



TECHNIUM
SOCIAL SCIENCES JOURNAL

www.techniumscience.com



Vol. 78/2025
A New Decade for Social Changes

PLUS
COMMUNICATION P



International
Communication & PR

Translating English Number Idioms into Arabic: A Neural Machine Translation Evaluation

Abdullah Sanad Mohammad Aldela'a¹, Mai Abdel karim Mohammad Malkawi²,
Ayman Altalihin³

¹²The Language Center, The Hashemite University, Zarqa, Jordan, ³Amman Arab University, Amman, Jordan

abdullahs@hu.edu.jo, Mmalkawi@hu.edu.jo, A.abdallah@aau.edu.jo

Abstract. Idiomatic expressions, particularly those involving numbers, pose significant challenges for neural machine translation (NMT) systems due to their non-literal meanings and cultural specificity. This study investigates the efficacy of NMT systems, specifically Google Translate and Chatgpt, in translating English number idioms into Arabic. A corpus of 30 English number idioms was investigated and their translations analyzed in terms of literalness, cultural appropriateness, and semantic accuracy. The findings reveal that while some idioms are accurately translated with functional equivalents, a substantial number are rendered literally. This leads to some problems in accuracy culturally and meaningfully. The study underscores the need for enhancing NMT systems with idiom-specific modules and culturally enriched datasets to improve translation quality.

Keywords. Neural Machine Translation, Number Idiomatic expressions, English-Arabic Translation, Google Translate, Chatgpt

Introduction

Idiomatic expressions are a fundamental component of natural language, encapsulating cultural nuances and figurative meanings that often defy literal interpretation. Among these, number idioms—phrases that incorporate numerical elements to convey idiomatic meanings—present unique challenges in translation due to their cultural specificity and non-compositional nature. For instance, the English idiom "at sixes and sevens" conveys a state of confusion, a meaning not directly inferable from the individual words.

Neural Machine Translation (NMT) systems, such as Google Translate and Chatgpt, have revolutionized the field of machine translation by leveraging deep learning techniques to produce more fluent and contextually appropriate translations. However, the translation of idiomatic expressions remains a persistent challenge for these systems, often resulting in literal translations that fail to capture the intended meaning (Baziotis et al., 2022).

This study aims to evaluate the performance of NMT systems in translating English number idioms into Arabic, a language with distinct linguistic structures and cultural references. By analyzing the translations produced by Google Translate and Chatgpt, the study seeks to

identify common pitfalls and propose strategies for improving the translation of idiomatic expressions.

Literature Review

The translation of idiomatic expressions has been a longstanding challenge in the field of machine translation. Traditional rule-based and statistical machine translation systems have struggled with idioms due to their fixed nature and non-literal meanings. With the advent of NMT, there has been a significant improvement in handling context and producing more natural translations. However, idioms continue to pose difficulties.

Baziotis et al. (2022) highlight that NMT models tend to produce literal translations for idiomatic expressions, leading to confusion and loss of meaning. Similarly, Dankers et al. (2022) observe that the compositional nature of transformer-based models contributes to the literal translation of idioms, as these models process idioms as compositional expressions rather than recognizing them as fixed units.

In his article, Ibrahim Jibreel (2023) investigated which tool will give more communicative, semantic or literal meaning for proverbs. The results were the following: “Bing is the most effective MT providing communicative proverbial equivalents. Bing and Google, in the same rank, provide semantic equivalents. Furthermore, the least effective MT among the five is Yandex. MT errors diverge between missing the implied meaning, weakly structured translations, wrong synonyms and meaning distorting.”

In the context of English-Arabic translation, Alkhawaja et al. (2020) conducted a fine-grained evaluation of Google Translate's output and found that while the system performs adequately on general texts, it struggles with idiomatic expressions, often resulting in mistranslations and loss of meaning. Mounassar (2024) also emphasizes the challenges faced by machine translation systems in translating English fixed phrases into Arabic, noting issues such as poor translations, loss of figurative meaning, and inappropriate equivalents.

Ben-Yahia, Moad (2024) examined in his paper the use of machine translation tools in translating English idioms found in American media. He used, for the purpose of the study, Google Translate, DeepL Translator, and Bing Microsoft Translator. The result of his study showed that idioms were translated with an average accuracy of 68.7%. Sometimes the translation was done by paraphrasing. According to Ben-Yahia, “Bing Microsoft Translator sometimes opts for using idioms in the target language that are similar in terms of meaning to the ones in the source language, which reveals new horizons for the development of high-tech online machine translation services that would be able to accurately and meaningfully translate cultural-specific lexis such as idioms.”

Palanichamy Naveen (2025) in their article focuses on the importance of grammatical correctness of the output of machine translation and the syntactic coherence as well while preserving the cultural aspects of the original. In addition, they gave special attention to the “challenges related to complex sentence structure and grammatical transformation.” They concluded that it is important “enhancing machine translation systems to better break language barriers, foster multicultural understanding, and support global collaboration”.

Blanka Klimova (2025) conducted a survey in the context of using machine translation in learning a foreign language. He asked the students about the tools they normally use for learning a foreign language, why they use them and what are the challenging faced during the education process. Klimova stated that: “A significant challenge identified was the inability of MT tools to accurately convey cultural nuances and context, with students frequently expressing the need for clearer institutional guidelines on the allowed use of MT in educational settings”.

Haider et al. (2025) , studied the perception and the judgment of translation students towards the AI-translated idioms and proverbs. The study showed that “ Participants who frequently encountered idiomatic content and received formal training in translation evaluation demonstrated greater skepticism toward AI’s handling of figurative meaning. While students acknowledged the efficiency and utility of AI tools, a strong preference remained for human translation when dealing with nuanced, culturally rich expressions.”. Thus, the study , while acknowledging the importance of AI tools, supports the growing need for AI literacy, critical evaluation skills, and ethical awareness in translator education.”.

Methodology

Corpus

A corpus of 30 English number idioms was compiled from reputable idiom dictionaries and linguistic corpora. The selection criteria ensured a diverse representation of idioms incorporating various numbers and covering a range of meanings and contexts.

Translation

Each idiom plus its context translated by two NMT systems: Google Translate and Chatgpt. The resulting Arabic translations were collected for analysis.

Process

Evaluation

The translations generated by the two systems were evaluated based on three criteria:

- Literalness: Assessment of whether the translation was a word-for-word rendering of the original idiom.
- Cultural Appropriateness: Evaluation of whether the translation employed an equivalent Arabic idiom or culturally appropriate expression.
- Semantic Accuracy: Determination of whether the translated idiom preserved the intended meaning of the original expression. Two bilingual linguists with expertise in English-Arabic translation independently evaluated the translations using a 3-point scale: 0: Inaccurate translation, 1: Partially accurate translation, and 2: Accurate translation. Inter-rater reliability was calculated using Cohen's kappa is also applied in order to ensure consistency among systems.

Criteria

Results and Discussion

The findings indicate that the performance of NMT systems differs in translating English number idioms into Arabic, primarily due to the literal translation approach and lack of cultural adaptation. The high rate of literal translations indicates that Google Translate fails to recognize idiomatic expressions. In general, we can say that ChatGPT has achieved remarkable success compared to Google Translate, as the number of idiomatic expressions that were correctly translated was 20, which represents a 67% of the total number of idiomatic expressions. Google Translate, on the other hand, achieved a low success rate, as it was unable to translate 21 idiomatic expressions out of 30, which constitutes a shortcoming in this program. The results of translation were recorded in the following tables (1 and 2)

Table 1 (accuracy ratio of Google Translate)

Google Translate	Inacurate	Partially accurate	Accurate	Total
Idiom’s number	21	4	5	30
Idiom’s ratio	70%	13%	17%	100%

As the table shows, 21 idioms out of 30 number idioms were not translated correctly. This is a low percentage. Also, 4 idioms were partially translated accurately, and only five idioms were translated correctly.

Chatgpt	Inaccurate	Partially accurate	Accurate	Total
Idiom's number	3	7	20	30
Idiom's ratio	10%	23%	67%	100%

Table 2 (accuracy ratio of Chatgpt)

Table 2 shows a great proficiency of Chatgpt in translating number idioms, with 20 out of 30 being correct, which is a high score. Seven phrases were partially translated and only three idioms were mistranslated.

The remarkable advantage of Chatgpt over Google Translate may be attributed to differences in training data and model architectures. However, the overall performance indicates a need for improvement in handling idioms. Integrating idiom recognition modules and expanding training datasets to include idiomatic expressions could enhance the performance of NMT systems. Additionally, incorporating cultural knowledge bases may aid in selecting appropriate equivalents in the target language. System comparison revealed that Chatgpt significantly outperformed Google Translate since the number of problems encountered by Chatgpt is less than google translate. These problems related to literalness, semantic accuracy and cultural Appropriateness as explained in the table and the examples in this section.

	Literalness	Semantic Accuracy	Culture Appropriateness	Total
Google Translate	14	2	5	21
Chatgpt	3	4	3	10
Total	17	6	8	31

Table (3) problems encountered by each system

Literalness

This problem forms the highest ratio (55%) of all inaccurate translations. It consists of 6 number idiomatic expressions which were inaccurate either completely or partially, where the translation captured some aspects of the original meaning but lacked contextual and semantic appropriateness. According to the results of the evaluation of neural machine translations systems, Google Translate has rendered 27 of the 30 images using literal translation. Only five of Google's 27 literal translations are accurate, while the other 22 renderings were inaccurate or word-for-word translations. Chatgpt opts for literal translation only 7 times out of 30 and among these seven translations were 5 correct translations. In some cases, both systems– Google translate and Chatgpt - produce a partially accurate rendition for the expression. An example of this is the idiom **one of these days** which is translated into **في أحد الأيام القادمة:** by Chatgpt, and **into أحد هذه الأيام** by google translate. Both renditions are partially correct in conveying the intended meaning. It is thought that both tools give an accurate rendition because the number

(one) found in the idiom is used and translated as an ordinary word not a number. Moreover, the word one is used as a modifier – as in this idiom- not as a number. Thus, it was not impossible for the two tools to grasp its meaning easily and without having to think of its connotations or shades. This is also clear in when translating the idiom **a thousand times** into ألف مرة by both tools. This is due to the literalness of the expression which leads to an accurate rendition just by having the denotative meaning of its lexical items.

When it comes to literalness, Google translate failed to give an accurate rendition for the expression **at sixes and sevens** as it opted for a literal inaccurate Arabic translation في السادسة والسابعة which does not convey its original meaning. On the other hand, Chatgpt, managed to give a partially accurate rendition في حالة من الفوضى / في حالة ارتباك that conveys the semantic meaning in Arabic. For example, *at sixes and sevens* was rendered as "في الستات والسبعات," a nonsensical phrase in Arabic that fails to convey the intended meaning of confusion or disorder.

Moreover, Google Translate produced a literal translation of the sentence "our team won after they scored a goal at the eleventh hour," rendering the idiom as "في الساعة الحادية عشرة", which does not convey the intended meaning. In the context of a football match—typically lasting no more than two hours—this expression should be interpreted pragmatically rather than literally. A more contextually appropriate translation is rendered by Chatgpt as "في الوقت القاتل". This example demonstrates that Google Translate tends to prioritize literal meaning over pragmatic or contextual interpretation. Such inaccuracy may result from the system's limited ability to comprehend the broader context or from insufficient training on idiomatic and context-dependent translations.

In addition to the previous example, the translation generated by google translate of the expression **I feel like a million dollars** as: أشعر وكأنني مليون دولار is also in accurate due to literalness which gives a translation far away from the intended meaning which must be اشعر اشعر بملايين من السعادة هائلة which means (I am extremely happy). However, Chatgpt opted for a fairly accurate rendition that conveys partially the meaning أشعر أنني في قمة الراحة. The same applies to the idiom *a nine-to-five job* , which is translated by both tools correctly as : وظيفة من التاسعة إلى الخامسة which literally means a full time job that goes from 9 – 5 . It is thought that the rendition is accurate because the idiom has no connotation or a hidden, unclear meaning. The meaning of the idiom is in fact literal and easily understood from the words constituting it.

One-horse town, is an idiom that is translated accurately by Chatgpt into بلدة صغيرة ومملة and مكان ممل وبسيط جداً which mean a place which is very small– that can be wandered using one horse - and simple that has no media for entertainment _ except for a horse which is the only way for entertainment. All these aspects, shades and connotations are conveyed in this translation. But when it comes to google translate, it gives a literal rendition that ignores all these essential components of the intended message.

When translating *to kill two birds with one stone*, Google translate opted for literal translation as: لقتل عصفورين بحجر واحد , which is not the natural rendition used in Arabic. Whereas Chatgpt's rendition يضرب عصفورين بحجر واحد is more accurate since it is the way how it is said in Arabic where the verb يضرب (to hit) is used instead of قتل which is the literal translation of the verb to kill.

Semantic Accuracy

This problem forms a ratio (19%) of all inaccurate translations. It consists of 6 number idiomatic expressions which were inaccurate either completely or partially, where the

translation captured some aspects of the original meaning but lacked contextual and semantic appropriateness.

Both systems encountered difficulties in achieving semantic accuracy when translating number idiomatic expressions. They render some examples using literal translation which is inadequate and neglects the semantic meaning of the SL expressions. For example, the idiom “to feel like a fifth wheel” conveys the sense of being superfluous or unnecessary (يشعر بأنه زائد عن الحاجة), whereas “she was talking nineteen to the dozen” signifies speaking rapidly or excessively (تتحدث بسرعة كبيرة).

Idioms containing numerical elements, such as “to feel like a fifth wheel” and “she was talking nineteen to the dozen,” illustrate how figurative meaning overrides literal interpretation, complicating the translation process. The idiom “to feel like a fifth wheel” literally refers to an extra wheel that is not required for a vehicle to function. Figuratively, it describes a person who feels unnecessary or excluded in a social context. Google translate rendered this idiom into Arabic, but all semantic elements of this idiom were neglected. The number in this idiom “fifth” and the object “wheel” lost their metaphorical meaning, produced a nonsensical translation. Instead, an equivalent idiomatic expression — يشعر بأنه زائد عن الحاجة (feeling redundant or unneeded)— better conveys the intended meaning by chatgpt, despite omitting the numerical imagery.

Similarly, “she was talking nineteen to the dozen” cannot be understood through its literal numerical reference. The idiom is traditionally British and means that someone is speaking very quickly or continuously. The phrase originated in the 19th century, possibly from industrial settings where “nineteen to the dozen” referred to rapid production rates. Translating it word-for-word into Arabic results in confusion, as the numerical relationship carries no cultural or semantic significance in Arabic. The closest equivalent would be تتحدث بسرعة كبيرة (she speaks very fast), which conveys the sense but not the vivid imagery or cultural color of the original expression. It is clear in this example that Google Translate maintained the form when rendering the example as كانت تتحدث تسعة عشر الى اثني عشر but it failed to provide accurate translation, as it concealed many important semantic details and meanings. Chatgpt, on the other hand, conveyed the dynamic and figurative meaning without rendering the form and numbers.

Cultural Appropriateness

The analysis shows that ChatGPT seems to be superior to Google Translate when it comes to translating NIEs that have cultural appropriateness. The superiority of Chatgpt is found in translating numerous examples, including “kill two birds with one stone,” which was translated accurately as يضرب عصفورين بحجر واحد that is suitable for the original. Google Translate was unable to provide this cultural aspect, as it translated “يقتل عصفورين بحجر واحد.”

Culture appropriateness is a serious problem faced by the two systems. There are eight number idiomatic expressions representing (26%) of all inaccurate translations that were translated using inappropriate cultural Arabic equivalents by both systems understudy. It is noted that Google translate has committed 5 inaccurate renditions whereas Chatgpt has committed only 3. This reflects the intelligence and accuracy of Chatgpt.

Idioms such as “three sheets to the wind” don’t literally refer to the number three or to sheets. The expression comes from sailing jargon, where loose sails cause a ship to wobble — metaphorically describing a drunk person. In Arabic, however, there’s no cultural or linguistic equivalent based on that metaphor. Thus, both systems find a natural equivalent (e.g. ثمل جداً) that conveys the intended meaning, not the literal image.

Conclusion

This study highlights the challenges faced by NMT systems in translating English number idioms into Arabic. The prevalence of literal translations and the lack of cultural adaptation result in translations that often fail to convey the intended meaning. To address these issues, NMT systems should incorporate idiom-specific modules and culturally enriched datasets. Further research is needed to develop models capable of recognizing and accurately translating idiomatic expressions across languages. The results and statistics reveal that ChatGPT - despite it faces some serious problems- is better than Google translate in translating number idiomatic expressions. Also, it can be a reliable option for translation such kind of idiomatic expressions if it has the required training, support and improvement.

References

- [1] Alkhawaja, L., Ibrahim, H., Ghnaim, F., & Awwad, S. (2020). Neural Machine Translation: Fine-Grained Evaluation of Google Translate Output for English-to-Arabic Translation. *International Journal of English Linguistics*, 10(4), 43–52. <https://doi.org/10.5539/ijel.v10n4p43>
- [2] Baziotis, C., Mathur, P., & Hasler, E. (2022). Automatic Evaluation and Analysis of Idioms in Neural Machine Translation. arXiv preprint arXiv:2210.04545. <https://arxiv.org/abs/2210.04545>
- [3] Ben-Yahia, Moad (2024). Examining the Efficiency of Machine Translation in Translating English Idioms used in American Media. *Journal of Translation and Language Studies*, 2024 Volume 5, Issue 2: 43–55
- [4] DOI: <https://doi.org/10.48185/jtls.v5i2.1070>
- [5] Dankers, V., Lucas, C. G., & Titov, I. (2022). Can Transformer be Too Compositional? Analysing Idiom Processing in Neural Machine Translation. arXiv preprint arXiv:2205.15301. <https://arxiv.org/abs/2205.15301>
- [6] Haider, A., Darwish, N., Saed, H., Dagamseh, M., Alantari, D., & Tannous, B. (2025). A reception study of AI-translated idioms and proverbs between Arabic and English. *Research Journal in Advanced Humanities*, 6(3). <https://doi.org/10.58256/k4d6pp20>
- [7] Ibrahim Jibreel (2023). Online Machine Translation Efficiency in Translating Fixed Expressions Between English and Arabic (Proverbs as a Case-in-Point). *Theory and Practice in Language Studies*, Vol. 13, No. 5, pp. 1148-1158, May 2023.
- [8] DOI: <https://doi.org/10.17507/tpls.1305.07>
- [9] Klimova, Blanka (2025) Use of machine translation in foreign language education.
- [10] *Cogent Arts & Humanities*, Volume 12, 2025 - Issue 1. <https://doi.org/10.1080/23311983.2025.2491183>
- [11] Mounassar, A. A. A. (2024). The Effectiveness of Machine Translation Systems in Translating English Fixed Phrases into Arabic. *University of Science and Technology Journal for Management and Human Sciences*, 2(2), 91–116. <https://doi.org/10.59222/ustjmhs.2.2.4>
- [12] Palanichamy Naveen1 and Pavel Trojovsky', Overview and challenges of machine translation for contextually appropriate translations. *iScience: Volume 27, Issue 10*, 18 October 2024, 110878.