ARTICLES

AI-based processing of poetic language and human translation in literary contexts

Rashad Ahmed [a], Abdu Alkadi [b, *], Jamal Kaid Mohammed Ali [c]

[a] *Jacksonville State University, USA*
[b] *University of Bergen, Norway*
[c] *University of Bisha, Saudi Arabia*

Received October 9, 2025; Revised November 3, 2025; Accepted November 12, 2025

Abstract. As Artificial Intelligence (AI) continues to redefine the boundaries of linguistic research, this study examines the extent to which machine translation (MT) and AI tools can go beyond literal meaning, push beyond surface-level syntax and semantics to process context-sensitive issues in literary translation. While traditional MT systems such as Google Translate and Microsoft Translator are optimized for direct source-to-target mapping, AI language models like ChatGPT represent a broader category of tools designed for general-purpose language generation, which includes but is not limited to translation. Using a 14-line Arabic poem, translations were generated by three MT systems, one AI model (ChatGPT), and two certified human translators. These outputs were evaluated against ten linguistic and stylistic dimensions: punctuation, layout, rhyme, mood, theme, logico-semantics, transitivity, field, tenor, and mode. The six versions of translation were compared through a framework grounded in systemic functional linguistics (SFL). The analysis also considers how humans process cognitive-linguistic features when rendering poetic language. Results indicate that ChatGPT outperformed both MT systems and human translators in structural and semantic coherence, as well as in preserving poetic features such as rhyme and mood. However, all automated systems struggled with context-rich dimensions like tenor and mode, underscoring the enduring value of human interpretive depth. The findings highlight the potential of AI language models to complement, rather than replace, human expertise in literary translation and advocate for hybrid approaches that integrate computational efficiency with poetic language and cultural sensitivity.

Keywords: ChatGPT, language processing, machine translation, literary translation.

Ахмед Рашад, Алькаді Абду, Мохаммед Алі Джамал Каід. Відтворення поетичної мови інструментами на основі Штучного Інтелекту та переклад людиною в літературних контекстах.

Анотація. Оскільки Штучний Інтелект (ШІ) дедалі більше розширює межі лінгвістичних пошуків, це дослідження вивчає, в якій мірі засоби машинного перекладу

* Corresponding author. Abdu Alkadi. 0000-0003-3805-7507, abdu.al-kadi@uib.no

East European Journal of Psycholinguistics, 12(2), 10–32, https://doi.org/10.29038/ahm

(МП) та ШІ здатні виходити за межі буквального значення, поверхневої синтаксичної та семантичної структури для обробки контекстно-залежних аспектів літературного перекладу. Хоча традиційні системи МП, такі як Google Translate та Microsoft Translator, оптимізовані для прямого перекладу з мови оригіналу на мову перекладу, мовні моделі ШІ, наприклад ChatGPT, представляють ширшу категорію інструментів загального призначення для мовного генерування, що також включає переклад, але не обмежується ним. Ґрунтуючись на прикладі 14-рядкового арабського вірша, переклади були згенеровані трьома системами МП, однією моделлю ШІ (ChatGPT) та двома сертифікованими перекладачами-людьми. Результати порівняли за десятьма лінгвістичними та стилістичними параметрами: пунктуація, організація, рима, спосіб, тема, логіко-семантика, транзитивність, предмет, характер та метод. Шість версій перекладу порівняли за допомогою моделі системно-функціональної лінгвістики. Аналіз також враховував, як перекладач-людина обробляла когнітивно-лінгвістичні особливості під час перекладу поетичної мови. Результати засвідчили, що ChatGPT перевершив обидві системи машинного перекладу та перекладачів-людей за структурною та семантичною узгодженістю, а також за збереженням поетичних особливостей, таких як рима та спосіб. Однак усі автоматизовані системи мали труднощі з контекстуальними аспектами, такими як характер і метод, що підкреслює незамінність тлумачення твору людиною. Результати дослідження підкреслюють потенціал мовних моделей як допоміжного методу, а не заміни людини в літературному перекладі, чим підтримують гібридні підходи, що поєднують ефективність машини з поетичною мовою та культурним контекстом.

Ключові слова: ChatGPT, обробка мови, машинний переклад, літературний переклад.

Introduction

With the rapid growth of digital technology, machine translation (MT) has changed the traditional concept of translation of converting texts from one language to another. Although translation is now married to digital tools and apps, including Google Translate (Lee, 2017) and ChatGPT (Sahari et al., 2023), the quality of MT has been, perforce, a debatable topic (Horbach et al. 2023; Macken et al., 2020; Visby, 2020). While traditional machine translation systems such as Google Translate and Microsoft Translator are optimized for direct source-to-target mapping, AI language models like ChatGPT represent a broader category of tools designed for general-purpose language generation, which includes but is not limited to translation. The quality of traditional MT is generally touted as less professional, compared to supervised human translation (HT) (Lee, 2017; Luo, 2018), especially when it comes to stylized translation, where these systems struggle with idiomatic expressions and maintaining coherence (Škobo & Petričević, 2023). Luo (2018) speculated that automatic translation neither jeopardizes nor exceeds human translators' existence, and “the performance of machine translators can never be primary and more creative than that of human translators” (p. 2).

Nevertheless, with the tremendous advances in AI, it is crucial to revisit the potential of such machine translators. Research on the current state-of-the-art in both MT and HT focused on improving the accuracy of machine translation (Lee, 2023; Macken et al., 2020; Maučec & Donaj, 2019; Paterson, 2023). Advances in Neural MT and other deep learning techniques have led to more accurate translations, and many MT systems now incorporate contextual information and other knowledge sources to improve the quality of their translations (Horbach et al., 2023; Wang, 2023). However, HT remains a preferred choice, particularly in fields where accuracy and cultural understanding are crucial, such as legal, medical, and technical fields.

Machine and human translations have been compared in various contexts. For instance, Maučec and Donaj (2019) showed that while MT has improved significantly, it still lacks the cultural understanding and nuanced language ability of human translators. Likewise, Sebo et al. (2024) compared the performance of three MT systems, DeepL, Google Translate, and CUBBITT, in translating medical abstracts from French to English. Findings showed that the three MT systems performed similarly according to ROUGE, but CUBBITT scored slightly higher than the others according to the human evaluation. With that said, MT has come a long way in recent years, and there are now many different MT systems in use. The quality of such MT tools is still debatable, with an argument that they are not yet as good as human translation in terms of cultural understanding and rhetorical nuances (Han, 2023; Musaad & Al Towity, 2023). Chomsky et al. (2023) expressed concerns about the sudden popularity of Large Laanguage Models (LLMs) (e.g., ChatGPT), arguing that such AI systems differ profoundly from how humans reason and use language, emphasizing that the human mind operates with small amounts of information and is an efficient system. However, their assertions have sparked debate, and a great deal of recently published research on ChatGPT has come up with promising results.

It is possible that the increased availability of MT models has led to a decline in the demand for human translators, and, as a result, many human translators may not be practicing their skills as frequently as they once did. This could lead to a decline in the quality of human translation, as well as a loss of cultural knowledge and depth of understanding. In the area of MT versus HT, comparative evaluation of MT quality for literary texts against supervised HT is a relatively unsaturated area of research (Cercone & Murchison, 1985; Correa-Díaz, 2024; Visby, 2020; Han, 2023; Škobo & Petričević, 2023). While the accuracy and cultural understanding of MT and HT have been compared (Alowedi & Al-Ahdal, 2023; Luo, 2018), it is unclear how widespread use of MT has an impact on the skills and abilities of human

translators. The impact of MT on the quality of human translation brings well-deserved attention to this matter.

As AI reshapes the boundaries of linguistic research, its impact on literary translation demands critical attention. Hence, this study rests on a key question in this evolving discourse: can AI rival human intuition in translating poetry? It draws on an array of previous studies (Alowedi & Al-Ahdal, 2023; Farghal & Haider, 2024; Hamat, 2024; Hatzidaki & Pothos, 2008; Xiao & Muñoz, 2020; Yuhan et al., 2024) to dig into how AI and humans handle poetic features such as mood, theme, and transitivity, framed as cognitive-linguistic processing.

The delineation of this study is reflected in the following two Research Questions:

1. To what extent can automated translation tools (viz. ChatGPT, Google Translate, Apple Translator, and Microsoft Translator) translate Arabic poems?
2. Compared to certified human translators, does ChatGPT exhibit outperformance across ten distinct criteria: punctuation, layout, rhyme, mood, theme, logico-semantics, transitivity, field, tenor, and mode?

Literature Review

Machine Translation

MT has emerged as a pivotal field of study over the years, with rapid advancements largely attributed to the integration of deep learning and LLMs (Lee, 2023; Macken et al., 2020; Paterson, 2023). In the quest to understand the dynamic between MT and HT, this literature review addresses several perspectives, evaluating MT's efficiency, its implications in education, and the evolutionary trajectory of translation technologies. The efficacy of MT in translation workflows was further assessed by Macken et al. (2020), who noticed tangible speed increases using Neural MT and Statistical MT. However, the enhancements were modest, suggesting that while MT holds promise in streamlining translation processes, its current capabilities might not be as transformative as one might anticipate (Macken et al., 2020). This notion was further reinforced by Alowedi and Al-Ahdal's (2023) exploration of MT in the realm of Arabic-to-English poetry. Their findings reinforced the limitations of MT in capturing the essence of intricate subjects like poetry, which require sensitivity and deep cultural understanding, traits currently still best found in human translators (Alowedi & Al-Ahdal, 2023). To Luo (2018) such human features are “nourished by layers of deep-mind imprints throughout frames of time and space” (p. 1).

Beyond these studies, the challenges presented by idiomatic expressions were highlighted by Musaad and Al Towity (2023). Their evaluation of popular MT systems in translating idiomatic expressions from English to Arabic reiterated the cultural nuances' complex nature and the current challenges MT systems face in their accurate translation (Musaad & Al Towity, 2023). Lee (2023) further shed light on the nuanced dynamics between educators and students regarding MT's effectiveness in foreign language education, emphasizing the need for alignment and understanding between both parties (Lee, 2023).

In the rapidly evolving landscape of technology-enhanced language education, MT tools like Google Translate and ChatGPT are being integrated, willingly or not, into language learning and translation practices. This integration, however, brings with it a myriad of pedagogical, ethical, and practical implications. According to Lee (2017) Google Translate which was first launched in 2006, "supports machine translation of over 90 languages" (p. 106). Then advances in digital technology and AI have spawned other translation tools including ChatGPT which first launched at the end of 2022 and dawn of 2023.

Paterson (2023) underscored the growing prevalence of machine translation in higher education, emphasizing the need to address the ethical and pedagogical consequences arising from their widespread use. At the tertiary level, despite the evident advancements in MT accuracy, there's a palpable tension between students' frequent consultation of these digital tools and the disapproval by many instructors. However, by analyzing the intersection of MT and learning, Paterson argues that MT should be recognized as an integral component of L2 users' academic environments. Beyond evaluating perceptions per se, the study delved into the broader policy considerations of MT, highlighting gaps in academic integrity legislation and the incongruities between internationalization goals and the continued marginalization of certain English varieties. Moreover, Paterson offers ways in which MT can augment the learning process in areas such as vocabulary acquisition, writing, and metalinguistic awareness. Through such a comprehensive exploration, the study signals a pressing need for reevaluating institutional and instructor attitudes toward MT, suggesting a more inclusive and strategic integration in higher education settings (Paterson, 2023).

Lee (2023) delved into the growing adoption of MT in foreign language education, noting an uptick in students utilizing MT for academic objectives. Despite many studies endorsing the pedagogical advantages of MT, many language educators remain wary due to concerns about its quality and efficacy. After reviewing 87 MT-related studies published between 2000 and 2019, Lee found a rising trend in publications and a noticeable enhancement in MT

quality. While many studies highlighted the positive impacts of MT, particularly for writing, students exhibited mixed emotions, and there existed perceptual gaps between educators and students (Lee, 2023).

AI-Based Translation

While MT's role in education is growing, the question of translation accuracy remains pivotal, especially in specific linguistic contexts (Sahari et al, 2023). Chomsky and his co-authors describe ChatGPT in 2023 as a "lumbering statistical engine for pattern matching," though this characterization overlooks its foundation in transformer-based neural architecture. They view it as a form of 'high-tech plagiarism' that does not contribute to linguistic and cognitive development. Nevertheless, several experts challenge their perspective, pointing out that ChatGPT's linguistic capabilities often surprise even machine learning specialists. While the authors claim that the child's operating system is different from that of a machine learning program, others argue that both learn from examples, albeit with varying amounts of exposure. Additionally, ChatGPT's ability to synthesize explanations from linguists' works suggests a level of sophistication.

Complementing these academic investigations, Organ (2023) offered an empirical lens into students' attitudes towards Google Translate. By employing ethnography, Organ scrutinizes online chatroom discussions predominantly among UK secondary school students, capturing their spontaneous reflections on Google Translate usage for L2 production, especially in the context of GCSE examinations. The results depict a nuanced picture. Over the past decade, the study finds a prevailing acceptance of Google Translate for assignments among these students. However, their attitudes are far from monolithic, fluctuating based on the evolution of the MT tool itself and shifts in UK examination requisites. The study serves as a testament to the permeation of MT tools into student practices and emphasizes the need for informed pedagogical responses to this burgeoning trend (Organ, 2023).

AI and Literary Translation

Literature accentuates the growing role of MT in diverse arenas. Literary work, poetry specifically, is rich in poetic devices that require a deep understanding of the source and target languages. Despite notable progress, machine translation often struggles with the nuance and cultural depth of genres like poetry. As Blum-Kulka (2004) argues, poetry stands apart from other literary genres, such as prose and drama, and every poet holds his or her own vision of poetry. It cannot be written down from the mind to the paper on a

predetermined model. This underscores the perennial importance of human touch and cultural sensitivity that human translators bring to the table. Recognizing this gap, the present study pivots to a three-way objective analysis, juxtaposing the translational prowess of ChatGPT, other commercial MT systems, and professional human translators, with a keen focus on the challenging domain of poetry translation (Tianrui, 2024). Through this lens, we aim to shed light on the evolving contours of translation quality and inform best practices in the intersection of technology and human expertise.

Using AI in literary research is not a novelty. It appeared in as early research as that of the 1980s. For instance, Cercone and Murchison (1985) explored the intersection of AI and literary research. It examined AI's actual and potential contributions to literary research. The authors determined whether AI techniques can simulate the intellectual processes of literary research and offer advantages beyond conventional data processing methods. Their survey of relevant literature reveals that literary researchers often use computation for statistical problems but may not fully recognize the relationship between AI techniques and literary research challenges. Likewise, Škobo and Petričević (2023) examined how teachers and literary translators can adapt to the new AI and machine translation era. The aim is to find a balance between human skills and machine abilities to produce faithful and accurate translations of literary works that maintain their artistic value. The paper compares four translations of a passage from the Serbian short story *Lake Como*, done by third-year Anglistics students (whose native language is Serbian) and by Chat GPT. This thorough assessment reveals the advantages and disadvantages of AI-powered language models for literary translation and helps better understand how AI affects this field.

Correa-Díaz (2024) highlighted the dynamic relationship between poetry, technology, and culture, where AI becomes a creative collaborator in the ever-evolving landscape of digital poetics. Within the context of Latin American digital poetic projects, the study explored the extent to which one can speak of Latin American digital poetic projects oriented toward the greater objectives of scientific and social cultures (especially in terms of efforts to digitize/virtualize the world–and our humanity/consciousness), raising thought-provoking questions about the extent to which these projects align with broader scientific and social cultures, especially in terms of digitization and virtualization efforts. On a related issue, Visby (2020) dealt with the evolving landscape where AI intersects with literary translation. While the question of how AI-based machine translation handles complex literature remains intriguing, the more immediate concern for contemporary literary translators lies in the commercial viability of AI-driven translations, especially for closely related languages.

According to Visby (2020), "whatever translation software can do; it is in various ways derived from a vast number of existing human translations" (p. 30).

Muhammad Ibn Idris Al Shafi'i's Poetry

Apart from establishing one of the four schools of Islamic jurisprudence, Al-Shafi'i showed interest and poetic genius. Most of his poems provide advice, wisdom, and rules of life. Although Al-Shafi'i lived in the eighth century, his poetry spans centuries and is still relevant to today's human issues. Perhaps, he is no less famous than many of modern Arab poets such as Nizar Qabbani (Tianrui, 2024). Because translating poetry is one of the issues that involve creativity and poetic abilities (Almaktary, 2022; Tianrui, 2024), it is important to understand AI's capabilities in translating poems of Al Shafi'i which are characterized by simplicity and clarity, the poem at hand is a working example.

It is within this landscape of research that the current inquiry substantiates prior research findings on the impact of AI on the field of translation. It specifically addresses the translation of poetry. The present inquiry reflects on Alowedi and Al-Ahdal's (2023) contention that MT struggles to convey meaning beyond its database and lacks cultural understanding, making it unsuitable for translating Arabic poetry to English. Perhaps, this is because, poetry, among all literary works, is rich in poetic devices that require a deep understanding of the source and target languages. The study intends to demystify how AI helps in getting poems of Arabic literature translated compared to that of human translators. It illustrates how translation tasks reveal the boundaries of artificial language processing and underscore the cognitive sophistication of human linguistic behavior. Primarily focused on translation studies and computational linguistics, the study also intersects with psycholinguistics by examining human translation as a cognitive process. It explores how human translators process and render poetic language and compares their output with that of AI-based translation tools. It stands out due to its focus on measuring the capacity of humans and computerized translation tools in rendering Arabic poetry using a systemic functional linguistics (SFL) framework.

Method

Adopting a multifaceted approach to translating Arabic poetry in the AI age, the study compared expert human translators' performance in translating a well-known poem in Arab literature and compared the outcome with automatic translation performed by four AI tools. The study compares four

automated systems: three traditional MT tools (Google Translate, Apple Translator, Microsoft Translator) and one AI-based language model (ChatGPT), which differs from MT systems in its architecture and broader linguistic capabilities. The rationale for this comparison is that literary texts normally pose a challenge to human translators, and even more to machine translation, since they typically contain elements that are specific to the source culture.

Materials and Analysis

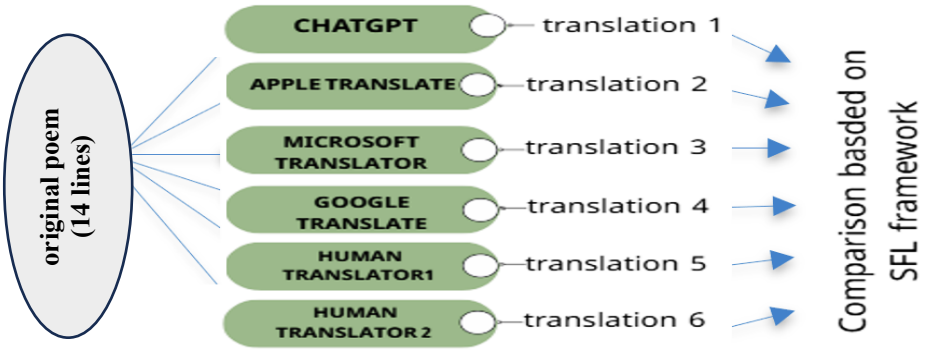
The effectiveness of ChatGPT and the other tools was assessed in comparison to human translation. For this purpose, an Arabic poem by Al-Shafi'i was selected as the basis for translation into English. The poem, which is the thrust of analysis in this study, is about traveling far away. It is one of Al-Shafi'i's best verses that is frequently cited in advisory contexts related to travel. It is concise and thematically centered on the concept of travel, and its meaning is reflected in the following lines of poetry.

The intelligent and refined find no rest in dwelling in one place,
So, leave your homeland and travel far away!
Travel and you will meet new people replacing those left behind,
And tire yourself out, because it makes life worth living!
I have seen that water stagnates when it stands still,
Yet when it runs, it is sweet and pure.
And if the lion left not its land,
it would not catch its prey
And if the arrow leaves the bow,
it would not hit its aim
And if the sun moved not across the horizon,
People the world over would have tired of the sky.

The original poem (see Appendix) is composed of 14 lines arranged in 7 rhymed couplets, following a monorhyme scheme typical of classical Arabic poetry, where each line maintains balanced meter and thematic parallelism. Its structure combines didactic tone and metaphorical imagery, reinforcing the central theme of movement and growth through rhythmic symmetry. As Figure 1 displays, the translations were generated by three traditional MT systems, an AI language model (ChatGPT), and two professional human translators who are native Arabic speakers. Given the architectural and functional differences among the translation tools, it is important to distinguish ChatGPT from the others. ChatGPT is an LLM trained in extensive multilingual corpora and designed for general-purpose language generation. In

contrast, Apple Translate, Microsoft Translator, and Google Translate are neural machine translation (NMT) systems that are optimized for direct translation tasks using neural networks. The translations were made anonymous on a form and later evaluated by the two experts in literary translation who hold a PhD degree in translation. They rated punctuation, layout, rhyme, mood, theme, transitivity, logico-semantics, field, tenor, and mode of each translation in Figure 1.

Figure 1
Analysis Framework of 14-Line Poem (from Arabic into English)



To ensure a fair and consistent assessment of translation quality, a standardized evaluation form was used. This form was structured as a rubric with ten criteria based on the SFL framework. Each criterion was scored on a 10-point Likert scale, where 1 indicated minimal fidelity and 10 represented high fidelity to the original. To maintain consistency, the form included operational definitions and examples for each criterion. Ratings were independently assigned by two experts in translation studies. Final scores were calculated by averaging the evaluators' responses, and a comparative analysis was conducted to identify performance trends across different translation systems. The ten criteria corresponded with the three meta-functions of SFL: ideational (e.g., transitivity, theme), interpersonal (e.g., mood, tenor), and textual (e.g., punctuation, layout). The framework of analysis (Figure 1) facilitated capturing both linguistic structures and contextual meaning in poetic translation.

In the vast and diverse realm of translation, understanding the efficiency and accuracy of various translation methods is paramount. The emergence of machine translation platforms, coupled with the age-old expertise of professional human translators, presents a plethora of options for those seeking translation services. However, the critical question remains: how do these methods fare across different linguistic and functional dimensions?

To provide a comprehensive answer, we delved deep into the comparative analysis of popular machine translation platforms and seasoned professional translators, evaluating their performance across ten criteria: punctuation, layout, rhyme, mood, theme, logico-semantics, transitivity, field, tenor, and mode are instrumental in determining the fidelity, fluency, and overall quality of literary translations. These criteria are grounded in the SFL framework developed by Halliday (1956, 1961), which organizes language into three meta-functions: textual, interpersonal, and ideational. This framework has been widely applied in translation studies (e.g., Wang, 2023; Hatzidaki & Pothos, 2008) to assess how meaning is constructed and conveyed across languages. The selected criteria are considered vital because they collectively capture both the structural and semantic fidelity of literary texts, especially poetry, which demands sensitivity to linguistic form, stylistic nuance, and cultural context. By operationalizing these dimensions, the study provides a systematic, linguistically informed basis for comparing human- and AI-generated translations.

The analysis of how AI and humans differ in handling poetic language is interpreted through a processing lens. The results and ensuing analysis aim to shed light on the strengths and challenges of each method, offering users informed insights to make optimal translation choices. These findings may be of interest to researchers in linguistics, translation studies, and related fields, or someone seeking translation services. The findings charted in the sections below promise to offer valuable clarity and direction. The following subsections present and analyze the responses of the evaluators for each of the 10 criteria.

Punctuation

The first criterion examined in this study is punctuation. Punctuation is an essential aspect of written language as it ensures clarity, maintains the original intent of the message, and determines the structure of written content. As shown in Figure 2, based on the evaluation of two translation experts, ChatGPT scored the highest with a total of 8 points out of 10.

ChatGPT, in comparison to other automated translation services and two professional human translators (who scored 4 and 6), was able to properly place commas and full stops based on the original Arabic text. Apple Translator (AT), with a total of 8 points, struggled with Arabic functional words. For example, the Arabic particle 'ما' [ma], which has multiple functions (e.g., negation, asking questions, relative pronoun, expressing surprise, etc.), was erroneously interpreted by AT as a question particle, when it was in fact a negation particle, which led the automated translation service to place a question mark at the end of the first part of the verse (see Table 1).

Figure 2
Comparative Scores for Punctuation Accuracy Across Translation Outputs

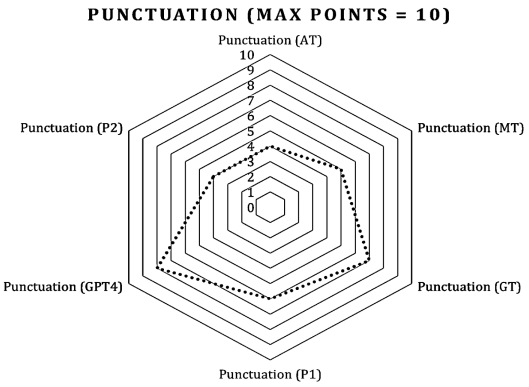


Table 1
Sample of Punctuation Use

Original Arabic Verse	ChatGPT	MT
ما في المَقامِ لِذي عَقلٍ وَذي أَدَبِ مِن راحَةٍ فَدَعِ الأَوطانَ وَاِغتَرِب	In the abode of the wise and the well-mannered, There is no comfort, so leave your homeland and travel.	What is in the place for the mind and the polite? From rest, leave the homelands and alienate

Layout

The layout in a translated poem refers to the organization and presentation of the verses and stanzas. The proper layout of a translated poem ensures that the translated content mirrors the structure of the original poem, ensuring that any inherent meaning derived from the original poem remains intact. Based on the responses of the evaluators, layout one area where all translations, excluding the outcome of Google Translate (GT), scored the maximum possible points. One of the evaluators granted GT 1 point out of 5, but no comments were provided regarding the reason for this evaluation (Figure 3).

Rhyme

Rhyme is a particularly challenging element to maintain when translating a poem. Rhyme is especially important in the translation of poetry, lyrics, and certain forms of prose since it contributes significantly to the aesthetic, rhythm, and sometimes even the meaning of the text. Based on the responses

of the evaluators, the output of ChatGPT scored the highest among all other outputs with 6 points out of 10 (see Fig. 4).

Figure 3
Evaluators' Ratings of Layout Fidelity in Translated Poem

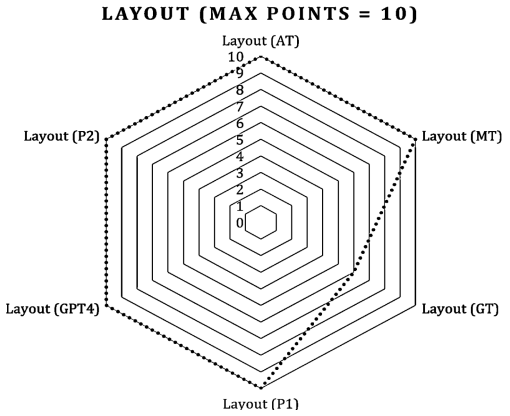
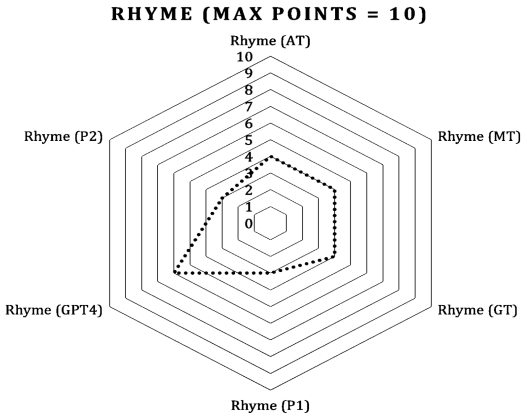


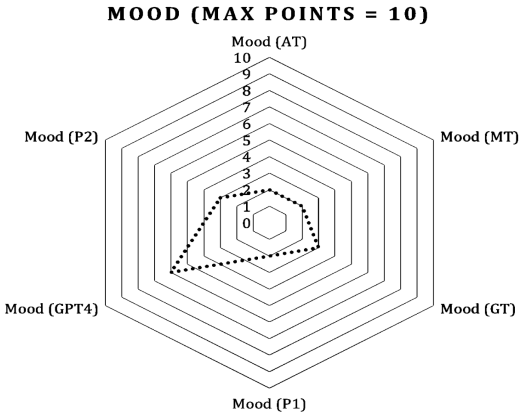
Figure 4
Evaluation of Rhyme Preservation Across Machine and Human Translations



Mood

The term 'Mood' refers to the grammatical structure of the clause, particularly focusing on the interpersonal relationships it conveys. It provides insights into the attitude of the speaker or writer, and how they present information or seek responses from others. Based on the evaluators' ratings, ChatGPT scored 6 out of 10 in preserving mood, outperforming other systems. Apple Translator and Google Translate scored 4, while Microsoft Translator received 5. These results suggest that while some systems captured basic interpersonal meanings, they struggled with subtler aspects such as tone, modality, and speaker stance (see Fig. 5).

Figure 5
Scores Reflecting the Preservation of Grammatical Mood and Interpersonal Meaning



Theme

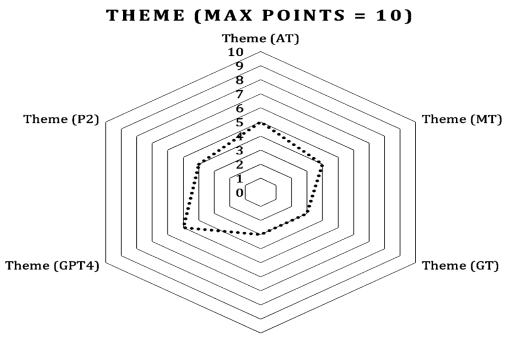
The theme of a piece of text refers to its central topic or message. In translation, maintaining the theme is crucial because it ensures that the intended message of the original text is communicated accurately in the translated version. A misinterpreted theme can result in a translation that diverges significantly in meaning or essence from its source. Scoring a 5 out of 10, Apple Translator and ChatGPT demonstrated moderate success in capturing the theme. While they might grasp the basic content accurately, there could be nuances or contextual elements that are occasionally missed, leading to slight deviations in the conveyed theme.

With a score of 4 out of 10, Microsoft Translate faced challenges similar to Apple Translate. While it provides generally coherent translations, there could be instances where the theme's finer points or underlying messages might not be captured perfectly. Similarly, achieving a score of 4, Professional 2 seems to have faced similar challenges as Microsoft Translate and Professional 1. This suggests potential complexities in the original text or conscious choices to adapt certain elements for the sake of clarity or resonance in the target language.

Google Translate achieved a score of 3. Given its extensive datasets and refined algorithms, it typically delivers contextually apt translations. However, like other machine translators, it may sometimes miss out on subtle thematic nuances. Similarly, with a score of 3 out of 10, Professional 1's performance in theme preservation aligns closely with machine translations. This might suggest that certain thematic elements in the text were inherently challenging to translate or that there were decisions made to adapt the content in a way

that felt more culturally or linguistically appropriate to the target audience, even if it shifted the theme slightly.

Figure 6
Ratings of Thematic Fidelity in Translated Versions



Logico-Semantics

In terms of logico-semantic fidelity, ChatGPT and Apple Translator retained approximately half of the original relationships, while other systems struggled to preserve coherence across clauses. This affected the clarity and interpretability of the translated verses. Achieving a score of 5 out of 10, traditional MT tools and ChatGPT managed to retain about half of the logico-semantic relationships. This implies that while many sentences might have been translated with appropriate logic and structure, there were instances where the relationships were disrupted or altered, leading to potential shifts in meaning. Scoring 4 out of 10, the output produced by Professional 2 and Apple Translator showed greater limitations in preserving logico-semantic relationships (see Fig. 7). It suggests that they might occasionally struggle to preserve the logical flow and relationships between entities in the original text, leading to translations that might be syntactically correct but semantically askew. Achieving a score of 3 out of 10, Professional 1 and GT seem to have faced challenges similar to Microsoft Translate. This score indicates that, while the translator was able to convey the general message, there might have been lapses in ensuring that the intricate logical and semantic relationships within the text were consistently maintained.

Transitivity

Transitivity was evaluated based on how well each translation preserved the relationships between verbs and their participants. ChatGPT scored highest, suggesting a stronger ability to maintain agent-action-recipient structures. In

contrast, Microsoft Translator and Professional 2 showed frequent lapses, which may have led to semantic distortions in the translated text. Scoring 5 out of 10, ChatGPT leads the machine translation category and the human professional translation category in preserving transitivity, perhaps due to its expansive training data. With a score of 3 out of 10, Apple Translator, Google Translate, and Professional 2 seem to face challenges in maintaining proper transitive relationships. Such a score suggests potential issues like omitting key participants in a sentence or misrepresenting the relationships between verbs and their subjects or objects. Scoring 2 out of 10, Microsoft Translate and Professional 2 faced occasional lapses in representing actions and their agents or recipients accurately (see Fig. 8).

Figure 7
Evaluation of Logical and Semantic Relationships Retained in Translation

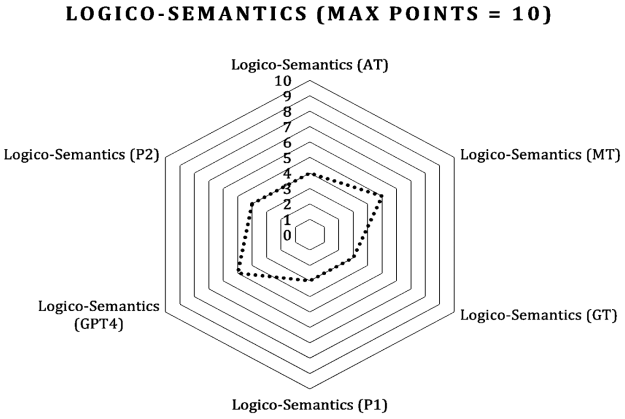
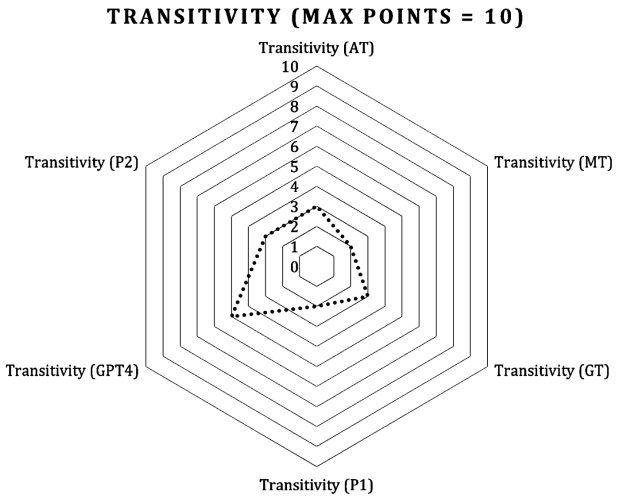


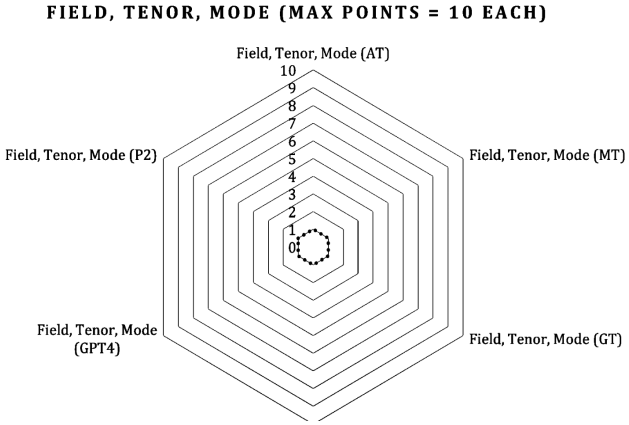
Figure 8
Scores Assessing the Accuracy of Transitive Structures in Translation



Field, Tenor, and Mode

In systemic functional linguistics, the concept of 'Field' refers to what is happening in a text, the type of activity or process described. Tenor refers to the participants in a communicative situation, capturing the roles and social relationships at play. The concept of 'Mode' relates to the degree of subjectivity or objectivity in a text and how distant or immediate the information is to the participants. These three areas proved problematic for machine translators and human translators alike, with each scoring only 1 point in each criterion (see Fig. 9).

Figure 9
Ratings for Contextual Dimensions Based on SFL



Discussion

The findings of this study reveal notable differences in how AI-based translation tools and human translators handle poetic language in the context of Arabic poetry. ChatGPT demonstrated relatively strong performance in structural and semantic coherence, outperforming both other MT and professional human translators in some linguistic dimensions. The superior performance of ChatGPT in several dimensions suggests that AI language models, while not traditional MT tools, offer enhanced flexibility and stylistic handling compared to systems optimized solely for translation. Building on the results, ChatGPT's relative strength in structural coherence and semantic fidelity suggests that LLMs may be capable of processing literary texts with a degree of sophistication, though limitations remain in capturing contextual dimensions, such as tenor and mode. This indicates a persistent gap between AI output and human interpretive depth. The comparative evaluation across the ten criteria in question, grounded in SFL, highlights the need for hybrid

approaches that integrate computational efficiency with human interpretive depth. While previous studies noted the challenges of translating idiomatic and culturally embedded expressions (Alowedi & Al-Ahdal, 2023; Musaad & Al Towity, 2023), this study demonstrates how AI tools perform when evaluated through a systemic functional linguistic lens. It quantifies those challenges across ten linguistic dimensions, offering a more granular view of where AI excels and where it falters.

The realm of translation, being an intersection of language, culture, and context, demands a nuanced understanding that transcends mere lexical substitution. In this study, MT platforms, whether it be Apple Translate, Microsoft Translate, Google Translate, or a more sophisticated AI tool (e.g., ChatGPT), show commendable competency in criteria that were deterministic and rule-based, like punctuation and layout. Yet, they stumbled when confronted with the multifaceted intricacies of mood, tenor, and mode. That is, algorithms, irrespective of their complexity, grapple with the depth and breadth of human expression embedded in cultural or context-rich dimensions. Human translation in the context of the study at hand reiterated human cognition and cultural acumen. Luo (2018) asserted that MT, thus far, cannot "uproot the occupation of human translators" (p. 1). While the machines grappled with nuanced dimensions, human translators, especially Professional 1, seemed adept at navigating these terrains. However, the variation in scores between the two professionals emphasized an essential aspect of translation: its inherent subjectivity and the pivotal role of individual expertise.

A resounding insight from this study is the multi-dimensional nature of translation. Mastery in one domain (e.g., punctuation) doesn't guarantee success in a more intricate domain (e.g., mood). To further contextualize these findings, it is important to consider the cognitive mechanisms that underpin human and AI translation processes (Yuhan et al., 2024). Hatzidaki and Pothos (2008) maintained that human translators engage in complex psycholinguistic processes that involve semantic processing, mental modeling, and working memory. These mechanisms allow them to navigate figurative language, cultural nuances, and emotional undertones inherent in poetry. Translators rely on their mental lexicon and experiential knowledge to interpret metaphors, idioms, and symbolic expressions. In contrast, AI tools simulate language processing neural architectures in transformer-based models trained on large-scale multilingual corpora. Eve advanced models (ChatGPT is a case in point) lack the depth of semantic understanding and contextual sensitivity that human cognition provides (Hamat, 2024). The divergence between AI and human translation highlights fundamental differences in language processing. Human translators exhibit dynamic, context-aware cognition, whereas AI

operates on probabilistic pattern recognition (Xiao & Muñoz, 2020). Human translators, with their innate understanding of language and context, offer a balance, excelling in some areas while underperforming in others.

The findings highlight the importance of a combined approach, utilizing both machine capabilities and human intuition, for optimal translation outcomes. As we advance in the realm of linguistics and AI, the harmonization of machine efficiency and human insight will be pivotal in bridging linguistic divides and fostering global communication. Whereas Visby (2020) warned that "literary translators may find themselves in an even more precarious position than their current one" (p. 30), a postulation endorsed by Alowedi and Al-Ahdal (2023) in that MT lacks depth of understanding of poetic discourse and cultural aspects. It struggles to convey meaning beyond its database. It may work well with specific terminologies, repetitive phrases, and everyday language use, but not with broader topics related to human emotions, cultures, and humanities (Luo, 2018). As AI continues to evolve, its role in literary translation should be viewed not as a replacement for human expertise but as a complementary tool that enhances productivity and consistency while leaving room for human creativity and judgment.

Conclusion

The contention of this comparative study is to explore AI's role in literary translation. It demystified noteworthy advances in the crux of MT, which has matured to such an extent that it competes with human intellectuality. The study signals a cautionary note that AI translation should not be taken for granted as a replacement of the human element, but a scaffolding tool to assist humans to better and speed up human work. Human translators still hold an edge in interpretive depth and cultural nuance, especially in dimensions like theme, tenor, and mode, where all systems scored poorly. Of importance, the study does not downplay the importance of AI in translation but highlights that AI is not a panacea that fits all genres and contexts. The study is essential for understanding the evolving landscape of AI in literary translation. It suggests that while AI can support certain aspects of literary translation, human expertise remains essential for preserving nuance and cultural resonance. Like with many evaluative translation studies, this study suffers from one of the problematic translation issues of this type. Hence, it provides more room for further research to dive into more rigorous and framed analytical research. Future research may benefit from integrating cognitive modeling into AI translation systems to enhance their interpretive capabilities and align them more closely with human psycholinguistic processes.

Acknowledgements

We would like to express our sincere appreciation to the human translators who contributed to this study by translating the selected poem from Arabic into English. We equally thank the two translation experts who contributed to the analysis of translations. Their careful attention to linguistic nuance and cultural context played a significant role in comparing the machine-generated and human-produced translations.

The authors are thankful to the Deanship of Graduate Studies and Scientific Research at the University of Bisha for supporting this work through the Fast-Track Research Support Program.

Disclosure Statement

The authors reported no potential conflicts of interest.

References

Almaktary, H. (2022). A poetic approach to teaching English: Personal account. *Journal of English Studies in Arabia Felix, 1*(2), 1–10. https://doi.org/10.56540/jesaf.v1i2.23

Alowedi, N. A., & Al-Ahdal, A. A. M. H. (2023). Artificial intelligence-based Arabic-to-English machine versus human translation of poetry: An analytical study of outcomes. *Journal of Namibian Studies: History Politics Culture, 33,* 1523–1538. https://doi.org/10.59670/jns.v33i.800

Blum-Kulka, S. (2004). Shifts of cohesion and coherence in translation. In L. Venuti (Ed.), The translation studies reader (2nd ed., pp. 1–15). Routledge.

Cercone, N., & Murchison, C. (1985). Integrating artificial intelligence into literary research: An invitation to discuss design specifications. *Computers and the Humanities, 19*(4), 235–243. http://www.jstor.org/stable/30200023

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). The false promise of ChatGPT. The New York Times. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

Correa-Díaz, L. (2024). Poetry and artificial intelligence. In S. Weintraub & L. Correa-Díaz (Eds.), Latin American digital poetics (pp. 37–54). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-26425-2_3

Farghal, M., & Haider, A. S. (2024). Translating classical Arabic verse: Human translation vs. AI large language models (Gemini and ChatGPT). *Cogent Social Sciences, 10*(1), Article 2410998. https://doi.org/10.1080/23311886.2024.2410998

Halliday, M. A. K. (1956). The linguistic basis of a mechanical thesaurus. Mechanical Translation, 3(3), 81–88.

Halliday, M. A. K. (1961). Categories of the theory of grammar. *Word, 17*(3), 241–292. https://doi.org/10.1080/00437956.1961.11659756

Hamat, A. (2024). The language of AI and human poetry: A comparative lexicometric study. *3L: Language, Linguistics, Literature: The Southeast Asian Journal of English Language Studies, 30*(2), Article 01. https://doi.org/10.17576/3L-2024-3002-01

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch? *Current Directions in Psychological Science, 10*(5), 164–168. https://doi.org/10.1111/1467-8721.00140

Horbach, A., Pehlke, J., & Laarmann-Quante, R. (2023). Cross-lingual content scoring in five languages using machine-translation and multilingual transformer models. *International Journal of Artificial Intelligence in Education, 34,* 795–834. https://doi.org/10.1007/s40593-023-00370-1

Lee, S. M. (2023). The effectiveness of machine translation in foreign language education: A systematic review and meta-analysis. *Computer Assisted Language Learning, 36*(1–2), 103–125. https://doi.org/10.1080/09588221.2021.1901745

Luo, X. (2018). Artificial intelligence and the crisis of translation. *Asia Pacific Translation and Intercultural Studies, 5*(1), 1-2. https://doi.org/10.1080/23306343.2018.1456440

Macken, L., Prou, D., & Tezcan, A. (2020). Quantifying the effect of machine translation in a high-quality human translation production process. *Informatics, 7*(2), Article 12. https://doi.org/10.3390/informatics7020012

Maučec, M. S., & Donaj, G. (2019). Machine translation and the evaluation of its quality. In A. Sadollah & T. S. Sinha (Eds.), Natural language processing: New approaches and recent applications (pp. 1–24). IntechOpen. https://doi.org/10.5772/intechopen.89063

Musaad, M. M. A. M., & Al Towity, A. A. (2023). Translation evaluation of three machine translation systems, with special references to idiomatic expressions. *Journal of Educational Sciences and Human Studies, 29,* 678–708.

Organ, A. (2023). Attitudes to the use of Google Translate for L2 production: Analysis of chatroom discussions among UK secondary school students. *The Language Learning Journal, 51*(3), 328–343. https://doi.org/10.1080/09571736.2021.1892171

Paterson, K. (2023). Machine translation in higher education: Perceptions, policy, and pedagogy. *TESOL Journal, 14*(4), Article e738. https://doi.org/10.1002/tesj.738

Sahari, Y., Al-Kadi, A., & Ali, J. (2023). A cross-sectional study of ChatGPT in translation: Magnitude of use, attitudes, and uncertainties. *Journal of Psycholinguistic Research, 52,* 2573–2592. https://doi.org/10.1007/s10936-023-10031-y

Sebo, P., & de Lucia, S. (2024). Performance of machine translators in translating French medical research abstracts to English: A comparative study of DeepL, Google Translate, and CUBBITT. *PloS One, 19*(2), e0297183. https://doi.org/10.1371/journal.pone.0297183

Škobo, M., & Petričević, V. (2023). Navigating the challenges and opportunities of literary translation in the age of AI: Striking a balance between human expertise and machine power. *Društvene I Humanističke Studij, 2*(23), 317–336. https://doi.org/10.51558/2490-3647.2023.8.2.317

Tianrui, M. (2024). Poetry and translation as bridges: Exploring Nizar Qabbani's translation theory. *Cogent Arts & Humanities, 11*(1), Article 2298551. https://doi.org/10.1080/23311983.2023.2298551

Visby, M. (2020). The future relationship of literary translation and AI: Reflections from CEATL president. Counterpoint, 4, 28–31.

Wang, Y. (2023). Artificial intelligence technologies in college English translation teaching. *Journal of Psycholinguistic Research, 52,* 1525–1544. https://doi.org/10.1007/s10936-023-09960-5

Xiao, K., & Muñoz, R. (2020). Cognitive translation studies: Models and methods at the cutting edge. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 19,* 1–24. https://doi.org/10.52034/lanstts.v19i0.559

Yuhan, N., Zhylin, M., Antonyuk-Kyrychenko, S., Popova, N., & Harbar, M. (2024). Cognitive aspects of translation: The latest research in psycholinguistics and cognitive science. *Forum for Linguistic Studies, 6*(4), Article 6367. https://doi.org/10.59400/fls.v6i4.6367

Appendix

Original Poem	**Apple Translate**	**Microsoft Translator**	**Google Translate**	**ChatGPT**
ما في المَقامِ لِذي عَقلٍ وَذي أَدَب *مِن راحَةٍ فَدَعِ الأَوطانَ وَاِغتَرِب* *سافِر تَجِد عِوَضاً عَمَّن تُفارِقُه* *وَانصَب فَإِنَّ لَذيذَ العَيشِ في النَصَب* *إِنّي رَأَيتُ وُقوفَ الماءِ يُفسِدُه* *إِن ساحَ طابَ وَإِن لَم يَجرِ لَم يَطِب* *وَالأُسدُ لَولا فِراقُ الأَرضِ ما اِفتَرَسَت* *وَالسَهمُ لَولا فِراقُ القَوسِ لَم يُصِب* *وَالشَمسُ لَو وَقَفَت في الفُلكِ دائِمَةً* *لَمَلَّها الناسُ مِن عُجمٍ وَمِن عَرَب* *وَالتِبرُ كَالتُربِ مُلقىً في أَماكِنِه* *وَالعودُ في أَرضِهِ نَوعٌ مِنَ الحَطَب* *فَإِن تَغَرَّبَ هَذا عَزَّ مُطلَبُه* *وَإِن تَغَرَّبَ ذاكَ عَزَّ كَالذَهَب*	What is in the place for the mind and the polite? From rest, leave the homelands and alienate Travel and find instead of those you leave. And the deliciousness of living in the monument I saw the water park ruining it. The yard is good, and if it doesn't run, it doesn't work. And the lion, had it not been for the separation of the earth, Mafterst And the arrow had it not been for the separation of the bow would not have been hurt.	There is no place for those who have a mind and literature From the comfort of the homelands and immigrate Travel and you will find compensation for those you leave Pour the delicious living in the monument I saw the standing of the water spoil it If it is good, it is good, and if it does not, it is not good And the lion, if it wasn't for the parting of the earth, wouldn't have preyed on And the arrow had it not been for the parting of the bow was not injured	What is in the position of a mind and a literature From comfort, leave the homelands and migrate travel to find a replacement for those you leave and erect, for it is pleasant to live in a monument I saw standing water spoil If he is good, and if he does not run, he is not good. And the lion, had it not been for the parting of the land, it would not have preyed And the arrow, had it not been for the parting of the bow,	What is in staying for one with reason and culture, Leave your homelands for a bit of comfort. Travel, you will find a substitute for what you left behind. And settle down, for living is delightful in stillness. I have seen standing water spoil it. Whether the pond is clear or not, it does not matter. The lion would not have feared if it did not depart from the earth. And the arrow would not have hit if it did not depart from the bow. And the sun, even if it stood still in the sky forever,

	And if the sun stands in the ark permanently I'm inspired by people from Ajam and Arabs And the righteousness is like the soil lying in its places And the oud in his land is a kind of firewood If you go, this is the height of his demand. And if you go away, glory is like gold.	And if the sun stood in the orbit permanently People from Ajam and Arabs filled it And the dust is like soil lying in its places And the oud in its land is a kind of wood If he goes abroad, this is his greatest demand And if that goes away, glory is like gold	would not have hit And if the sun were to stand permanently in the orbit People of non-Arabs and Arabs filled it And dust is like dust lying in its places And the oud in his land is a kind of firewood If this exalted his request And if that goes away, that is like gold.	People would get tired of it, both Arabs and non-Arabs. And the ashes are like soil falling in their places. And the oud in its land is a type of fuel. So if this departs, its goal becomes glorious. And if that departs, it becomes glorious like gold.