate greater idiomatic accuracy, others rely more heavily on paraphrase and contextual interpretation [8], [9]. Skacheva emphasizes the fact that there are still considerable problems in training AI systems in this specific domain, as shown in [10]. Furthermore, some studies, such as that by Fadaee, Bisazza and Monz, indicate that context does not always improve translation quality but may also lead to deviations in meaning [11]. Other scholars, including E. Mamaeva, also point to the "impossibility of achieving high-quality results without human involvement, particularly from a person capable of deeper contextual analysis and greater creativity." [12]

Building on the above, this paper focuses on the adequacy of translating set expressions in Serbian media headlines into German and Russian using artificial intelligence tools, with special attention to the relationship between literalness and expressiveness.

A specific motivation for this study arises from the observed inconsistency and inadequacy of set-expression translation in practice, particularly in media discourse, where meaning is often shifted in relation to the source expression.

2. RESEARCH OBJECTIVE

The aim of this study is to examine the degree of adequacy in translating phraseological units in news headlines based on a corpus of 21 examples, through a comparative analysis of translation solutions produced by three artificial intelligence-based tools (ChatGPT, Gemini, and Google Translate).

It is also important to note the difference between the first two and the latter. Modern machine translation systems and large language models differ in architecture and functional design. Google Translate is a dedicated neural machine translation (NMT) system based on transformer architecture, specifically optimized for translation tasks. In contrast, ChatGPT and Gemini are general-purpose large language models (LLMs), also transformer-based, in which translation is an emergent capability rather than a specialized component.

This distinction is methodologically relevant for the present study: NMT systems are typically designed to maximise lexical and syntactic alignment, whereas LLMs tend to exhibit stronger performance in contextual interpretation and idiomatic adaptation, often relying on paraphrasing strategies rather than direct equivalence.

The present analysis focuses on the relationship between literalness and expressiveness in translation.

3. HYPOTHESES

Based on the stated research objective, the following hypotheses are formulated:

1. It is assumed that the tools will recognise more frequent phraseological units and successfully convey the meaning of the headlines.
2. It is assumed that the tools will translate phraseological units using equivalent expressions in cases where such equivalents exist in the target language.
3. Highly colloquial and/or more recent lexical items, which are less conventionalized or do not form part of established set expressions, are expected to be more prone to translation errors.
4. LLM-based models (ChatGPT, Gemini) demonstrate better contextual processing but more frequently resort to paraphrase instead of direct equivalents.

4. METHODOLOGY

The study employs a contrastive and qualitative analysis.

Corpus:

We collected and analysed 21 news headlines from Serbian-language online media containing various types of phraseological units. The headlines were sourced from a range of news portals (kurir.rs; noviglas.info; novosti.rs; dnevnik.rs; b92.net; republika.rs; rts.rs; 021.rs; nin.rs; slobodnaevropa.org; politika.rs; sport.alo.rs; vecernjenovosti.ba; informer.rs; pulsonline.rs) and cover topics including politics, economics, sports, and society. During the selection process, headlines containing offensive or inappropriate content were excluded. While tabloid media tend to feature a high density of phraseological units and highly expressive lexis, such features are less common on more specialized portals (e.g., those focused on economics). The headlines were collected during the final week of March 2026.

Selection: headlines containing highly idiomatic language, in most cases—phraseological units and

Procedure: translation using three AI tools—ChatGPT, Gemini and Google Translate. The free versions of the tools were used, and the same prompt was applied to each headline in both languages: "Translate the following headlines into Russian/Serbian." We did not specify that the texts were news headlines, nor did we request literal translations or the preservation of expressiveness. Instead, our aim was to examine the type of output an average user would obtain, rather than that of a professional translator.

Analysis criteria:

- semantic accuracy
- stylistic adequacy

5. RESULTS AND DISCUSSION

We assessed translation quality based on a three-level evaluation of the obtained translations: excellent translation, good translation, and incorrect translation. The criteria for these categories are as follows: excellent translation—fully conveyed meaning; good translation—partially conveyed meaning; poor translation—incorrect translation, i.e., the meaning of the headline differs from its meaning in Serbian.

It should be noted that the "partially conveyed meaning" category includes examples in which idiomaticity or naturalness is violated, but the intended meaning remains clear. For instance, the rendering "сбрасывать пыль в глаза" (Google Translate; literally "to throw/drop dust into [one's] eyes") is semantically transparent, although it violates standard collocational norms, where "пускать пыль в глаза" would be expected. Occasional errors in proper names (observed in Google Translate outputs) were not taken into account, as they fall outside the scope of this study.

Based on the table above, we can conclude that news articles were translated with a high degree of success into both German and Russian.

The success rate of translation from Serbian into Russian and German was assessed based on the proportion of excellent, good, and incorrect translations.

The analysis of the results shows that translation into Russian is at a high level of success across all three tools. Gemini achieved the best results, with 90.48% excellent translations and no incorrect ones, indicating exceptionally high quality. ChatGPT also demonstrates a very high level of success (85.71% excellent translations and a small percentage of errors), while Google Translate performed somewhat worse due to a higher proportion of incorrect translations (19.05%).

On the other hand, translation into German shows a significantly lower level of success. "Gemini and ChatGPT show relatively satisfactory results, with 71.43% and 66.67% of titles translated excellently, respectively. However, Google Translate yielded a significantly weaker result, as the percentage of excellent translations is low, standing at 33.33%."

Table 1. Assessment of the Accuracy of the Obtained Translations

		ChatGPT	Gemini	Google Translate
Russian	Excellent translation	85.71%	90.48%	71.43%
	Good translation	9.52%	9.52%	9.52%
	Incorrect translation	4.76%	0%	19.05%
German	Excellent translation	66.67%	71.43%	33.33%
	Good translation	33.33%	28.57%	33.33%
	Incorrect translation	0%	0%	33.33%

Based on the above, we conclude that translation into Russian is significantly more successful compared to translation into German.

Differences in the degree of literalness and translation adequacy between Russian and German can be explained by the interaction between linguistic proximity and the operational mechanisms of AI tools. While the close relationship between Serbian and Russian often allows for literal translations to remain acceptable, the greater structural and phraseological distance between Serbian and German results in greater variability of strategies and a lower overall level of translation adequacy—a pattern consistently observed across all three systems analyzed.

This analysis aimed to examine the extent to which the tools are able to recognize and convey the expressiveness carried by both the original and translated headlines. Only correctly translated headlines were included in it.

Examples in which a degree of ambiguity is present but which nevertheless aim at expressiveness were also considered acceptable. Such ambiguity is reflected in the previous table, not in the present one. For instance, "все министры — под прицелом" (Gemini; literally, "all the ministers are in the crosshairs") can be interpreted both as being under scrutiny and as being in potential danger. Despite this ambiguity, the translation remains expressive and is therefore classified under "Preserved expressiveness."

The success rate of translating headlines from Serbian into Russian and German in this table can be examined based on the ratio between preserved expressiveness and literal translation using the selected tools (ChatGPT, Gemini, and Google Translate).

The analysis shows that in translations into Russian, literal translation predominates across all groups. For ChatGPT, 70% of translations are literal, while expressiveness is preserved in 30% of cases. A similar pattern is observed with Gemini (71.43% literal and 28.57% expressive translations), as well as with Google Translate (70.58% literal and 29.4% expressive). These results indicate that the examined AI tools more frequently resort to literal transfer of meaning than to the preservation of stylistic and expressive elements.

In the case of German, literal translation also prevails in the ChatGPT and Gemini groups (61.90% compared to 38.10% expressive translations), although the difference is smaller than in Russian. In translations produced by Google Translate, a balanced distribution is observed, with 50% literal translations and 50% translations preserving expressiveness, indicating greater variability in translation output.

Based on the above, we conclude that translation into Russian is more strongly oriented toward literalness, with all tools showing similar results. In contrast, translation into German shows a more balanced relationship between literal and expressive translation, particularly in the case of Gemini, which may indicate a more flexible translation approach. Interestingly, more recent or slang-leaning expressions such as "neopjevani blam," "kako da vam se boja uhvati od prve," "Srbi odlepili za Teslom," and "nemam pare da se bockam" proved more challenging to translate, particularly into German. However, in fairness to AI, it should be noted that in such cases it sacrifices a specific lexical item in order to preserve the overall message. This can be observed, for example, in the renderings of "neopjevani blam" and "nemam pare da se bockam", which are translated as "eine peinliche Blamage" (ChatGPT, literally "an awkward embarrassment") / "Unfassbare Blamage" (Gemini, literally "an unbelievable embarrassment"), and "ich kein Geld für Beauty-Injektionen habe" (ChatGPT, literally "I have no money for beauty injections") / "mir das Geld für die Spritze fehlt" (Gemini, literally "I lack the money for the injection").

In other words, the AI appears to recognize that a literal translation would be awkward or unintelligible, and therefore selects a term that elicits a similar response from the reader in German.

However, the observed differences do not indicate the superiority of any single tool and can be explained by the interaction of linguistic proximity and the functioning of AI tools: while the relatedness of Serbian and Russian allows a literal translation to often be correct, the greater structural and phraseological distance relative to German results in higher variability of strategies and a lower degree of overall translational adequacy—a finding consistently confirmed across all three analyzed systems.

Table 2. Assessment of the Accuracy of the Obtained Translations

		ChatGPT	Gemini	Google Translate
Russian	Preserved expressiveness	30%	28.57%	29.4%
	Literal translation	70%	71.43%	70.58%
German	Preserved expressiveness	38.10%	38.10%	50.00%
	Literal translation	61.90%	61.90%	50.00%

The results largely confirm the hypothesis that translation tools successfully convey the meaning of frequent phraseological units, although this is not always accompanied by idiomatic accuracy. The assumption that equivalent expressions would be consistently employed is not supported, as literal translation predominates. The hypothesis regarding the increased difficulty of translating recent and highly colloquial expressions is clearly confirmed. While LLM-based systems demonstrate superior contextual performance, the expectation that they would rely more heavily on paraphrase is not borne out by the data.

6. CONCLUSION

The analysis shows that, while AI-based translation tools generally succeed in conveying the core meaning of Serbian news headlines, they tend to favour literal translation over the preservation of expressiveness. This tendency is particularly pronounced in translations into Russian, where all tools display similar patterns, whereas translations into German reveal a more balanced relationship between literalness and expressiveness. The findings also confirm that highly colloquial and newly emerging expressions pose greater challenges, often resulting in reduced translation quality. Although LLM-based systems demonstrate strong contextual capabilities, they do not consistently employ idiomatic or equivalent expressions, and their performance varies across languages, with no single system achieving clear overall superiority.

When all aspects of the translations are considered and the tools are compared, it becomes clear that the results produced by ChatGPT, Gemini, and Google Translate are largely similar. No single tool can be said to be significantly superior or preferable to the others. Overall, artificial intelligence can be seen not only as a means of facilitating the understanding of foreign-language texts, but also as a valuable aid for professionals; however, all translations still require careful review.

Future research should include the analysis of larger corpora, the inclusion of additional languages, the investigation of post-editing strategies, and the development of AI models specialised in the translation of phraseological.

REFERENCES

[1] H. Burger, Phraseologie: eine einführung AM beispiel dês Deutschen, Berlin: Erich Schmidt Verlag, 2015.

[2] W. Fleischer, Phraseologie der deutschen Gegenwartssprache, Tübingen: Niemeyer, 1997.

[3] E. Piirainen, Widespread idioms in Europe and beyond: toward a lexicon of common figurative units, New York: Peter Lang, 2012.

[4] E. A. Nida, Toward a science of translating: With special reference to principles and procedures involved in Bible translating, Leiden: Brill Archive, 1964.

[5] M. Baker, In other words: A coursebook on translation, London: Routledge, 1992.

[6] C. Nord, Translating as a Purposeful Activity, Manchester: St. Jerome, 1997.

[7] A. Ziem, Frames und sprachliches Wissen, Berlin: De Gruyter, 2018.

[8] V. Dankers, C. Lucas and I. Titov, "Can Transformer be Too Compositional? Analysing Idiom Prosessing in Neutral Machine Translation," in *In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), Dublin, 2022. doi:10.48550/arXiv.2205.15301

[9] C. Baziotis, P. Mathur and E. Hasler, "Automatik Evaluation and Analises of Idioms in Neutral Machine Translation," Dubrovnik, 2023.

[10] N.V. Skacheva, "Analysis of Idioms in Neural Machine Translation: A Data Set," *Software systems and computational methods*, vol. 3, p. 55–63, 2024. doi:10.7256/2454-0714.2024.3.71518

[11] M. Fadaee, A. Bisazza and C. Monz, "Examining the Tip of the Iceberg: A Data Set for Idiom Translation," in *In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, 2018.

[12] E. Mamaeva, "Artificial intelligence in translating phraseological units of a literary text," *Philology and Culture*, vol. 4, pp. 96-103, 2024. doi:10.26907/2782-4756-2024-78-4-96-103