University "St. Kliment Ohridski" – Bitola
Faculty of Education – Bitola
ISSN: 1857- 8888
DOI: 10.20544/teacher.29.07

AI IN NONLITERAL LANGUAGE TRANSLATION: TRANSLATING MACEDONIAN PROVERBS AND IDIOMS

Elena Shalevska
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola
elena.shalevska@uklo.edu.mk

Bisera Kostadinovska-Stojchevska
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola
k_bisera@yahoo.com

Violeta Janusheva
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola
violeta.janusheva@uklo.edu.mk

Maja Janusheva
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola
maja_robinson@hotmail.com

Marija Stojanoska
Faculty of Education – Bitola, University "St. Kliment Ohridski" – Bitola
marija.stojanosaka@uklo.edu.mk

Marija Talevska
Bitola, North Macedonia
m.talevska1@yahoo.com

Abstract

Translating proverbs and idioms poses a significant challenge for translators worldwide. Thus, it is no wonder that AI models seem to struggle, too. These expressions often rely on cultural context, figurative meaning, and linguistic creativity – things that AI models cannot really replicate. Recognizing this, this study examines how three AI models—ChatGPT, Gemini, and DeepSeek—handle the translation of Macedonian proverbs and idioms into English. The findings reveal several patterns: Explanation accuracy varies though all AI models consistently provide a two-part translation; they all favor direct translations over idiomatic equivalents, provide uniform translations of simpler proverbs, but show significant variability in word choice for more complex ones, and fail to maintain stylistic elements such as rhyme, alliteration, and wordplay.

Keywords: Artificial Intelligence; Machine Translation; AI Translation; Nonliteral Language; Proverbs; Idioms.

1. Introduction

Proverbs and idioms are integral components of any modern language in use. They show speakers' linguistic creativity, and culture, as well as their societal norms, and collective experiences. Both proverbs and idioms often convey meanings that transcend their literal interpretations. This makes them rich in connotative and figurative meaning (Sazdovska Pigulovska, 2020). This also makes them challenging for translators.

One can easily see why translating such nonliteral language elements may pose significant challenges: To translate these expressions accurately, linguistic proficiency is simply not enough. One must also understand one country's history, culture, traditions, language changes, attitudes, and overall day-to-day life.

Enter, the machines. The advent of Artificial Intelligence (AI) and Neural Machine Translation (NMT) systems has revolutionized the field of translation. However, the efficacy of these systems in handling idiomatic and proverbial expressions remains a subject of ongoing research. Studies have shown that while machine translation excels in translating literal language, computers often struggle

with nonliteral expressions due to their reliance on contextual and cultural nuances (Gjurchevska Atanasovska & Sazdovska Pigulovska, 2017).

The Macedonian language, with its vast number of proverbs and idioms, presents a unique challenge for even the most advance AI models of now. These expressions are deeply rooted in our nation's history, traditions, and societal values, and they often lack direct equivalents in English. Recognizing this, this paper aims to evaluate the performance of contemporary AI translation models in translating Macedonian proverbs and idioms into English.

Subheading 1.1. Machine Translation and AI Translation: An Overview

Machine Translation (MT) has evolved significantly since its inception. It has undergone a major transformation from early rule-based systems to more advanced artificial intelligence-driven models. Old-school, traditional MT relied on linguistic rules and dictionaries to translate text between languages. It was revolutionary at the time, though lacking from a modern perspective. The original approaches often failed to capture context, idiomatic expressions, and cultural context (Hutchins, 2005). The introduction of statistical machine translation (SMT) in the early 2000s improved translation quality significantly. This type of machine translation used large bilingual corpora to predict the most probable translations (Koehn, 2012). Though better than earlier MT attempts, SMTs still struggled with syntactic flexibility and contextual adaptation. This led to the rise of neural machine translation (NMT). NMT, powered by deep learning and artificial neural networks, has become the dominant approach in modern AI translation. Systems like Google Translate use NMT to process entire sentences holistically, rather than word-for-word. This, in turn, improves fluency and contextual understanding (Bahdanau et al., 2015). Despite these advancements, many challenges persist in MT – challenges such as translating figurative language, cultural references, and ambiguous expressions, among else.

AI translation has further evolved with the emergence of large language models (LLMs) such as GPT-4, Bard, and Mistral. These models use transformer-based architectures (Kaplan, 2024) to enhance translation accuracy. Unlike traditional NMT systems, LLMs can incorporate broader contextual knowledge, refine their translations based on user input, and generate more natural-sounding text (Brown et al., 2020). These models, however, still exhibit limitations. For example, AI often defaults to literal translation rather than recognizing the deeper cultural meanings behind select phrases (May Pulido, 2024). Ethical concerns such as bias, and misrepresentation, are also important to discuss in this regard (Shalevska, 2024). And while AI has undeniably improved global communication and accessibility, its effectiveness in accurately translating culturally- and nuance-rich texts—such as proverbs and idioms—is still being evaluated.

2. Literature review

The advent of Artificial Intelligence (AI) in the field of translation has revolutionized the way languages are interpreted and understood. However, the translation of culturally rich and linguistically complex elements, brings about unique challenges in AI, and machine translation, in general. Especially when it comes to proverbs.

Proverbs are succinct expressions that convey cultural wisdom, values, and norms. Their meanings often extend beyond the literal interpretation of individual words, making them challenging to translate, especially for AI systems. Wang et al. (2025) investigated the proficiency of Neural Machine Translation (NMT) systems and Large Language Models (LLMs) in translating proverbs across four language pairs. The study found that while AI models performed adequately between languages with similar cultural backgrounds, they struggled with proverbs deeply rooted in specific cultural contexts. Similarly, Romaniuk-Cholewska (2024) evaluated the translation quality of 18 proverbs using various machine translation tools, including Google Translate, DeepL, PONS, ChatGPT-3.5, and Google Bard, in the English-Polish language pair. The findings revealed that AI tools exhibited varying degrees of accuracy, with ChatGPT achieving 61% accuracy in Polish-to-English translation and 38% in English-to-Polish translation.

Similar challenges are seen in the machine translation of wordplay. Humor and wordplay, which are sometimes included in proverbs and idioms, often rely on linguistic creativity, cultural references, and contextual understanding, posing significant challenges for AI translation systems. Rudenko and Semenova (2024) conducted a study comparing AI-generated translations of pun-based jokes to those produced by professional human translators. The study revealed that AI translations were often perceived as complex due to literal interpretations and calques, whereas human translators employed holistic techniques, demonstrating greater flexibility in interpreting imagery and syntactic structures. This disparity highlights the creative freedom human translators possess – something that AI cannot compete with. Creativity is also something that Basiotis et al. (2023) stress in their study on NMT translation of idioms. By examining translations of 155 idioms by different machine and AI translation systems, the authors find that machine translation systems still struggle with nonliteral language, often defaulting to literal translations that may not convey the intended meaning.

3. Methods

To assess the efficacy and accuracy of AI translations of proverbs and wordplay-based phrases within them, this study uses the qualitative design and the general qualitative interpretative-inductive method used by Kahlke (2014) and Harding & Whitehead (2016). At the very core of the study lies a self-compiled corpus of 10 proverbs and idioms in Macedonian, chosen by the authors:

Selo gori baba se cheshla

So muvata na kapata

Brat brata ne rani, teshko koj go nema

Ne frlaj po Gospod so kamenja

Utresh kukja ne krepi

Ako ne zavrni, kje kapni

Vo Prilep majmun ne igra

Ne umri magare do zelena treva

Krushata pod krusha pagja

Lozje ne bara molitva, tuku motika

These proverbs and idioms have then been translated into English using three commercially available AI models: ChatGPT, Gemini and DeepSeek, prompted by the authors. The output i.e. translation is then carefully analyzed to determine patterns and inconsistencies.

4. Results and discussion

The analysis of the AI-generated translations by the three models (available below, in full, in Table 1) provides some truly interesting insights.

Table 1: AI-generated translations of Macedonian proverbs and idioms

	AI Translations		
Proverb / Idiom	**ChatGPT**	**Gemini**	**DeepSeek**
Selo gori baba se cheshla	The village is burning, but grandma is combing her hair. (Used to describe someone ignoring urgent problems and focusing on trivial matters.)	"The village is burning, and the old woman is combing her hair." This idiom describes someone who is oblivious or indifferent to a serious problem happening around them.	The village is on fire, and the old lady is combing her hair. (Meaning: Ignoring a major problem while

			focusing on something trivial.)
So muvata na kapata	With a fly on one's hat. (Refers to someone who is unnecessarily defensive or feels accused without reason.)	"With a fly on the hat." This idiom describes someone who is arrogant or conceited.	With a fly on the hat. (Meaning: Being overly meticulous or nitpicky about small details.)
Brat brata ne rani, teshko koj go nema	A brother doesn't feed a brother; woe to the one who has no one. (Family ties alone don't provide sustenance; one must fend for themselves, but it's even harder for those who have no family at all.)	"A brother doesn't feed a brother, woe to him who has none." This expresses the idea that family support is essential, even if it's not always practical. It emphasizes the importance of having family.	A brother doesn't harm a brother, woe to the one who doesn't have one. (Meaning: Family sticks together, and it's hard for those without family support.)
Ne frlaj po Gospod so kamenja	Don't throw stones at God. (Don't challenge fate or divine will.)	"Don't throw stones at God." This idiom warns against challenging or blaming a higher power or fate.	Don't throw stones at God. (Meaning: Don't challenge or disrespect higher powers or fate.)
Utresh kukja ne krepi	Tomorrow doesn't hold up the house. (Procrastination won't solve problems; action is needed.)	"Tomorrow doesn't build a house." This means that procrastination gets you nowhere; you need to act now.	Don't build your house for tomorrow. (Meaning: Don't postpone important tasks or preparations.)
Ako ne zavrni, kje kapni	If it doesn't pour, it will drip. (Even if something big doesn't happen, a smaller version of it eventually will.)	"If it doesn't rain, it will drip." This means that something bad is going to happen one way or another.	If it doesn't turn, it will drip. (Meaning: If you don't take action, things will go wrong.)
Vo Prilep majmun ne igra	In Prilep, the monkey doesn't dance. (In some places, tricks and deception don't work.)	"A monkey doesn't dance in Prilep." This means that there are no foolish actions or tricks in Prilep, or that people from Prilep are serious and not easily fooled.	In Prilep, a monkey doesn't dance. (Meaning: Certain things are not done in certain places or situations.)
Ne umri magare do zelena treva	Don't die, donkey, until the green grass comes. (Used when waiting for something that may never come.)	"Don't die, donkey, until the green grass." This is a cynical expression about false hope or a plan that will never work.	A donkey doesn't die near green grass. (Meaning: People won't give up when

			there's an opportunity or solution nearby.)
Krushata pod krusha pagja	The pear falls under the pear tree. (Children often take after their parents.)	"The pear falls under the pear tree." This means that children often resemble their parents.	The pear falls under the pear tree. (Meaning: Children often resemble or follow their parents.)
Lozje ne bara molitva, tuku motika	A vineyard doesn't need prayer, but a hoe. (Hard work, not wishful thinking, brings results.)	"A vineyard doesn't need prayer, but a hoe." This means that hard work, not just wishing or praying, is what brings results.	A vineyard doesn't need prayers, but a hoe. (Meaning: Hard work is more effective than just hoping or praying.)

The first noteworthy matter is that all tested AI models tend to produce a two-part translation: a direct definition followed by an additional explanation of the meaning. This shows that all of the AI models tested "understand" that the sentences provided include nonliteral language use and understanding them may require an additional elaboration.

Secondly, all of the AI models tested seem to struggle with handling homonyms and polysemous words—terms that have multiple meanings depending on context. This is particularly evident in the translations of "Ne umri magare **do** zelena treva" and "Brat brata ne **rani**, teshko koj go nema". The preposition "Do" in Macedonian, can mean both "next to" and "until", while the verb "rani" can mean both "feed" and "harm/wound". Some models default to one meaning, while others opt for the alternative, leading to inconsistent translations. These inconsistencies suggest that AI models still rely on word-frequency biases rather than deeper contextual understanding when processing translating such nonliteral examples.

For some, more general proverbs and idioms, the translations seem identical across all models. This likely occurs because these proverbs have clearer, more straightforward structures and resemble idioms found in English. For instance, all three models translate:

- "Krushata pod krusha pagja" → "The pear falls under the pear tree"
- "Ne frlaj po Gospod so kamenja" → "Don't throw stones at God"
- "Lozje ne bara molitva, tuku motika" → "A vineyard doesn't need prayer, but a hoe".

The translations of other proverbs and idioms, however, show greater variety: In "Ako ne zavrni, kje kapni", the verb "zavrni" is translated into "pour", "rain", and "drop"; in "Selo gori baba se cheshka", the noun "baba" is translated into "old woman"; "old lady", and "grandma"; in "Utresh kukja ne krepi", the verb "krepi" is translated into "hold" by ChatGPT and "build" by DeepSeek and Gemini.

Finally, all models seem to use a neutral, explanatory role when translating the proverbs and idioms. The lack of creativity, thus, strikes as the most striking and noteworthy observation. Gone is the creativity seen in the original text – the alliteration that adds a specific sound quality in some instances, or the rhyme – in others. AI translations strip away these stylistic elements. The translations sound "mechanic", and even "strange" word-by-word translations in English. It is also interesting to note that not a single instance of idiom-to-idiom or idiomatic equivalence translation occurs in the corpus. For instance, "Nero fiddles while Rome burns" would have been a great equivalent to "Selo gori baba se cheshla", and "The apple doesn't fall far from the tree" would have captured the essence of "Krushata pod krusha pagja". Yet, no such examples are found in the corpus of translations.

Although the AI-generated explanations are not direct translations, per se, they do play a role in understanding the proverbs. Speakers of both Macedonian and English will notice that a significant

number of these explanations are accurate. This best shows the progress AI models have made in recent years. In terms of accuracy, AI models show:

- ChatGPT: 9/10 accuracy
- Gemini: 6/10 accuracy
- DeepSeek: 5/10 accuracy.

For non-Macedonian speakers, the accurate explanations are marked in green in the table above. The significant gap between ChatGPT and the other models tested suggests that AI development in this area is uneven, with some models excelling in contextual interpretation while others lag behind. It's important to note though, that although accurate in a broad sense, the explanation for "Vo Prilep majmun ne igra" doesn't fully convey the cultural meaning of the expression. The people of the town of Prilep are said to be bright and cunning – which is why tricks and deception just "don't fly" there.

The most problematic proverb seems to be "Ne umri magare do zelena treva" – a proverb that, according to the Explanatory Dictionary of the Macedonian Language (n.d.) means that one has to *"To endure or persist until one finally receives what has been promised to them,, or what they expected".* All of the tested models struggle grasping this meaning. However, none of the AI models correctly grasp this meaning. Instead, they either: translate it literally, producing an awkward or nonsensical English sentence; misinterpret "do" as "until" instead of "next to"; and struggle to recognize the metaphorical significance of "zelena treva" (green grass), which represents prosperity. This case best shows AI's limitations in handling culturally embedded metaphors in Macedonian, particularly those that lack direct analogs in English – which goes in line with previous studies done in other language pairings (Wang et al., 2025; Rudenko and Semenova, 2024; Basiotis et al., 2023 etc.)

5. Conclusion

This study has revealed that while AI models have made impressive strides in translating proverbs and idioms, particularly in Macedonian, they still fall short in areas requiring deep cultural understanding, figurative meaning, and linguistic creativity. The data obtained for the purpose of this study clearly shows that AI is improving in recognizing idioms, but, in many instances, it still struggles with polysemy, cultural context, and figurative language. The models still favour literal translations over idiomatic equivalence, leading to awkward phrasing. And while some – such as ChatGPT – show better results, there's much room left for improvement.

Overall, this study highlights both AI models' progress and their ongoing challenges, pointing toward future improvements that could make AI translation of nonliteral language significantly better.

References.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473.*

Baziotis, C., Mathur, P., & Hasler, E. (2023). Automatic Evaluation and Analysis of Idioms in Neural Machine Translation. *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics*. https://doi.org/10.18653/v1/2023.eacl-main.267

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877-1901.

Gjurchevska Atanasovska, K., & Sazdovska Pigulovska, M. (2017). When translation goes wrong: Translating idiomatic expressions. In *English Studies at the Interface of Disciplines: Research and Practice.* Blaze Koneski Faculty of Philology

Harding, T., & Whitehead, D. (2016). Analyzing data in qualitative research. In: Schneider, Z. & Whitehead, D. (eds.), *Nursing and midwifery research: methods and appraisal for evidence-based practice, 5th ed*. Elsevier.

Hutchins, W. J. (2005). Machine translation: A brief history. In *Concise history of the language sciences* (pp. 431-445). Retrieved from: Machine Translation: a brief history on 18.3.25.

Koehn, P. (2012). *Statistical machine translation.* Cambridge University Press.

Kahlke, M. R. (2014). Generic qualitative approaches: pitfalls and benefits of methodological mixology. *International Journal of Qualitative Methods, 13*(1), 37–52.

May Pulido, S. (2024). Reinventing Translation: AI and Human Collaboration. *Multilingual*. Retrieved from: https://multilingual.com/reinventing-translation-ai-and-human-collaboration/ on 17.3.25.

Romaniuk-Cholewska, D. (2024). The machine translation (MT) of proverbs in the ENG-PL language pair. *Studies in Logic, Grammar and Rhetoric, 69*(1), 553-570. The Machine Translation (MT) of Proverbs in the ENG-PL...

Rudenko, E. S., & Semenova, M. Y. (2024). Artificial vs. human intelligence: A case study of translating jokes based on wordplay. *Research Result. Theoretical and Applied Linguistics, 10*(4), 53-93. https://rrlinguistics.ru/en/journal/annotation/3675/

Sazdovska Pigulovska, M. (2020). Definition and classification of phraseological expressions and problems of idiomatic equivalence in translation. *Palimpsest: International Journal for Linguistic, Literary and Cultural Research.*

Shalevska, E. (2024). Human Rights in the Age of AI: Understanding the Risks, Ethical Dilemmas, and the Role of Education in Mitigating Threats. *Journal of Legal and Political Education 1*(2):38-52. https://doi.org/10.47305/JLPE2412038sh.

Tolkoven rechnik na makedonskiot jazik / Explanatory dictionary of the Macedonian language (n.d.) *Umre*. Retrieved from: умре | Македонски.гов.мк on 15.3.25.

Wang, M., Pham, V.-T., Moghimifar, F., & Vu, T.-T. (202*5). Proverbs run in pairs: Evaluating proverb translation capability of large language models. *arXiv preprint arXiv:2401.12345*. https://www.researchgate.net/publication/388317930_Proverbs_Run_in_Pairs_Evaluating_Proverb_Translation_Capability_of_Large_Language_Model