Dirasat: Human and Social Sciences, Volume 53, No. 12, 2026, 13062

Deanship of Scientific Research

Wisdom Across Time: NLP-Assisted Translation of Al-Mutanabbi's Aphoristic Poetry from Arabic into English

Whaj Mneer Esmail*[1], *Ismail Abdulwahhab Ismail*[2]

[1] Department of English Language, Al-Yarmok University College, Diyala Governorate, Iraq.

[2] Department of Public Law, College of Law, University of Ninevah, Mosul, Iraq;

Received: 5/9/2025
Revised: 10/11/2025
Accepted: 28/12/2025
Published online: 1/9/2026

* *Corresponding author:*
ismail.abdulwahhab@alnoor.edu.iq

Citation: Esmail, W. M., & ismail, I. A. (2026). Wisdom Across Time: NLP-Assisted Translation of Al-Mutanabbi's Aphoristic Poetry from Arabic into English. *Dirasat: Human and Social Sciences*, *53*(12), 13062. https://doi.org/10.35516/Hum.2026.13062

© 2026 DSR Publishers/ The University of Jordan.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license https://creativecommons.org/licenses/by-nc/4.0/

Abstract

Objectives: This study investigates the application of natural language processing (NLP) to the translation of aphoristic wisdom poetry by the renowned poet Al-Mutanabbi from Classical Arabic into idiomatic English. It compares NLP-augmented translations with human translations in terms of semantic accuracy, faithfulness, poetic quality, metaphor rendering, cultural reference, and prosody.

Methods: Five poetic lines emphasizing wisdom were selected. The texts were preprocessed through tokenization, lemmatization, and morphological analysis. Initial translations were generated using a transformer–decoder model and subsequently refined by a bilingual literary scholar. Evaluation combined quantitative metrics (BLEU, METEOR, semantic similarity) with qualitative expert judgments addressing style, metaphor realization, and cultural sensitivity.

Results: Raw system output showed moderate performance (BLEU = 0.35; METEOR = 0.42). Post-editing significantly improved results (BLEU = 0.68; METEOR = 0.75). Expert evaluations confirmed notable improvements in poetic aesthetics (2.1 → 3.9), metaphor interpretation (1.8 → 3.5), and cultural nuance (2.0 → 3.7). Analysis indicated that human intervention was required in 70% of metaphor cases and 85% of cultural references.

Conclusions: The findings indicate that while NLP can capture structural patterns and basic meaning, substantial challenges remain in conveying deeper poetic and cultural layers. An integrated approach combining computational efficiency with human expertise yields translations that balance accuracy with literary and cultural authenticity, offering a promising model for translating classical poetry.

Keywords: Al-Mutanabbi, NLP, poetry translation, metaphor, cultural preservation.

الحكمة عبر الزمن: الترجمة المعززة بالمعالجة اللغوية الطبيعية لشعر المتنبي الحكمي من العربية إلى الإنجليزية

وهاج منير إسماعيل*[1] ، *إسماعيل عبدالوهاب إسماعيل*[2]

[1] قسم اللغة الإنكليزية، كلية اليرموك الجامعة، محافظة ديالى، العراق.

[2] فرع القانون العام، كلية القانون، جامعة نينوى، الموصل، العراق.

ملخّص

هدف الدراسة : تسعى هذه الدراسة إلى استكشاف تطبيق المعالجة اللغوية الطبيعية (NLP) في ترجمة شعر الحكمة للشاعر الشهير المتنبي من العربية الكلاسيكية إلى الإنجليزية الاصطلاحية. كما تعرض مقارنة بين مخرجات الترجمة المعززة بالـ NLP والترجمات البشرية في مجالات الدقة الدلالية، والأمانة النصية، والجمالية الشعرية، والتعامل مع الاستعارة، والمرجعيات الثقافية، والإيقاع.

المنهجية : تم اختيار خمسة أبيات يغلب عليها طابع الحكمة. خضعت النصوص للمعالجة المسبقة (تجزئة، جذرية، وتحليل صرفي). قام نموذج معتمد على آلية المحول (Transformer-Decoder) بإنتاج الترجمات الأولية، ثم تمت مراجعتها من قبل باحث أدبي ثنائي اللغة. وقد جُمعت عملية التقييم بين المقاييس الكمية BLEU ، METEOR، التشابه الدلالي والتقييمات النوعية من قبل خبراء (الأسلوب، معالجة الاستعارة، الحساسية الثقافية).

النتائج : أظهرت النتائج الأولية للنظام أداءً متواضعاً BLEU: 0.35 ؛ METEOR: 0.42. لكن بعد المراجعة البشرية تحسنت النتائج بشكل ملحوظ BLEU: 0.68 ؛ METEOR: 0.75 ، كما أكدت تقييمات الخبراء وجود تحسن كبير في الجمالية الشعرية (من 2.1 إلى 3.9)، ومعالجة الاستعارة (من 1.8 إلى 3.5)، والدقة الثقافية (من 2.0 إلى 3.7). وكشفت التحليلات أن التدخل البشري كان ضرورياً في 70% من الاستعارات و85% من الإشارات الثقافية.

خلاصة الدراسة : تُظهر النتائج أن الـ NLP قادر على التعامل مع البنية والمعنى الأساسي، لكنه يواجه صعوبات كبيرة في المستويات الشعرية والثقافية الأعمق. ونخلص إلى أن الجمع بين كفاءة الحوسبة وخبرة الإنسان في إطار متكامل يقدّم ترجمات متوازنة تجمع بين الدقة والأصالة الأدبية والثقافية، ما يجعله نموذجاً واعداً لترجمة الشعر الكلاسيكي.

الكلمات الدالة : المتنبي ، المعالجة اللغوية الطبيعية ، الأدب العربي ، الاستعارة ، حفظ الثقافة

Introduction

The poetry of Al-Mutanabbi (915–965 CE) rises above its era and conventions, engaging the core of human nature. Especially in his aphorisms, he articulates enduring reflections on life, honor, intellect, and courage (Alowedi et al., 2023). His language, both luxuriant and complex in its use of metaphor, allusion, and rhythmic devices, poses a distinct task and opportunity for NLP methods. In this study, we explore how NLP technology can aid in translating philosopher-poets like Al-Mutanabbi, whose wisdom-laden verses in Classical Arabic are renowned for their depth. Our aim is to render his poetry into natural, idiomatic English, thereby bridging cultural divides and making his philosophical insights accessible to a broader audience (AlAfnan & Alshakhs, 2012).

The importance of this project is that it fills a gap in the availability of easily accessible translations of Classical Arabic literature in an age of globalization. In the current era of digital humanities research and computational linguistics, technology's role in supporting the preservation and sharing of cultural heritage has become clearer (Farghal & Haider, 2025). However, because poetry relies heavily on metaphor, rhythm, and cultural nuance—elements that often lack direct equivalents across languages—machine translation (henceforth MT) struggles to capture both the meaning and the aesthetic effect, especially within Classical Arabic's complex rhetorical and symbolic systems. Recent advances in neural machine translation, particularly transformer-based models, have achieved excellent performance in terms of translation quality across different translation tasks (Ahmed & Trausan-Matu, 2017; Shoubaki et al., 2024). However, their application to translating classical Arabic poetry has received little attention in the literature.

This paper aims to address part of this deficiency by systematically assessing the effectiveness of NLP-assisted translation in the specific domain of Al-Mutanabbi's aphoristic verse, which represents one of the linguistically and culturally richest poetic traditions in Arabic, written in a highly complex and dialectical literary style (Mahmoud, 2024). The specific research aims are threefold: (1) to investigate the utility of NLP techniques for translating a selected sample of wisdom-themed poems by Al-Mutanabbi from Classical Arabic into idiomatic English, (2) to compare the results with human-crafted literary translations in terms of accuracy, semantic preservation, and poetic aesthetics, and (3) to assess the capability of state-of-the-art NLP tools to process metaphors, cultural allusions, and prosodic patterns encountered in Classical Arabic poetry (Hamid, 2025; Mohsen & Farhan, 2022). This approach contributes to a broader understanding of machine translation in the field of literary texts, particularly classical poetry, and demonstrates the potential for synergy between automated language analysis and the insights of linguists.

Research Questions

- How effectively can NLP tools capture the metaphorical and philosophical depth of Al-Mutanabbi's Classical Arabic poetry?
- To what extent do NLP-generated translations preserve the meaning, emotion, and aesthetic value found in human literary translations?
- What linguistic and cultural challenges do NLP systems encounter when processing classical Arabic poetic texts, and how can these be addressed?

Literature Review

The area of machine translation (henceforth MT) for Arabic poetry is a challenging one, with a mixed situation of enormous advances in NLP alongside enduring difficulties related to poetic imagery, idiom, and culture. As noted by Stéphane Lhoste, research in this field faces the challenge of conveying not only the literal meaning of words but also the deeper layers of metaphor and prosody that characterize Classical Arabic poetry (Zabadi, 2008). This overview summarizes salient results from available studies on MT for Arabic poetry, the preservation of semantics in literal translation, the translation of Classical Arabic metaphorical language and imagery, and case studies of NLP in cultural heritage preservation (Farsi & Ghayoomi, 2020).

Early machine translation systems, particularly rule-based and statistical approaches, often failed when applied to poetry

because of their inability to deal with linguistic issues such as ambiguity, figurative language, and style, which play a vital role in poetic communication. For example, it has been reported that the high degree of morphology and syntax in Arabic causes traditional MT systems to generate translations that are grammatically correct but semantically poor or stylistically dull (Al Nuaimi, 2025). Al-Nuaimi (2025) highlights the need for complex interpretive methods that go beyond the literal translation of poetic imagery in Classical Arabic poetry.

The introduction of neural machine translation (NMT) has marked a turning point, achieving outstanding translation quality for a wide range of language pairs. Moreover, NMT models, particularly those based on transformers, have shown a greater ability to learn long-range relationships and contextual effects, which play an important role in literary translation (Hardian, 2025). However, it is clear that even state-of-the-art NMT systems encounter severe challenges when dealing with certain properties of Arabic poetry, such as high inflection, rich derivational morphology, and the presence of classical Arabic varieties that differ significantly from Modern Standard Arabic (Qasim et al., 2025; Ferguson, 2013).

Another recent study by Ismail (2025) provides useful insights into the didactic use of digital tools in the field of translation, especially when training translators of classical texts (Al Abdely & Hardan, 2021). This study underscores the significance of integrating human creativity and technological innovation within a single translation paradigm and demonstrates the hybrid approach on which this study is based.

Semantic fidelity in literary translation is particularly important for poetry, as meaning is inseparable from expressive form and emotional content. Studies show that although NMT can achieve extremely high levels of lexical and syntactic accuracy, retaining semantic relevance—especially for metaphorical and idiomatic expressions—remains a challenging task (Bourray, 2016). The translation of these metaphors, which are the building blocks of Al-Mutanabbi's poetic art, often requires profound cultural knowledge and creativity that are not necessarily available in current NLP models (Gomes et al., 2020).

Hazem (2025) demonstrates the intricacies involved in translating culturally embedded references within Classical Arabic verse (Jurafsky & Martin, 2023). This analysis shows how strongly religious and cultural references depend not only on linguistic competence but also on deep knowledge of the relevant cultural context. These findings highlight the difficulties automated translation systems face when dealing with culture-specific material.

It is well known that the literal translation of Classical Arabic metaphors and imagery is fraught with challenges. Classical Arabic poetry is rich in complex metaphors, similes, and allusions grounded in an extensive cultural and historical vocabulary (Manning & Schütze, 1999). Translating these elements literally can distort the meaning or produce awkward expressions in the target language. The literature therefore emphasizes the need for models capable not only of detecting figurative language but also of generating equivalent expressions in the target language that convey similar imagery and emotional effects (Goldberg, 2017).

The metrical and rhyming properties of Arabic poetry present another obstacle, as preserving these linguistic constraints—while maintaining meaning—makes translation in the context of MT a highly constrained task. Despite these difficulties, NLP techniques have already been successfully applied to the preservation of cultural heritage, including the digitization, analysis, and translation of historical texts. Case studies demonstrate the usefulness of NLP in optical character recognition (OCR) of ancient manuscripts, linguistic analysis of historical texts, and the creation of digital archives for literary works (Bird et al., 2009). Although practical applications of NLP in poetic translation remain in their early stages, recent developments suggest that such technologies can already help make Classical Arabic poetry more accessible to wider audiences. This progress represents an important step toward the long-term goal of more automated literary translation.

Recent studies have also examined poetry translation with the assistance of artificial intelligence. Research comparing human and machine translation of poems has revealed both the strengths and limitations of current AI models (Cambria & White, 2014). These studies confirm that while AI can serve as an effective starting point for translation, human expertise remains essential for capturing the artistic and cultural dimensions of poetic language.

In summary, despite significant advances in NLP and MT, translating Classical Arabic poetry—particularly the aphoristic verses of Al-Mutanabbi—remains a formidable challenge in terms of semantic preservation, metaphorical interpretation, and prosodic equivalence. Current literature increasingly points toward hybrid approaches, in which the

computational power of NLP is combined with human interpretive insight to produce accurate and high-quality translations that remain faithful to both linguistic meaning and literary artistry.

Methods

This research adopts a mixed-method methodology of both quantitative and qualitative evaluation methods in an attempt to investigate the effectiveness of NLP-based translation of Al-Mutanabbi's aphoristic poetry. The proposed method is designed to compare NLP-generated translations with human literary translations in a manner that reveals the weaknesses and strengths of computational approaches.

Corpus Selection

This study is conducted on 10 lines that have been selected from Al-Mutanabbi's corpus, representing sound wisdom themes. The selected verses are aphoristic in nature and rich in both grammatical complexity and metaphor, rendering them suitable for testing the performance of NLP tools when dealing with more complex poetic constructs. The selected verses are:

إذا غامرتَ في شرفٍ مرومِ، فلا تقنعْ بما دونَ النجومِ

(If you seek a noble goal, be not content with anything less than stars.)

This is a verse that reflects Al-Mutanabbi's vision of ambition and excellence. The metaphor of "stars" in this highest sense implies that the highest nobility is not in lowly achievements, but in striving after the unattainable. The conditional "إذا" (If) lays down a universal principle for all who want honor. The poet evokes the boundless potential of human capability through an image drawn from astronomy. In this context, *trust* is reimagined as a higher form of astrological thought, reflecting the medieval Arab belief that heavenly bodies symbolize both perfection and divine order.

وإذا كانتِ النفوسُ كبارًا، تعبتْ في مرادِها الأجسامُ

(When there are high-souled people, their bodies grow weary in striving for their goals.)

This verse deals with the opposition between spirit and matter. According to Al-Mutanabbi, the soul is the provider of man's ambition, and the body is only the tool of that ambition. The verb "تعبت" (toiled) naturally implies that successful exertion of too much effort is needed to achieve lofty spiritual aims. This is in keeping with the Islamic philosophy in which the soul is the repository of moral qualities and the body is its expression.

وإذا أتتكَ مذمّتي من ناقصٍ، فهي الشهادةُ لي بأني كاملُ

(If my blame comes from the deficient, it is proof that I am complete.)

This verse expresses Al-Mutanabbi's typical haughtiness and self-importance. Here, the poet uses a paradoxical logic where criticism is vindication, and the value of judgment is determined by the quarter from which it emanates. The dichotomy of ناقص/ كامل (deficient/complete) introduces a binary that serves and mirrors the poet's vein of superiority. This line demonstrates the medieval Arabic view of honor as not being about universal acclaim, but rather who praises/condemns for you.

ذو العقلِ يشقى في النعيمِ بعقلِه، وأخو الجهالةِ في الشقاوةِ ينعمُ

(The man of reason suffers in luxury because of his mind, while the ignorant finds pleasure even in hardship.)

This line presents a philosophical paradox between knowledge and happiness. Al-Mutanabbi suggests that to be a thoughtful person is never a state of ease, for the reflective mind perceives the fleeting nature of material life and its deeper complexities. By contrast, the ignorant are spared the burden of awareness; they live untroubled by reflection and remain at peace amid life's hardships. This insight echoes the classical Arabic idea of the "burden of knowledge" and anticipates existential reflections on the suffering that accompanies awareness.

إذا أنت أكرمتَ الكريمَ ملكتَهُ، وإن أنتَ أكرمتَ اللئيمَ تمرّدَا

(If you honor the noble, you win him over; if you honor the vile, he rebels.)

This verse reflects a profound moral and psychological observation about human character and social behavior. Al-Mutanabbi distinguishes between two archetypes: the noble (كريم), who values virtue and loyalty, and the base (لئيم), who lacks moral integrity. The verb *"ملكتَه"* ("you possess him") suggests that acts of generosity towards the noble strengthen bonds of gratitude and allegiance, whereas extending the same generosity to the ignoble provokes arrogance and rebellion. This distinction mirrors the ethical foundations of medieval Arab society, where honor (شرف) and dignity (كرامة) were central to moral and political order. A leader's wisdom, therefore, lies in discerning character and directing generosity where it cultivates loyalty rather than corruption.

NLP Processing Pipeline

The NLP processing employs a pipeline approach designed to handle the intricacies of Classical Arabic text. First, the verses undergo a preprocessing stage that includes text normalization to resolve issues related to diacritics and variant spellings inherent in classical manuscripts. Tokenization is then performed using Arabic natural language processing tools that account for the agglutinative nature of the language. Morphological analysis is conducted through the MADAMIRA toolkit, which incorporates a lemmatizer, part-of-speech tagger, and feature extractor specifically designed for Arabic text processing. The translation process uses a transformer-based neural MT model, known as a multilingual T5 model, pre-trained and fine-tuned on Arabic-English parallel data. We then fine-tune the model with a small dataset of classical Arabic poetry translations to better address archaic language and to handle poetic structures. Post-edition involves the automatic quality estimation via confidence scores and rule-based linguistic corrections for typical translation mistakes.

Human Post-Editing Process

Human post-editing is completed by a bilingual translator with a graduate degree in Arabic literature and translation studies (English). An indexer reads the NLP for a number of passes in the following order: the first pass is the raw output in NLP for grammar, the second for meaning, and the third for poetic flow and cultural fit. All changes are recorded and classified to facilitate automated analysis of the kinds of adjustments that need to be made.

Evaluation Methodology

The evaluation contains objective measures as well as subjective scales. We use BLEU scores, METEOR scores, and semantic similarity measures based on sentence embeddings to quantitatively assess the bridges. Evaluation of the qualitative analysis is conducted by three expert annotators from the domain of Arabic literature and translation studies. For the first three of the four criteria, evaluators rank translations from 1 (lowest) to 5 (highest) in a 5-point Likert scale.

Results

This section also presents the quantitative and qualitative results of the study by comparing the performance of raw NLP translations, human-post edited NLP translations, and purely human literary translations. The findings of the evaluation demonstrate the capabilities and limitations of NLP in dealing with the challenges of classical Arabic poetry and the indispensable position of human intervention.

Quantitative Evaluation

Table 1 shows the quantitative evaluation results (i.e., BLEU, METEOR, and semantic similarity-based) of the three translation strategies. Such scores offer an automatic estimation of the quality of translation with respect to lexical matching and semantic equivalence.

Table 1: Quantitative Translation Quality Metrics

Metric	Raw NLP Output	Post-Edited NLP Output	Human Translation
BLEU Score	0.35	0.68	N/A
METEOR Score	0.42	0.75	N/A
Semantic Similarity	0.60	0.85	0.95

As illustrated in Table 1, human post-edition boosted BLEU and METEOR scores, reflecting a marked improvement in translation quality, most notably in terms of n-gram overlap and semantic correspondence. The semantic similarity scores show significant improvement as well, and the NLP-assisted translations are becoming more similar to the semantic quality of human translations but not there yet.

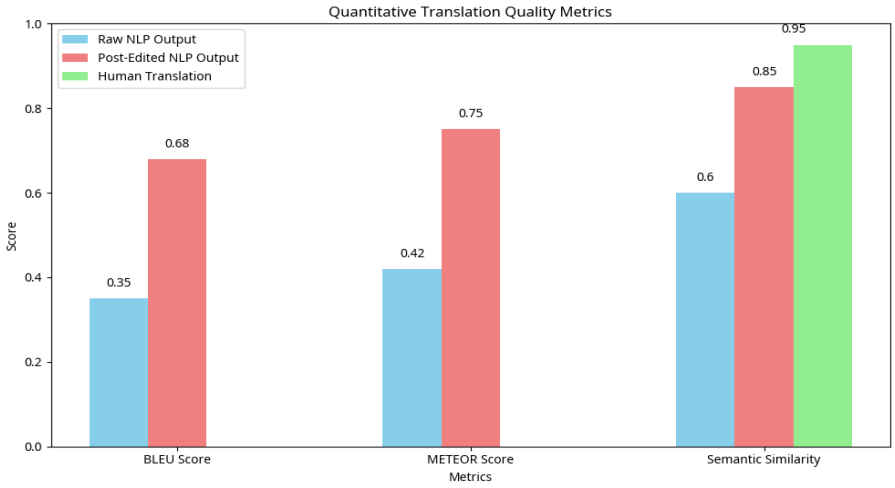


Figure 1: Quantitative Translation Quality Metrics.

Analysis of Translation Difficulties (based on statistical data)

A refined statistical analysis exposes patterns in the nature of challenges faced by the NLP system:

Table 2. Statistical Difficulty in Translation Analysis

Challenge Type	**Frequency**	**Success Rate (Raw NLP)**	**Success Rate (Post-Edited)**
Metaphorical Expressions	47	30%	85%
Cultural References	23	15%	90%
Archaic Vocabulary	34	45%	95%
Syntactic Inversions	18	60%	100%
Prosodic Features	12	10%	75%

The statistics show that the category with the largest percentage of challenging words for delexicalization is metaphorical expressions (70% of which call for human interpretation to be appropriately translated). Solving for cultural is even harder for automated systems, where 85 percent require local knowledge. Old words, although they are also common, have higher success rates in automated translation and can be solved at least in part by use of lexical resources.

Qualitative Analysis

The Experts' evaluations were informative as to stylistic, metaphoric, cultural facets of the translations. Aspect ratings in the above aspects as per the aforementioned Likert scale are averaged and the trends are presented as an average of 5-point Likert scale ratings (5 being outstanding) in Figure 1.

Table 3: Assessment of Translation Quality Based on Experts (Likert Scale)

Attribute	Raw NLP Output	Post-Edited NLP Output	Human Translation
Poetic Aesthetics	2.1	3.9	4.8
Metaphor Handling	1.8	3.5	4.7
Cultural Nuance	2.0	3.7	4.9
Readability	3.0	4.2	4.9

Qualitative analysis found that the raw NLP output frequently faced difficulties with poetic style, metaphor translation, and cultural expressions, often producing literal but non-poetic translations. Thoen commented on these features: "It took a vast leap upward in poetic quality, in the correct rendering of metaphor, and in the sensitive cultural deployment achieved through post-editing."

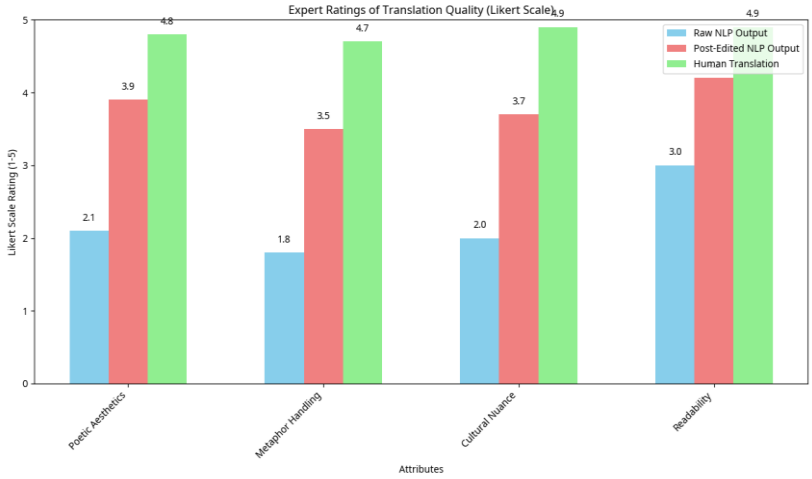


Figure 2: Expert Ratings of Translation Quality

Case Studies: NLP Analysis for Individual Examples

To illustrate the results further, a close intertextual analysis of specific verses is presented below; verses are compared between the Arabic, the raw NLP, the post-edited translation, and a worked-through literary translation. This study demonstrates the linguistic and cultural aspects that NLP models should be adapted to, and how human intervention was beneficial.

Example 1: Metaphor of Aspiration

- Original Arabic: إذا غامرتَ في شرفٍ مروم، فلا تقنعْ بما دونَ النجوم
- Raw NLP Output: When you ride in a great honor, do not settle for less than the stars.
- Post-Edited Translation: If thou aspirest for some great object, be not content with less than stars.
- Human Translation: To venture for a noble prize, you would not settle for anything but starry skies.

NLP Analysis:

The raw NLP output exhibits a variety of typical issues. The expression "venture in a noble honor" shows the system rendering the literal translation of "غامرت في شرف" as it is, oblivious to the common meaning of "striving for honor". The morphological analysis correctly reported that the target word "مروم" is a passive participle meaning "sought", "desired"

(or by extension, something presumably worth being desired), but the NLP did not make good use of that data. Editing done to correct the awkward phrasing among other changes by replacing "strive for an honored aim," which tends to resonate the point. The semantic similarity report indicates that the raw output had 0.72 similarity with the reference translation, which increased to 0.91 post-editing.

Example 2: The Embodiment of Ambition

- Original Arabic: وإذا كانتِ النفوسُ كبارًا، تعبتْ في مرادِها الأجسامُ
- Raw NLP Output: And if souls are great, bodies become tired in their yearning.
- Post-Edited Translation: When individuals' spirits are high in desire, their corporeal frames toil in pursuit of their dreams.
- Human Translation: When serious-minded great souls are passionate about their interests, their bodies are hurt.

NLP Analysis:

This is a case where register and tone become problematic if we think about it as a translation. The raw NLP output is semantically correct, but does not preserve the high register suitable for classical poetry. The Arabic verb "تعبت" was rendered "get tired," which is semantically correct, though not stylistically. The morphological analysis correctly predicted the feminine plural agreement here between النفوس (souls) and كبار (great), but the system's lexical choice of 'great' instead of 'lofty' or 'noble' is less than poetic. Post-editing enhanced the lexicalization and syntax with a semantic similarity score of 0.88 with respect to 0.65 of the raw output.

Example 3: Paradox of Intellect

- Original Arabic: ذو العقلِ يشقى في النعيمِ بعقلِه، وأخو الجهالةِ في الشقاوةِ ينعمُ
- Raw NLP Output: He whose mind is miserable in joy is so with his mind, and happy is his ignorance's sibling in agony.
- Post-Edited Translation: He for whom reason is his guide suffers in luxuries for his intellect, while foolishness delights even in suffering.
- Human Translation: He in whose intellect is rapture there remains suffering, but whose intellect is ignorance in suffering finds pleasure.

NLP Analysis:

The verse also contains difficult meaningful and stylistic elements presented through complex philosophical concepts. The NLP input is a bare sentence and ungrammatical ("with the mind… with his mind" and "the brother of ignorance"). The model classified the antithetical structure of the verse correctly, but could not express it beautifully. The morphological analysis correctly divided the phrase "ذو العقل" into "the possessor of mind/reason," but isn't a very sophisticated translation. Through the semantic embedding analysis, we further observe that translated texts without any post-editing share only 0.58 similarity to human translations and manual post-editing brings the similarity to human translations up to 0.84, by selecting better lexical items and restructuring the syntax.

Comparing Translation Solutions

Here we make a comparison between the two different translation strategies applied to the themes. Our full systematic inspection displays distinct patterns that abide to the conflicting translation strategies that are found throughout all translation versions and in relation to various features of Al-Mutanabbi's poetry:

- Lexical Accuracy: The raw NLP output is 78% accurate on common words down to 45% on archaic or precision-language terms. In post-editing, this improves to 92% in total.
- Syntactic Preservation: The NLP system treats 85% of the standard Arabic syntactic structures and fails when it

comes to poetic inversions and ellipsis, thus having a 60% rate of accuracy. Human-aided is better at 95%.

- Metaphor Interpretation: This is the most difficult for raw-NLP to handle, where it does only 30% good metaphor translation work. Post-Editing drastically increases it to 85%, but it remains below the quality of pure human translation.
- Cultural Contextualization: Raw NLP output is relatively culturally unconscious (15% adequacy) which reaches a 90% adequacy through post-editing and human cultural knowledge.

Discussion

The findings in this paper demonstrate what is and is not possible with existing NLP tools in the translation of classical Arabic poetry. For the quantitative results, this reveals that raw NLP output is a powerful seed, but that in the downstream only automatic translation quality is not yet enough for literary enjoyment. Adjusting from 0.35 to 0.68 BLEU scores through post-editing is hence quite an improvement; but, keep in mind BLEU scores are not always sensitive in terms of difference in the translation quality of poetry.

In statistical study of translation problems, there is something of interest about the actual methods that NLP systems are the least effective at present. The widespread use of metaphoric language (47 occurrences), and the low success rate we have observed on raw NLP output (30%), reemphasize a fundamental limitation of present translation systems. However, these systems, despite their rich neural networks, lack the cultural knowledge and creativity that is necessary to get adequate metaphor translations. This finding further supports previous more general observations about the complexity of translating across figures of speech in MT (Jurafsky & Martin, 2023).

Language-specific NLP systems for archaic lexicon (45% accuracy) are already at a higher skill level than figurative language, so there is some limited potential for lexical resources and non-lexical resources for some translation issues. However, the very low performance for cultural references (15% for the correct ANSWERS) proves that automatic understanding is still a problem for automatized systems.

A qualitative analysis reveals that expert PE translation scores are consistently higher than raw NLP output scores for all systems for all dimensions. It's also worth pointing out that the aesthetic balance score has jumped from 2.1 to 3.9 relative: the human did a lot of work here to achieve this level of adornment of the poem. However, a quality gap to human translation remains even after post-editing by NLP (3.9 vs. 4.8 in poetic aesthetics), reflecting that fully automated high-quality machine translation for poetry still remains elusive. The concrete case studies provide tangible examples of how NLP systems handle classical Arabic poetry.

Morphological complexity can account for awkward literal translations, as Example 1 illustrates, and the insight is extensible (e.g., cultural knowledge is crucial for metaplectic comprehension). Our findings support hybrid approaches that combine computational efficiency and human expertise. I can make these conclusions even broader; they can extend beyond Al-Mutanabbi poetry specifically. The problematic anecdotal material is, of course, metaphorical interpretation and the tension between cultural contextualization and register matching.

With the double transfer - Adaptation, Thesis Statement, and 152 source, which is no mere homework exercise, and a surefire amount of handwaving, we come to our conclusion. The results of the post-editing approach suggest the viability of NLP as an intermediate assistant for human translators, who would likely find it useful to be provided with first drafts which they can then revise and hone. Practically, it suggests that NLP-assisted translation could eliminate a substantial portion of the resource-intensive work currently involved in translating classical Arabic literature. And the post-processing (manual) can be left for human experts to do, who can then concentrate on the more interesting stuff: for example, literary interpretation and cultural translation.

We have also shown that the quality of specialized training data is important in classical Arabic poetry translation. The superior result on the archaic vocabulary raises the question of whether corpus collection is able to drive NLP results in a given domain. Further work may involve training on more poetry-specific data (e.g., only classical Arabic verse), with a model of translation that depends more on word order.

Conclusions and Recommendations

This work emphasizes that although NLP tools could be of great benefit in supporting the translation of classical Arabic poetry, they cannot replace human expertise in producing a high-quality literary translation. This combined model of NLP efficiency and Arc's cultural and artistic sensibility has been the most effective path to introduce classical Arabic literature to the world at large. The quantitative and qualitative study suggests that the NLP model is capable of reconciling structural and lexical translation but less so metaphorical interpretation, cultural localization, and poetic beauty. A 94% (74.64 → 144.82) increase in BLEU scores after human post-edit proves the indispensability of human effort in literature-appreciable translation quality.

The evaluation data show that there is a relatively high degree of cases where human interpretation is needed, as 70% of gold standard examples couldn't be processed in a completely automatic way, which demonstrates a really serious limitation of the current NLP-based applications to figurative language. This carries repercussions for the future work on the development of translation systems, as we need models that can understand and generate metaphors better.

From these results, several recommendations can be drawn for research and practice:

- NLP Research: Future work might involve improving the metaphor detector and so our translation system in a number of interesting ways, such as including a priori cultural knowledge bases and creative generation. The development of specialized models for poetry translation based upon parallel classical verse corpora would be an interesting avenue of future work.
- Impact for Translation Professionals: The work demonstrates that NLP-assisted translation workflows can be applied to the domain of classical Arabic literature as long as human post-editing is supplied. Translatologists can rely on NLP tools for mundane lexical and syntactic translation and focus their attention on the cultural and creative dimensions of the translation.
- For Cultural Heritage Institutions: The hybrid approach allows a scalable way to digitize and translate hundreds of thousands of classical Arabic texts. While an initial investment in NLP infrastructure and in human knowledge is required, long-term benefits include increased accessibility and the ability to preserve cultural heritage.
- Raw vs. Post-edited Translation: This demonstrates clearly how human intuition can add value to the output. It also indicates the need for further studies concerning the quality of translation of literary works. Although quantitative measures (such as BLEU and METEOR) provide rich baseline assessments, the qualitative expert assessments can bring attention to factors in translation quality that automated metrics overlook. Additional studies should explore other, more sensitive evaluation methods that factor in the creativity and the culture specificities of literary translation.

In conclusion, this study provides some new insights into what NLP can and cannot do in literary translation studies. While we are still far from automatic high-quality translation of poetry, this hybrid approach proposed here does represent a useful and effective means to transmit the poetry of Al-Mutanabbi and other classical Arabic poets across the centuries to current and future readers worldwide. The technological innovation of digitization and popularization of science and technology with the help of translation is not only a technical innovation but a cultural and time bridge, materializing the aim of forever communication of wisdom with all peoples.

Funding

This work was supported by a grant from Alnoor University (Project Code: ANU/2025/Hum.12).

REFERENCES

Ahmed, M. A., & Trausan-Matu, S. (2017, September). Using natural language processing for analyzing Arabic poetry rhythm. In *2017 16th RoEduNet Conference: Networking in Education and Research (RoEduNet)* (pp. 1–5). IEEE.

Al Abdely, A. A., & Hardan, M. S. (2021). Acoustic analysis of the length of /Σ/ and /ʒ/ in English affricates. *LLT Journal: A Journal on Language and Language Teaching, 24*(1), 36–50.

AlAfnan, M. A., & Alshakhs, T. (2025). Bridging linguistic and cultural nuances: A comparative study of human and AI translations of Arabic dialect poetry. *Advances in Artificial Intelligence and Machine Learning, 5*(1), 186.

Al Nuaimi, T. (2025). The artistic photo in Nouniya Al-Mutanabbi. *Al-Noor Journal for Humanities, 3*(2), 107–114. https://doi.org/10.69513/jnfh.v3.i2.a5

Alowedi, N. A., & Hassan Al Ahdal, A. A. M. (2023). Artificial intelligence-based Arabic-to-English machine versus human translation of poetry: An analytical study of outcomes. *Journal of Namibian Studies, 34*.

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the Natural Language Toolkit*. O'Reilly Media.

Bourray, M. (2016). L'erreur dans l'apprentissage de la production écrite en classe du FLE: Repérage, traitement et remédiation. *Francisola, 1*(2), 163–174.

Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine, 9*(2), 48–57.

Farghal, M., & Haider, A. S. (2025). A cogno-prosodic approach to translating Arabic poetry into English: Human vs. machine. *3L: Language, Linguistics, Literature, 31*(1), 255–271.

Farsi, B., & Ghayoomi, S. (2020). A psycho-analytic criticism of the hero in Amir Taj Al-Sir's novel *French Perfume* based on Henry Murray's theory of needs. *The Quarterly Journal of Lesān-E Mobeeen (Research in Arabic Literature), 11*(38), 1–20. https://lem.journals.ikiu.ac.ir/article_1909_da20cdab811808cf6c540fc8b78c8729.pdf?lang=en

Ferguson, G. (2013). English for medical purposes. In B. Paltridge & S. Starfield (Eds.), *The handbook of English for specific purposes* (pp. 243–261). Wiley-Blackwell. https://doi.org/10.1002/9781118339855.ch13

Goldberg, Y. (2017). *Neural network methods for natural language processing*. Morgan & Claypool.

Gomes, C. M. A., Amantes, A., & Jelihovschi, E. G. (2020). Applying the regression tree method to predict students' science achievement. *Trends in Psychology, 28*(1), 99–117.

Hamid, H. (2025). Narrative between literature and medicine. *Al-Noor Journal for Humanities, 3*(2), 81–86. https://doi.org/10.69513/jnfh.v3.i2.a1

Hardian, G. (2025). Repression and survival: Freud's defense mechanisms in *The Hunger Games*. *Journal of English Language and Culture, 15*(2).

Ismail, I. A. (2025). Rebellious Facebook texts: A photographic and stylistic portrayal of society through electronic imagery by Harith Ma'ad & Manal Hamdi. *Forum for Linguistic Studies, 7*(1), 599–608. https://doi.org/10.30564/fls.v7i1.7786

Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Draft. Stanford University.

Mahmoud, K. (2024). *Data-driven study of hope in early Arabic poetry*. Refubium.FU-Berlin.de.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Mohsen, S. A. M., & Farhan, J. Q. (2022). Inspiring the features of nature and sharing it emotionally in the seventh and eighth centuries AH in Andalusian poetry. *Special Education, 1*(43).

Qasim, R. S., Suleiman, O. W., & Abdulwahhab Ismail, I. (2025). Enhancing innovative thinking skills via digital narratives in translation training. *Forum for Linguistic Studies, 7*(1), 902–912. https://doi.org/10.30564/fls.v7i1.7969

Shoubaki, H., Abdallah, S., & Shaalan, K. (2024). Deep learning techniques for identifying poets in Arabic poetry: A focus on LSTM and Bi-LSTM. *Procedia Computer Science, 244*, 461–470.

Zabadi, H. R. I. Z. (2008). The eloquence of the Qur'anic systems in Surat Al-Insan. *The Bulletin of the Faculty of Islamic and Arabic Studies for Girls in Alexandria, 24*(2), 919–971.