Research Paper

Volume 9, Issue 1
Winter, 2025

Journal of Teaching English Language Studies

Accepted: April, 2025 Published: July, 2025

Research Article

Translation Quality of Three MT Systems

Atefe Mazidi1, Guiti Mortazavizadeh2 (Corresponding Author)*

1Department of English Language, Ya.C., Islamic Azad University, Yazd, Iran

Email: mazidii@iau.ac.ir

2Assistant Professor, Department of English Language, Ya.C., Islamic Azad University, Yazd, Iran

Email: Mortazavizadeh@iau.ac.ir

Orcid ID: 0000-0003-4688-6540

ABSTRACT

This quantitative study performed a machine translation (MT) quality assessment on the accuracy and fluency of three MT systems: Google Translate, ChatGPT, and Xerac. The corpus of the study was The Alchemist, from which the total number of 173 idioms were chosen and translated by the MT systems. The translations were rated through Morgan's (2004) rubrics by two experts in translation. The data were analyzed through SPSS 28. The findings indicated that Google Translate significantly outperformed the other two systems. It provided translations that were not only more relevant to the original text, but also exhibited higher fluency for readers in their first language. The evaluation highlighted that Google Translate effectively applied linguistic rules pertinent to the target language, resulting in translations that felt more natural and contextually appropriate. In contrast, both ChatGPT and Xerac fell short in meeting the standards set by Google Translate, with evaluators consistently favoring the former for its superior performance. This suggests that Google Translate can be the best reliable MT system for translating texts, making it a valuable tool for users seeking accurate and fluent translations. Overall, the study underscores the importance of evaluating MT systems based on both accuracy and fluency, providing insights into their relative effectiveness in real-world applications. The findings of the study could have implications for trasnaltors and translator course designers and trainers.

Keywords: Accuracy, Fluency, Idiom, MT quality assessment, Google translate, Chat GPT, Xerac

1. INTRODUCTION

Since the onset of the twenty-first century, translation has begun to play a key role in supporting the globalization process and the use of the internet, Machine Translation (MT) and technology made the exchange of all sorts of information possible (Ceramella, 2008). Translation is both a cognitive procedure which occurs in a human being's, the translator's, head, and a social, cross-linguistic and cross-cultural practice. Any valid theory of translation must embrace these two aspects. To do this, a multidisciplinary approach to translation theory integrating these aspects in a plausible manner is needed (Sennrich et al., 2016). Furthermore, a theory of translation is not possible without a ref lection on the role of one of its core concepts: equivalence in translation and looking at equivalence leads directly into a discussion of how one would go about assessing the quality of a translation. Translation quality assessment can thus be said to be at the heart of any theory of translation (Daxbock, 2010).

According to Thao (2023), MT is the use of computational techniques to translate text or speech from one language to another, including the contextual, idiomatic and pragmatic nuances of both languages. Early approaches were mostly rule-based or statistical. These methods have since been superseded by neural MT and large language models such as DuPont 16 (Shardlow, 2014). The origins of MT can be traced back to the work of Al-Kindi, a ninth-century Arabic cryptographer who developed techniques for systemic language translation, including cryptanalysis, frequency analysis, and probability and statistics, which are used in modern MT (Tiedemann, 2018). The idea of MT later appeared in the 17th century. In 1629, René Descartes proposed a universal language, with equivalent ideas in different tongues sharing one symbol (Toral & Apidianaki, 2016).

According to Qian and Liu (2019), MT systems are particularly effective in translating from English to languages such as Chinese, Spanish, French, and others, producing high-quality translations. However, when translating from English to Persian, there are noticeable gaps, errors, and limitations. It is uncertain whether these systems can accurately translate culturally nuanced content. According to Ronagh Zadeh et al. (2021), one challenge in MT between English and Persian is idioms, as their meanings are not literal but culturally significant.

Idiom is a multiword expression and its meaning cannot be translated verbatim because idioms are fixed non-compositional expressions that belong to the intersection between language and culture (Bashir et al., 2021). Baker (1991) claims that the main problems that idiomatic and fixed expressions pose in translation relates to two main areas: the ability to recognize and interpret an idiom correctly and the difficulties in rendering various aspects of meaning that an idiom or a fixed expression conveys into the target language.

Creative translation is about capturing the essence and style of the original content and adapting it in a way that resonates with the target audience. This is only possible when the skill, creativity, and knowledge of the local culture are provided by a human translator. However, the use of MT in cross-lingual communication is growing (Conneau & Lample, 2019). MT is one of the resources with the greatest potential for providing different types of audiences with mutual access to information. Translation machines have certainly made the process of translation faster and more convenient (Conneau & Lample, 2019). Nevertheless, some of the translated books, articles, and texts are confronted with unsuitable and incomprehensible translations, which are due to the use of translation machines (Ronagh Zadeh et al., 2021). In the words of Madsen (2009), there are different problems encountered by MT. Some problems cannot be solved at all, and high-quality translation done solely by machines is not possible and machine-translated

texts will continue to be plagued by errors in the future, ranging from eccentric turns of phrase to grave distortions of meaning.

These problems are especially true for literary works and complex texts where a deep understanding of the source language is required. Some problems like lack of equivalence in the target language as languages vary and they express meanings using different linguistic means such as fixed and frozen expressions, idioms, etc. Therefore, it is very hard to find equivalents in the target language. Enhancing idiomatic MT quality has significant social implications because high-quality translation can preserve cultural nuances, ensuring that translated texts retain their intent and emotional resonance, thus fostering better cross-cultural communication. Idiom translation requires knowing both the source and target culture and mastering different translation strategies and is even difficult for human. According to Moon (1998), 3 out of 10 sentences are estimated to contain idioms.

Despite the advancements in MT, current MT systems often struggle to preserve the idiomatic writing style in translated texts. The challenge stems from accurately capturing the cultural and emotional essence embedded in idioms, which are highly context-dependent and vary across languages, and cultures (Thao, 2023). The majority of users understand the capabilities and limitations of MT, and they know that MT will be less than perfect results. Some also realize the challenges of MT such as name handling, idioms, and context-sensitive translation (Wang & Wang, 2019). MT systems focus only on linguistic analysis at morphological and lexical levels, and provide limited syntactical analysis, far from sufficient to support sophisticated translation, not to mention any kind of cultural understanding. Moreover, MT doesn't understand the context and emotions behind the words. In literary texts, tone and emotion are crucial, but MT tends to produce literal translations that lack such nuances (Wu & Wang, 2020).

These issues may result in errors and misunderstandings in the translated text, affecting the overall quality and authenticity of the translation. There is thus a pressing need to learn about the types and levels of mistakes and errors that MT systems make when deployed in translating idioms in the literary domain. Further research and development are necessary to enhance the precision and efficiency of MT systems in accurately conveying the intended meaning of idioms in literary texts.

Furthermore, considering that the need for translation in society is increasing day by day, it should be noted that translation machines are not of proper quality in checking and translating terms (Liu & Liu, 2020). Human translation also has its problems and limitations. From another point of view, it should be noted that translation machines lack context and are not familiar with the translation culture of each society, which cause many problems during translation (Kafi et al., 2018).

Although in recent years many works have reported on the evaluation of MT some of which use automatic evaluation systems (e.g., Popovic & Ney, 2011; Ma & Huang, 2019), only a few of them have worked on specific terms like register, lexis, or idioms, just to name a few, (e.g., Qian & Liu, 2019). Therefore, it seems that more work is necessary to be done in these domains. The main aim of the present study, as a result, was to evaluate and compare the performance of three online MT systems namely Goole translate, xeric, and Chat GPT in translating English idioms into Persian in the book titled *The Alchemist*. To this end, the following research questions were formulated:

1) What is the performance of three online MT systems (Google translate, Chat GPT, Xerac) regarding the accuracy and fluency of MT systems in translating English idioms in Alchemy into Persian?

2) What is the best MT system (Google translate, Chat GPT, Xerac) for translating English phrases in Alchemy to Persian?

2. LITRATURE REVIEW

2.1 MT (Machine Translation)

Machine translation (MT) has undergone transformative changes in recent years, evolving from rule-based systems to complex neural networks that leverage vast amounts of data and sophisticated algorithms. This evolution is crucial in an increasingly globalized world where the demand for real-time translation services is surging. The inception of neural machine translation (NMT) marked a paradigm shift in MT technology. The introduction of sequence-to-sequence (seq2seq) models by Sutskever et al. (2014) laid the groundwork for subsequent developments. By 2015, researchers began to explore the capabilities of NMT, demonstrating its superiority over traditional statistical methods. Cho et al. (2014a) highlighted how NMT reduces the need for extensive feature engineering, enabling more fluent and contextually appropriate translations.

By 2016, Google had announced the implementation of NMT in its translation services, a move that underscored the technology's potential and efficacy. Dabre et al. (2017) reported significant improvements in translation quality, particularly in handling long-range dependencies within sentences—a common challenge in MT. This advancement not only enhanced user satisfaction but also spurred further research into optimizing NMT architectures. The period from 2017 to 2019 witnessed a surge in research focusing on refining NMT systems. One notable advancement was the introduction of transformer models by Vaswani et al. (2017). This model eliminated the reliance on recurrent neural networks (RNNs) and allowed for parallel processing of data, which significantly improved training efficiency and translation quality. The transformer architecture has since become the backbone of many state-of-the-art MT systems (Vaswani et al., 2017).

Further explorations in NMT during this period also included the integration of attention mechanisms, which enabled models to focus on specific parts of the input sentence while generating translations. The work by Bahdanau et al. (2016) demonstrated that attention mechanisms could address the limitations of fixed-length encodings in RNNs, resulting in more accurate translations (Bahdanau et al., 2016). The attention mechanism's effectiveness has led to its widespread adoption, influencing various applications beyond MT, including natural language processing (NLP) tasks such as text summarization and sentiment analysis.

As we move into 2024, the landscape of machine translation continues to evolve, influenced by advancements in artificial intelligence and machine learning. The emergence of large language models (LLMs) has further enhanced MT capabilities, providing more contextually aware translations. Additionally, research into explainable AI in MT is gaining momentum, aiming to make the decision-making processes of translation models more transparent and interpretable.

It can be noted that the advancements in machine translation from 2015 to 2024 represent a remarkable journey characterized by significant technological innovations and a deeper understanding of linguistic complexities. As the field continues to evolve, ongoing research will be vital in addressing the challenges and ethical considerations that accompany these advancements. The future of machine translation holds promise, with the potential to bridge linguistic divides and foster communication across cultures.

2.2 Empirical Studies

In one study, Dweik and Thalji (2016) explored the translation of English idioms into Arabic, highlighting the complexities involved and the pedagogical implications. It emphasized the importance of cultural awareness and effective translation strategies when dealing with idioms. The findings revealed that context alone may not always lead to accurate interpretation, and students should be aware of potential misinterpretations. The study recommends incorporating cultural translation into the curriculum, minimizing the use of paraphrasing, and providing training on effective translation strategies for idioms. Overall, the study underscores the significance of cultural awareness and effective translation strategies and offers valuable recommendations for language instruction and translation to enhance students' proficiency in translating idioms and culturally bound expressions.

In another study, Al-Khresheh and Almaaytah (2018) examined the challenges of translating English proverbs into Arabic using MT. The study highlighted the limitations of MT in accurately translating proverbs, such as incorrect equivalents, literal translation, and grammatical errors. The study also stressed the need for human intervention to address accuracy issues, as MT alone cannot effectively handle the complexities of translating ambiguous expressions like proverbs. The findings align with pedagogical implications, emphasizing the cultural and linguistic disparities between English and Arabic and the importance of providing accurate functional equivalents. The study recommends improving MT accuracy through human intervention, comprehensive databases, and contextual analysis to enhance the quality of online translation.

In another study, Taleghani and Pazouki (2018) discussed the evaluation of four free online translators in terms of their ability to translate English idioms into Persian. The study compared the performance of www.bing.com, www.translate.google.com, www.freetranslation.com, and www.targoman.com in translating idiomatic phrasal verbs from the book "Oxford Word Skills: Idioms and Phrasal Verbs." The results show that www.targoman.com performs better in translating idioms from English to Persian, making it the preferred choice for this specific task. Overall, the study highlights the shortcomings of free online translators and the need for further improvement in this area.

In another study, Al-Rushaidi (2017) investigated the challenges faced by Omani undergraduate students majoring in English language teaching and literature when translating idiomatic and culturally bound expressions from English into Arabic. It also aimed to identify the strategies used by the students to overcome these challenges and suggested ways to improve their translation skills. The research involved administering a translation test and survey to 60 students, and the findings highlighted the difficulties encountered and the coping strategies employed by the students. The study also offers recommendations and pedagogical implications to address these challenges and enhance the students' translation competencies.

Another research entitled *strategies used to translate idioms and proverbs which are influenced by region* has been done by Rushaid (2010). The purpose of study was to investigate the way the Persian idioms were translated into English based on Baker's translation strategies for idioms. The researcher's concentration was on the Persian idioms which included elements such as name of animals, plants, wind, rain, sun and so on in their lexical constructions. To do the study, she collected several idioms and proverbs of Persian language which were mostly used in two provinces of Iran, Yazd and Mazandaran.

3. METHODOLOGY

3.1. Design of the Study

This study was conducted within a qualitative descriptive approach based on machine translation quality assessment (TQA) based on the rubrics proposed by Morgan (2004).

3.2. Corpus of the Study

The Alchemist book was selected as the corpus in this research. The Alchemist is a novel written by Paulo Coelho (Coelho, 2002), which narrates the story of a man who is in search of a valuable thing in his life. After that, sampling was done from within this novel. Sampling in this research is purposive. The purposive method was chosen because this research was only looking for sentences in which there were idioms. The Alchemist was investigated for the idioms and the total number of 200 idioms were found, some of which were repetitive. By counting the type of the idioms found in the book, the total number of idioms selected from the book was 173 that went under investigation. In this study, the sampling method applied was purposeful sampling. In fact, the idioms were considered in the study and their number was counted by looking through The Alchemist. Purposeful sampling, also known as intentional sampling, is a non-probability sampling technique widely used in qualitative research to select participants who possess specific characteristics or qualities relevant to the research question. This strategy allows researchers to gain in-depth insights by focusing on individuals who are most likely to provide rich, relevant information (Creswell & Poth, 2016).

3.3. Data Collection and Analysis Procedure

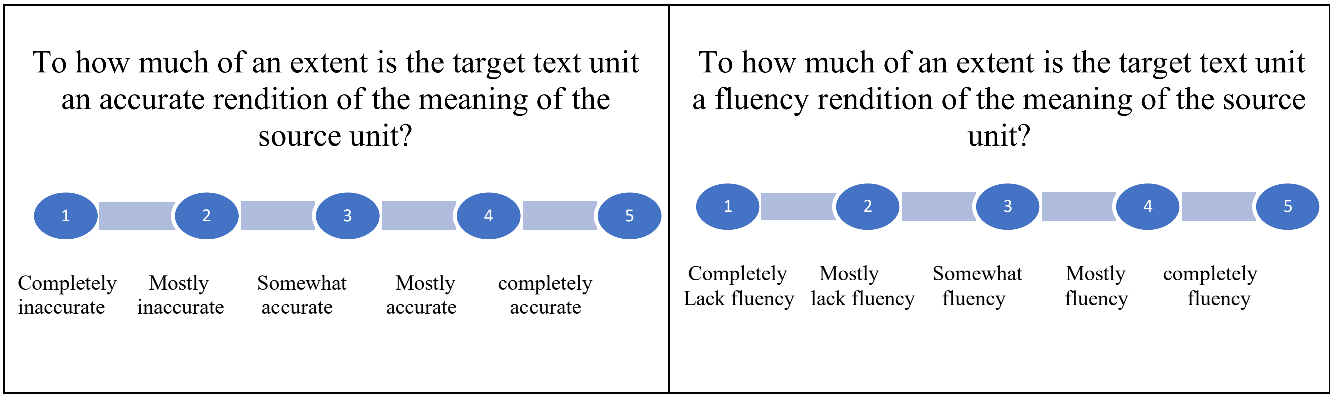
The data collection procedure for this study involved the purposive selection of 173 idioms from the novel *The Alchemist*. In fact, the total number of idioms selected from the book were 200 that some of them were repetitive. By considering the type of idioms, 173 ones were left and investigated in the study. For data collection, the whole text of *The Alchemist* was looked into and two-word terms were extracted and written in a table. As the next step, the idioms were given to MTs and the translations were provided in front of each idiom in the table.

Then idioms were given to Google Translate, Chat GPT ,and Xerac and the translation of these tools were listed (see Appendix). The obtained items were recorded for further analysis in terms of accuracy and fluency of translation. Two separate lists containing Englisg idioms and their translations by Google Translate, Chat GPT and Xerac, were given to two expert raters. The raters were experts in translation, who had great experience. They worked based on the scale adopted from Morgan (2004), whose scoring rubrics was based on the scale of 1-5. Number 1 as a score in this rubric means that the translation lacked fluency and/or was totally inaccurate. Number 5, on the other hand, showed that the idiom was translated totally accurately and/or fluently. They provided the aforementioned scale for the translations and the data were imported into SPSS for further analysis.

The translations were conducted with a focus on maintaining the contextual and cultural significance of the original idioms. Careful attention was given to capturing the nuances and metaphorical meanings embedded within the extracted idioms during the translation process. The evaluation of the translation was carried through using the translation quality assessment (TQA) method. One of the key questions in translation studies is how to effectively assess translation quality. The form used for TQA in the study is provided in Figure 1 below.

Figure 1

TQA Form for the Evaluation of the Quality of Translations



According to House (2009), TQA will mean a constant to and from a macro-analytic approach, wherein questions of ideology, function, gender, or register are considered, to a micro-analytical one in which the value of collocations and individual linguistic units are considered. TQA typically focuses on accuracy and fluency, readability, comprehensibility, usability, and acceptability of translations. To evaluate the accuracy and fluency of the translated idioms, two Ph.D. holders who were experts in translation were selected. The raters separately evaluated the translated texts and provided the results. The results were later analyzed by SPSS 28 and the reports were presented in tables and figures for further discussion.

In this study, the accuracy and fluency of translations were rated based on the rubric presented in the Figure 1, based on which the raters worked separately on the data and provided their opinions. For the purpose of measuring the agreement between the raters two Kohen's k test were conducted through SPSS 28 to estimate the agreement on accuracy and fluency, respectively. To compare the performance of MT systems in terms of accuracy and fluency, two Kruskal-Wallis tests and two Mann-Whitney tests were conducted.

4. RESULTS AND DISCUSSIONS

To answer the first research question How does the performance of the three online systems regarding the accuracy and fluency vary on translating idioms in *The Alchemist*?, the data was investigated through the descriptive statistics comparing the outputs from Google Translation, Chat GPT and Xerac. The findings of the study in this regard are presented in Table 1 below.

Table 1

Descriptive Statistics of the Data regarding the Three MT Systems

Online MT Systems	N	Number (%) of idioms with the highest score	Median score (IQR)	Mean Rank	Sig
Google Translation	173	10	7	380.77	

Chat GPT	173	4	5	278.13	
Xerac	173	1	3	121.10	*<0.001

According to the Table 1, the mean rank of fluency scores was 380.77 for Google Translate, 278.13 for Chat GPT, and 121.10 for Xerac respectively. This indicates that Google Translate tended to get a score higher than Chat GPT and Xerac (median = 7 for Google Translation, vs. 5 for Chat GPT, and 3 for Xerac; p-value < .001).

However, for the purpose of identifying whether the observed differences were statistically significant, a Kruskal-Wallis test was conducted, Table 2 presents the results.

Table 2

Kruskal-Wallis Test Comparing the Three Translation Machines regarding their Fluency

	Score
Chi-Square	49.13
df	2
Asymp. Sig.	.000

A significant Kruskal-Wallis test indicates that at least one sample stochastically dominates another sample ($p < .001$). However, the test does not specify where this stochastic dominance occurs, or for how many pairs of groups it applies. To analyze the specific sample pairs for stochastic dominance, pairwise Mann-Whitney tests with Bonferroni correction were conducted. Post hoc comparisons were conducted using Mann-Whitney Tests with a Bonferroni adjusted alpha level of .016. the results of the Mann-Whitney U-tests are provided in Table 3.

Table 3

Comparison of the Fluency of the Translation Machines Through Mann-Whitney U-Tests

	Z	**P-Value**
Google Translate vs. Chat GPT	-10.296	.000
Google Translate vs. Xerac	-10.076	.000
Chat GPT vs. Xerac	-8.790	.056

The difference in Fluency Median Score was found to be statistically significant between Google Translate and hat GPT ($z = -10.296$, p-value < 0.001), and also between Google Translate and Xerac ($z = -10.076$, p-value < 0.001). However, there was no significant difference between the fluency score the raters provide for Chat GPT and Xerac ($Z = -8.790$, $p = .056$).

To investigate the accuracy of the translation provided by MT systems, the accuracy scores given to the three MT systems were compared and the descriptive statistics are provided in Table 4.

Table 4

Descriptive Statistics of the Accuracy of MT Systems

Online MT Systems	**N**	**Number (%) of idioms with the highest score**	**Median score (IQR)**	**Mean Rank**	**Sig**
Google Translate	173	15(8.7)	6(4-8)	323.17	
Chat GPT	173	6(3.5)	2(4-6)	230.01	
Xerac	173	7(4)	2(4-6)	225.14	<0.001

The comparison between the mean ranks of the MT systems showed that the highest rank belonged to the Google Translate regarding the accuracy of translation (mean rank = 323.17, $p < .001$). The mean rank of the Chat GPT put it in the second order (mean rank = 230.01, $p < .001$), and Xerac stood in the third rank (mean rank = 225.14, $p < .001$).

For comparing the accuracy scores, a Kruskal- Wallis test was conducted, whose results are provided in Table 5.

Table 5

Kruskal-Wallis Test Comparing the Three Translation Machines regarding their Accuracy

	Score
Chi-Square	269.23
df	2
Asymp. Sig.	.000

As the Table 5 indicates, there was a statistically significant difference between MT systems in the scores they gained for their accuracy ($H(2, N=519) = 269.23$, $p < .001$).

Nevertheless, the pair-wise comparison of the MT systems required the conduction of the Mann-Whitney U tests, whose results are provided in Table 6.

Table 6

Comparison of the Accuracy of the Translation Machines Through Mann-Whitney U-Tests

	Z	***P*-Value**
Google Translate vs. Chat GPT	-7.623	.023
Google Translate vs. Xerac	-9.236	.001
Chat GPT vs. Xerac	-0.790	.139

The pair-wise comparison of the MT systems showed that Google Translate was statistically different from Chat GPT (Z = 7.623, p = .023) and Xerac (Z = 9.236, p = .001). however, there was no statistical difference between the estimated accuracy of Chat GPT and Xerac (Z = .790, p = .139).

To address the second research question, What is the best MT system for translating English phrases in *The Alchemist* to Persian?, one must consider both accuracy and fluency of the MT systems investigated in the study, which are provided in Table 7 below.

Table 7

Comparison of the Fluency and Accuracy of the MT Systems

Evaluation criteria	**Google Translation Median(IQR)**	**Chat GPT Median(IQR)**	**Xerac Median(IQR)**
Fluency	7(6-9)	5(4-6)	3(2-3)
Accuracy	6(4-8)	2(4-6)	2(4-6)

The results indicated that Google Translate (fluency = 7; accuracy = 6) outperformed Chat GPT (fluency = 5; accuracy = 2) and Xerac (fluency = 3; accuracy = 2) in terms of fluency and accuracy and achieved higher scores in presenting translations that were more natural and coherent.

The results of the analysis indicated that the answer to the second research question, investigating the best MT system, is Google Translate since it achieved the highest scores on both accuracy and fluency.

The first research question posed in the study explored the accuracy and fluency of English-to-Persian translations produced by three different machine translation systems, namely Google Translate, ChatGPT, and Xerac. The findings of the study gained through the analysis of the data collected by translating the *The Alchemist* via three mentioned AI-based translators showed that Google Translate was the most accurate MT system in translating idioms.

The findings of the study were in line with the one conducted by Zhang (2024). He found out that Google Translate generally performs well in terms of accuracy, especially when translating scientific texts.

In addition, Khoshafa (2023) figured out that Google Translate is a system, which relies on NMT and a large bilingual corpus, therefore, it can capture many technical terms and scientific concepts effectively.

However, certain domain-specific phrases and complex sentence structures in scientific articles present challenges while being translated by Google translate. In some instances, Google Translate struggles with technical jargon, providing awkward or imprecise translations. Moreover, some scientific terms lack contextual accuracy, leading to slight misinterpretations or ambiguity (Rahman & Saputra, 2021).

The findings of the current study showed that ChatGPT and Xerac demonstrated the same level of accuracy in translating text from English to Persian in *The Alchemist*. These findings have consistency with the work carried out by Aghai (2024), who believed that Chat GPT, being based on advanced deep learning models, displays a high degree of accuracy in understanding context and producing more nuanced translations. They mention that it demonstrates strength in translating more complex, idioms found in scientific writing. However, there are occasional instances where the translation is slightly off due to the model's reliance on general knowledge rather than specialized scientific vocabulary (Kafi et al., 2018). This occasionally leads to translation errors, particularly when the input includes specialized terms that are less common in its training data. Despite these instances, Google Translate often provides translations that better reflected the meaning of the original text (Rahman & Saputra, 2021), which is in consistence to the findings of the study relying on the analysis of data. Furthermore, Budianto (2020) mention that Xerac, as a more domain-specific translation tool, has shown promising accuracy in translating scientific texts into Persian that could be a good explanation for phenomenon that in the study, the mentioned MT system did not perform as well as the two other ones under investigation.

Xerac is designed with scientific and technical translation in its codes and it excels in translating highly specialized language and technical terminology. Aghai (2024) also mention that ChatGPT and Xerac are found to be more accurate than Google Translate in terms of preserving the meaning of scientific concepts, though each have its strengths and weaknesses depending on the domain and context of the text.

Regarding the fluency of the translations provided by the three AI-powered translated under investigation, the findings of the study showed that Google Translate had the highest fluency in translation followed by Chat GPT having Xerac in the last rank in fluency. In this study, fluency referred to how natural and smooth the translated text read in the target language. A translation was considered fluent if it adhered to the syntactic, grammatical, and stylistic norms of the target language. Based on the findings of the study, Google Translate had the highest fluency among the other three. Chat GPT had the second rank followed by Xerac standing in the third rank based on the quality of fluency in translation.

The findings were in line with Khoshafa’s (2023) study, in which they declared that, in terms of fluency, Google Translate has made significant strides over the years, particularly with the integration of NMT. Nevertheless, Aghai (2024) state that while the Persian translations are generally grammatically correct, the fluency of the text is sometimes problematic due to literal translations of idioms or syntactic structures that are not typical of Persian. In literary texts, this results in some awkward phrasing or slightly erroneous language. While Google Translate is useful for everyday translation, its output in literary contexts often feel mechanical or overly simplistic (Khoshafa, 2023).

According to Aghai (2024), ChatGPT's translations are generally more fluent than those of Google Translate, which is in contrast with the findings of the study. This is largely due to its ability to generate more contextually appropriate translations and its understanding of discourse and sentence flow (Khoshafa, 2023). ChatGPT's neural network allows it to handle more intricate sentence structures and produce translations that sound natural and are more stylistically aligned with the Persian language, according to Rahman and Saputra (2021). However, according to Khoshafa (2023), the fluency of ChatGPT's translations is sometimes affected by issues related to idioms, which occasionally results in slightly awkward or unnatural phrasing, which supports the results gained from the data analysis in this study.

Based on the findings of the study, the fluency of Xerac was lower than the other three AI-translators, but the accuracy was lower than Google Translate and identical with ChatGPT. According to Aghai (2024), Xerac performs the best with more colloquial phrases. In such cases, the fluency of the translation suffers, as the system tries to adapt its highly formal register to less formal English structures ((Khoshafa, 2023). Budianto (2020) believe that Xerac is specialized in scientific texts, generally providing the most fluent translations, particularly for technical language and formal writing. The fluency of the Persian output is in line with the conventions of scientific writing, maintaining formal tone and technical precision.

Regarding the second research question, posing a question on the quality of translation, the results of this study showed that Google Translate had the best results in both accuracy and fluency compared to the other two ones while investigating the challenges of translating a literary content, where accuracy is critical, however, fluency is also important for readability. While each translation system performed admirably in certain aspects, Google Translate outperforms the others across all measures of accuracy and fluency, which is in line with the study conducted by Khoshafa (2023). Contrary to these findings, Budianto (2020) believe that the importance of the translator's domain expertise becomes evident, as specialized tools like Xerac provide better accuracy in the scientific domain but sometimes lack fluency in more nuanced contexts.

Google Translate is one of the most widely used machine translation tools, powered by NMT and vast bilingual corpora. In terms of idiomatic translation, Google Translate performs reasonably well with common idioms found in everyday language, according to Zhang (2024), that supports the discussion provided in this study. Based on the work conducted by Khoshafa (2023), Google Translate typically struggles with idioms that have no direct equivalent in the target language, often resorting to literal translations that do not make sense in Persian. Aghai (2024) state that while some idioms are accurately translated due to the availability of equivalent phrases in Persian, many do not carry the same philosophical or metaphorical weight as the original English terms.

The study that supports the findings on the current research is the one conducted by Khoshafa (2023). They believe that even ChatGPT occasionally faced challenges with less commonly used idioms. In these cases, it sometimes generated translations that were partially inaccurate or lacked full alignment with the original conceptual meaning.

Nevertheless, contrary to the claim of the study, ChatGPT, based on advanced deep learning models, outperformed Google Translate in terms of understanding the context and meaning behind idioms, according to Aghai (2024). According to Boulton and Vyatkina (2021), unlike Google Translate, which often defaults to literal translations, ChatGPT is capable of interpreting idiomatic phrases in context, making it better suited for rendering expressions that require a deeper understanding of metaphor or philosophical concepts that are contrary to what the current findings showed.

According to Budianto (2020), Xerac, which specializes in scientific and technical translations, showed mixed results when translating idioms. Its strength lies in handling domain-specific terminology, and it often produced accurate translations of technical phrases and concepts. Moreover, based on the study conducted by Aghai (2024), Xerac struggles with idioms, particularly those outside its core domain of

scientific language. The system tends to translate idioms too literally and fails to capture the subtleties or philosophical implications behind the idioms.

According to Ronaghzade et al. (2021), idioms in *The Alchemist* often involve cultural and historical references that are not easily understood outside of context. For a machine translation system to convey the full meaning of an idiom, it must not only recognize the metaphorical significance, but also understand the historical and cultural context in which it was used (Khoshafa, 2023). When translating idioms, machine translation systems often face the risk of semantic loss, where the metaphorical or symbolic meaning of the idiom is not accurately transferred to the target language.

The results of the analysis also showed that regarding accuracy, fluency and applying Persian norms and rules, Google Translate outperformed the rest in providing a translation for *The Alchemist*. Since some studies found contrary results compared to this study, the reasons for the contrast were discussed in the previous subsection, 5.2.2.

The findings are supported by Zhang (2024), who claims that Google Translate performs reasonably well for more general or widely used idioms, particularly when the terms have equivalent or widely accepted translations in Persian. Additionally, Google Translate's continuous updates and large data corpus allow it to adapt somewhat to newer phrases and idioms, offering basic translations that might work for everyday or introductory texts on literary texts (Khoshafa, 2023).

Nevertheless, research by Dicks (2018) highlights Google Translate's utility for general language but notes its limitations in handling idioms. This is discussed to be due to its reliance on large corpora and a one-size-fits-all approach to translation. Furthermore, although Aghai (2024) claim that ChatGPT outperforms other machine translators, they believe that it is not flawless. It may also generate translations that, while contextually appropriate, are not always grammatically perfect or stylistically suited to the formal tone often required in the text.

Research by Boulton and Vyatkina (2021) and Hadi and Ghorbani (2021) suggest that ChatGPT performs exceptionally well in idioms translations due to its ability to grasp meaning within context, but it still requires fine-tuning for accuracy in more arcane or niche subject areas. However, in this study we found out that Google Translate performed better in this realm.

According to Aghai (2024), Xerac's major limitation is its struggle with figurative or metaphorical language as Budianto (2020) claimed that Xerac is a machine translation system designed specifically for technical and scientific translations. Xerac's performance has been critically evaluated by Aghai (2024), who highlighted the system's strength in technical translations but acknowledged its limitations when faced with non-literal, abstract concepts.

5. CONCLUSION

In general, the results gained through the analysis of data demonstrated that Google Translate performed best in terms of fluency, offering translations that were stylistically smoother and more natural in Persian, especially in complex contexts. Xerac stood out in its accuracy for technical terms, but at times lacked the fluency that Google Translate could offer. ChatGPT, while accurate in many cases, fell short in terms of accuracy, especially in translating idioms where its literal translations often disrupted the natural flow of the Persian language.

The translation quality of English idioms in *The Alchemist* into Persian by Google Translate, ChatGPT, and Xerac revealed both strengths and weaknesses in the systems. Google Translate outperforms the other two systems in terms of capturing the contextual and metaphorical meaning behind idioms. ChatGPT and Xerac, on the other hand, often fail to produce fluent and accurate translations of idioms, particularly those unique to *The Alchemist*. This highlights the need for ongoing advancements in machine translation, particularly for specialized domains that require nuanced understanding and cultural sensitivity.

To summarize the results, we can say that Google Translate offers the most promising results due to its contextual understanding and flexibility in translating metaphorical and symbolic language. Its capacity to adapt to specialized contexts makes it a strong contender, although occasional translation errors can occur with highly obscure terms. ChatGPT is suitable for basic, technical translations but fails when dealing with figurative or culturally rich terms. Xerac while excellent for technical terms, falls short when translating the metaphorical and symbolic aspects of the language. Its rigid approach makes it less suitable for this highly nuanced domain.

Therefore, Google Translate emerges as the best machine translation system for translating English phrases in *The Alchemist* into Persian, particularly when considering its ability to handle context, metaphor, and the nuanced nature of the language. However, it is essential to note that no machine translation system currently excels in all aspects, and further development and fine-tuning of specialized MT systems are necessary for optimal translation quality in this unique field.

The findings of this study highlight the importance of improving machine translation systems for specialized fields and genres. While current systems like Google Translate, ChatGPT, and Xerac are powerful tools, they still face significant challenges when it comes to translating idioms that are deeply embedded in cultural, historical, and philosophical contexts. Therefore, the implications of the study will be for AI-based program designers, who can figure out what features are necessary to be added to the currently-used translators benefiting operators online or offline.

The findings of this study have some implications for those dealing with translation of texts from English to Persian. As idioms often pose significant challenges for machine translation systems due to their non-literal meanings and cultural specificity, this study focused on comparing how well these systems handled idioms in literary texts, particularly those related to *The Alchemist*. By analyzing their ability to convey both the intended meaning and cultural context, this research provides insights into the strengths and limitations of each system. Therefore, the people working with them might figure out how much they can trust the translations provided by the translators.

REFERENCES

Aghai, M. (2024). ChatGPT vs. Google Translate: Comparative analysis of translation quality. *Translation Studies, 22*(85), 85-100. DOR: 20.1001.1.17350212.1403.22.1.9.2.

Al-Khresheh, M. H., & Almaaytah, S. A. (2018). English Proverbs into Arabic through Machine Translation. *International Journal of Applied Linguistics & English Literature, 7*, 158-166.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *Proceedings of the 3rd International Conference on Learning Representations (ICLR).* https://arxiv.org/abs/1409.0473

Baker, M. (1991). *In Other Words: A Coursebook on Translation.* Routledge.

Bashir, A., Aleem, M., & Anjum, M. A. I. (2021). Pakistani translators' strategies for translating English idioms. *Ilkogretim Online, 20*(5), 5110-5116.

Boulton, A., & Vyatkina, N. (2021). Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology, 25*(3), 66–89. https://psycnet.apa.org/record/2022-10654-005

Budianto, L. (2020). EFL learner's perception about utilizing technology-driven learning in the midst of Covid-19 outbreak. *International Journal of Innovation, Creativity and Change, 14* (4), 69-80. http://repository.uin-malang.ac.id/8023/

Ceramella, S. (2008). The role of translation in globalization. *International Journal of Translation Studies, 2*(3), 45-60.

Cho, K., van Merrienboer, B., Gulcehre, C., Bougares, F., Schwenk, H., and Bengio, Y. (2014a). Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*.

Conneau, A., & Lample, G. (2019). Cross-lingual language model pretraining. *Advances in Neural Information Processing Systems, 32*, 7057-7067. https://proceedings.neurips.cc/

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches (4th ed.).* Sage Publications.

Dabre, R., Cromieres, F., & Nakazawa, T. (2017). Neural machine translation: Basics, practical aspects and recent trends. *Proceedings of the 8th International Joint Conference on Natural Language Processing*, pages 11–13, Taipei, Taiwan.

Daxbock, B. (2010). Text and Context: The Role of Equivalence in Translation Quality Assessment. *Journal of Translation Studies, 4*(2), 123-137.

Dicks, B. (2018). The impact of AI-powered translation tools on language learning and practice. *Language and Education, 32*(1), 50-65. https://doi.org/10.32996/ijllt.2025.8.3.30

Dweik, B. S., & Thalji, M. B. (2016). Strategies for translating proverbs from &English into Arabic. *Academic Research International, 7*(2), 120-127.

Hadi, G., & Ghorbani, S. (2021). Comparative analysis of Google Translate and ChatGPT in translating idiomatic expressions. *Journal of Applied Linguistics and Translation, 22(*1), 89-105. DOI: https://doi.org/10.33474/j-reall.v6i1.23252

House, J. (2009). *Translation.* Oxford University Press.

Kafi, M., et al. (2018). Challenges in machine translation of Persian: A review. *Persian Studies Journal, 12(*4), 75-89.

Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. https://doi.org/10.21203/rs.3.rs-2814154/v1

Liu, Q., & Liu, H. (2020). Improving machine translation quality with unsupervised pre-training. *Journal of Artificial Intelligence Research, 68,* 1143-1165.

Ma, J., & Huang, J. (2019). On the impact of training data quality on neural machine translation. *Journal of Computer Science and Technology, 34*(2), 245-263. https://doi.org/10.1007/s11390-019-1940-6

Madsen, H. (2009). The challenges of literary translation and machine assistance. *Translation Journal, 13*(4), 66-79.

Moon, R. (1998). *Fixed expressions and idioms in English*. Clarendon Press.

Popovic, M., & Ney, H. (2011). Towards the evaluation of machine translation quality: A study of the impact of fluency and fidelity. *Computational Linguistics, 37(*4), 799-806. https://doi.org/10.1162/COLI_a_00055

Qian, D., & Liu, Y. (2019). Evaluating Machine Translation of Cultural Terms: A Case Study of English-Persian. *International Journal of Language Resources and Evaluation, 53*(3), 367-383.

Rahman, F., & Saputra, N. (2021). English as international language revisited: Implications on South Korea's ELT context. *Journal of English Language Teaching, 6*(1), 8-15. DOI: http://dx.doi.org/10.30998/scope.v6i1.9383

Ronagh Zadeh, S., et al. (2021). Challenges in translating idioms from English to Persian. *Journal of Language and Cultural Studies, 15(*2), 112-129.

Rushaid, A. (2010). The evolution of translation technology: From rule-based to neural machine translation. *Journal of Language and Computer Science, 5*(2), 77-92.

Sennrich, R., Haddow, B., & Birch, A. (2016). Neural machine translation of rare words with subword units. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL),* 1715-1725. https://doi.org/10.18653/v1/P16-1162

Shardlow, M. (2014). *Neural Machine Translation: A Review. Journal of Machine Learning, 1*(1), 1-10.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. *Advances in Neural Information Processing Systems, 27*, 238-245.

Taleghani, M., & Pazouki, E. (2018). Free Online Translators: A Comparative Assessment in Terms of Idioms and Phrasal Verbs. *International Journal of English Language & Translation Studies. 6*(1). 15-19.

Thao, N. (2023). An Investigation into Google Translates Translation of English Proverbs into Vietnamese. *International Journal of Social Science and Human Research 7*(2), 89-97.

Tiedemann, J. (2018). Neural machine translation: A review. *Journal of Natural Language Engineering, 24*(1), 1-30. https://doi.org/10.1017/S135132491800014X

Toral, A., & Apidianaki, M. (2016). The role of post-editing in the evaluation of machine translation systems. *Machine Translation, 30*(3), 257-274. https://doi.org/10.1007/s10590-016-9208

Vaswani, A., Shard, N., et al. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems, 30*, 1-19.

Wang, Y., & Wang, Y. (2019). Evaluating the fluency and adequacy of machine translation: A review and a new framework. *ACM Transactions on Intelligent Systems and Technology, 10*(4), 1-30. https://doi.org/10.1145/3310561

Wu, H., & Wang, Y. (2020). Challenges in literary machine translation. *Translation Studies, 13*(2), 210-226.

Zhang, M. (2024). A study on the translation quality of ChatGPT. *International Journal of Educational Curriculum Management and Research, 5*(4), 25-32. Doi: 10.38007/IJECMR.2024.050121.