



RESEARCH ARTICLE

Section: *Literature, Linguistics & Criticism*

Translating ambiguity: A comparative evaluation of AI Chatbots in rendering Arabic full homonyms into English

Abdullah Al Momani¹, Mohammad Al Taher¹ & Sultan Alali^{2*}¹Department of English Language & Translation, Faculty of Arts & Humanities, Applied Science Private University, Jordan²Department of English Language, Faculty of English Language and Translation, University of Jeddah, Jeddah, Saudi Arabia*Correspondence: sialali@uj.edu.sa

ABSTRACT

Homonymy is one of the most persistent challenges in translation, particularly for AI-based systems. This study examines the abilities of three prominent AI chatbots, ChatGPT 5, DeepSeek (V3.1), and Gemini (2.5 Flash), in the English translation of Arabic homonyms. The dataset comprised Qur'anic verses and classical poetry. The homonyms were first collected, then analyzed for contextual meaning, and then translated by the chatbots. The data were presented without diacritics to replicate the conditions of authentic Arabic text. After the items were translated, the data were assessed using the TAUS Adequacy/Fluency model to determine the extent to which the systems preserved meaning while maintaining linguistic naturalness. The results showed that while all chatbots often achieved high fluency, their adequacy slightly differed across text types and homonym categories. Gemini outperformed the other models overall, followed by ChatGPT and DeepSeek. The Qur'anic translation performed best due to the high prevalence of such texts in online corpora. However, the performance declined noticeably in the poetic and classical domains. While large language models (LLMs) represent a significant improvement over traditional machine translation (MT) systems in handling lexical ambiguity, they remain unable to reliably disambiguate complex Arabic homonyms without contextual or cultural grounding. The study concludes that AI systems require substantially more training, particularly in Arabic-language contexts. The study therefore recommended further research on homonym translation, focusing on different contexts as well as the rendition of homonyms outside of pairings.

KEYWORDS: AI chatbots, Arabic homonyms, figurative language, lexical ambiguity, machine translation

Research Journal in Advanced Humanities

Volume 7, Issue 4, 2026

ISSN: 2708-5945 (Print)

ISSN: 2708-5953 (Online)

ARTICLE HISTORY

Submitted: 29 July 2026

Accepted: 14 September 2026

Published: 14 October 2026

HOW TO CITE

Al-Momani, A., Al Taher, M., & Alali, S. (2026). Translating ambiguity: A comparative evaluation of AI Chatbots in rendering Arabic full homonyms into English. *Research Journal in Advanced Humanities*, 7(4). <https://doi.org/10.58256/tmmtvm27>



Published in Nairobi, Kenya by Royallite Global, an imprint of Royallite Publishers Limited

© 2026 The Author(s). This is an open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

The translation process begins with understanding the source text and deciphering its meanings. This indicates that lexical ambiguity can increase the challenge of translation (Al-Darabee et al., 2025; Li et al., 2024). An aspect of language that can increase lexical complexity is homonymy. Research carried out in classrooms showed that even human translators, especially students, face certain obstacles when handling homonyms without guidance (Rabadi, 2015). In this study, full homonymy refers to lexical forms that are identical in form but convey distinct, unrelated meanings. It is distinguished from polysemy, in which meanings are semantically related, and from homophony, in which words share pronunciation but differ in spelling or meaning. Context-dependent rhetorical wordplay is treated separately and is not classified as full homonymy unless the lexical item meets these criteria. Related corpus-assisted discourse research by Haider (2019a, 2019b, 2019c) illustrates how contextual interpretation and corpus-based evidence can inform the analysis of meaning in Arabic discourse, although these studies did not investigate machine translation or homonym disambiguation.

Arabic translation requires attention to semantic, pragmatic, and cultural factors that shape context-dependent meanings (Aburahme, 2021; Ammari & Al-Ahmad, 2023; Fraihat, 2024; Ja'afreh, 2023; Khatatbah & Ashour, 2024; Meqdad et al., 2023; Naib, 2025; Al-Ezzi & Al-Qudah, 2024; Al-Qudah, 2022; Alqudah, 2025). The morphological richness of the Arabic language, the absence of diacritics in most written texts, and the context-dependent change of semantics can lead to full homonyms, which are words that are similar in form but differ in meaning, becoming problematic. Furthermore, the triangular root grammar in the Arabic language yields many homonymous forms. Meanings with a mutual root may be extracted from the phonological differences in vowelization, affixation, and syntactic positions. Due to the deletion of diacritics in most written Arabic forms, the thing that evidences meaning becomes context. Translators should thus extract the meaning from contextual clues, a process where learners can improve, yet machines typically display incapability. Haider and Hussein (2022) also investigated intralingual Arabic subtitling, drawing attention to how linguistic choices convey culturally sensitive meanings. In a study of English–Arabic subtitle corpora, Haider, Alzghoul, and Hamadan (2023) examined the role of parallel audiovisual data in translating culturally specific language.

As a result, conventional machine translators, such as Google Translate and SYSTRAN, are unable to resolve this ambiguity, producing results that are decipherable but meaningless (Aldawsari, 2023). Recent work, however, has shown that larger language models (LLMs), including GPT and Gemini, are more accurate at handling vague expressions (Farghal & Haider, 2024; Obeidat et al., 2024; Tannous & Haider, 2025); nevertheless, some failures still persist. Artificial Intelligence chatbots are now being used in learning, customer care, and real-time translation, both in the workplace and in social interactions (Al-Salman & Haider, 2024; Haider, 2026a, 2026b; Haider et al., 2025). Many people around the world now use these multi-purpose systems daily, rather than dedicated translation engines such as Google Translate or DeepL. They vary in how well they handle linguistic structures, which determines their effectiveness. Though the grammatical basis of Arabic, including root-and-pattern morphology, derivational systems, and the codification of syntax, is widely known in the literature (Holes, 2004; Ryding, 2005; Versteegh, 1997; Wright, 1896/2011), these rules still need to be stated and placed in the proper context. Bringing AI levels up to cope with these types of structures and homonyms, in particular, would not only improve translation quality but also enhance the effectiveness of AI chatbots as a cross-cultural communication tool. The work is thus at the border of computational linguistics and applied translation, specifically focusing on Arabic homonyms. More recently, Haider, Obeidat, et al. (2025) compared cloud-based and traditional subtitling workflows, providing a related perspective on how translation technologies are evaluated in practice. Research on English subtitling of Egyptian film idioms further illustrates the difficulty of transferring context-dependent figurative meanings across languages (Haider & Shohaibar, 2024).

Recent studies highlight the role of AI and linguistic context in processing Arabic meaning and ambiguity (Hammouri, 2025; Jaradat et al., 2026; Alkarazoun & Rizi, 2025; Rizi & Hamdan, 2023). To a great extent, especially with the emergence of larger language models (LLMs), AI-based translation systems have made significant progress in fluency, contextual reasoning, and cross-linguistic generalization. In spite of these developments, they still exhibit systematic deficiencies in their treatment of Arabic homonyms - lexical items that have the same form but differ in meaning. Such homonyms can significantly alter meaning, particularly in sensitive areas such as education, books, and religious writings (Al-Abbas & Haider, 2020; Al-Jarf, 2025; Bin

Hariz, 2023).

Recent studies establish that although larger language models (LLMs) like GPT and Gemini are better at vague or underspecified expressions, the improved linguistic performance is not consistently extended to the case of dense homonymy in Arabic. Partly, this is because most AI chatbots are trained on large-scale, general-purpose corpora that do not adequately reflect Arabic's morphological and semantic diversity. This, in turn, creates a gap between users' expectations of correct AI-mediated translation and the ambiguity of homonymous phrases, leading to mistranslations or semantic loss. Without special effort to support Arabic homonym disambiguation, AI chatbots tend to lead to misinterpretation and disrupt cross-linguistic communication.

Although research on machine translation and word-sense disambiguation is increasing, little has been done to examine AI chatbots' performance with Arabic homonyms. The available literature is inclined to discuss homonyms indirectly. For example, comparative analyses of Google Translate and SYSTRAN (Aldawsari, 2023) or larger studies of the LLM's performance on lexical ambiguity (Aldawsari, 2025) do not examine homonyms as a phenomenon that warrants individual examination. The problems with homonyms in Arabic are mentioned in human-centered research (Rabadi, 2015), though there is no evidence available on how AI chatbots can replicate, reduce, or increase these problems.

Technical advancements in Arabic word-sense disambiguation have been made over the last 6 years (e.g., Wael et al., 2024), but the models have not been systematically deployed in translation pipelines or tested in chatbots. Moreover, research on the translation of the Qur'an highlights the theological and interpretative implications of poor handling of homonyms Bin Hariz (2023), but no research has evaluated the ability of modern AI chatbots to prevent these traps. The failure to address this gap and the lack of specific empirical studies indicate a definite gap between Arabic homonyms, AI chatbot behavior, and Arabic-English translation. The need has become a necessity to fill this gap to improve the process of translation and the effective application of AI-mediated language tools in practice.

This study aims to answer the following questions:

- How accurately and fluently do ChatGPT, DeepSeek, and Gemini translate the selected Arabic homonyms into English?
- What descriptive differences are observed among the three systems in translating the selected Arabic homonyms?

2. Literature Review

2.1. Theoretical background

Klepousniotou (2002) asserted that lexical ambiguity can be attributed to two main types, homonymy and polysemy. Furthermore, Ali (2022) outlined the theoretical distinctions between homonymy and polysemy and other lexical relationships, and discussed the common errors in Arabic-to-English translation that may be explained by homonymy. Types of errors are multiple-root homonyms, homonyms extended metaphorically, homonyms which are particles of speech, cross-part-of-speech homonyms, and confusion with polysemy. The piece is a compilation of classroom-based suggestions of how to instruct disambiguation strategies and strengthen language-structure awareness.

Yang and Zhang (2024) asserted that language ambiguity, including homonymy, especially in the form of homographs, can pose challenges to machine translation. Additionally, Khudjaeva (2023) argued that the translation of homonyms can be challenging for both human and machine translators; however, disambiguation can aid with this issue. Furthermore, Garcia (2021) determined that the context of a term can reduce the ambiguity of homonyms and aid people, to some extent, and machines in understanding their intended meanings. Furthermore, Hauer and Kondrak (2020) indicated that homonyms exhibit distinct behaviors in translation. Ali (2020) asserted that homonyms can complicate translation, as it is essential to properly determine their intended sense to avoid mistranslation, as awareness of homonymy is necessary in maintaining the message of the ST.

2.2. Empirical studies

Wael et al. (2024) tested transformer-based tools (sentence transformers based on Siamese networks and SetFit few-shot learning) on an Arabic word sense disambiguation (WSD) dataset, named SALMA. Their plan scored

71 on F1 and was the second-best, which is much better than the traditional WSD baselines. The article identifies Arabic-specific causes of ambiguity, such as homophones, homographs, and the lack of diacritics. It shows that contrastive training and few-shot classification can be used to improve predictive performance on small datasets. Khasawneh and Alsharif (2025) investigated MT's capabilities in translating complex items by comparing how four MT systems translated AV political idioms against human reference. Gemini's strategies were closest to the human translator, while Matecat achieved the best translation quality. Microsoft Bing performed the worst, though all systems produced similar translation lengths.

Al-Jarf (2025) examined the performance of AI translation systems (e.g., ChatGPT, Google Translate, QuillBot, Smartling, DeepL) on polysemous words in full-text Arabic academic articles in the field of education. Despite the natural, stylistically correct sound of the outputs, the study recorded systematic semantic mistakes in which polysemous items were interpreted in the wrong domain (e.g., *قيد* as validity rather than honesty; *ميكحتل* as peer review rather than arbitration; *قلاسر* as thesis rather than a message). The findings indicate that AI tends to mistranslate terms by employing literal equivalents rather than conceptual ones. However, it was shown that these issues can be reduced through post-editing and the incorporation of standard technical terms. The article warns against using AI to provide full-text translations and offers realistic insight into how the outputs could be improved in specialized academic fields.

In addition, Abdulmughni & Alrefaee (2025) adopted a mixed-methods approach, using computational methods to analyze the interaction between the translator and the AI, comparative analysis of human and AI-generated translations, and interviews to evaluate AI translation. The study concluded that while basic linguistic processing has been enhanced using NMT based on transformers, it has yet to tackle issues with respect to cultural sensitivity, style, figurative language, and background knowledge in the case of Arabic literature. Abdulmughni & Alrefaee (2025) have suggested some improvements such as cultural knowledge modules and improvement of Arabic morphology processing and a hybrid human-AI workflow that preserves the authors' voices and cultural authenticity. Lastly, they reached the conclusion that AI should not replace professional literary translation, but rather be used as a tool to assist them.

Moreover, Aldawsari (2023) assessed the intelligibility and accuracy of Google Translate and SYSTRAN output in a controlled test of sentences containing homonyms, heteronyms, and polysemes across various domains. The mean scores were below 40, and intelligibility exceeded accuracy. The performance of Google Translate was usually better than that of SYSTRAN, but the systems failed most on heteronyms, in part because Arabic diacritics and discretization errors were poorly modeled by the MT systems. The article highlights the importance of more context modeling and Arabic-specific design to reduce errors caused by ambiguity. Even though the scope of research on machine translation and word-sense disambiguation has grown in recent years, little work has examined the performance of modern AI chatbots on Arabic homonyms.

In addition, Aldawsari (2025) compared how GPT, Gemini, Google Translate, and SYSTRAN translated 16 Arabic sentences that were designed to be lexically ambiguous. The experiment achieves huge improvements in the precision and readability of the bigger language models (LLMs). The experiments reveal that the contextual thinking of LLMs is better than a classical MT pipeline, but it has a few limitations. The article assumes that LLMs represent a step-change in disambiguating Arabic ambiguity and suggests that the next step is to consider such aspects as robustness, timeliness in the design, and the size of the evaluation.

Moreover, Ramadhan (2023) compared the use of polysemy in English and Arabic in various parts of speech. Both languages have polysemy, particularly in nouns, verbs, and prepositions. However, English has the tendency to generalize polysemy by use of adjectives and phrasal verbs, but Arabic mostly does this by adverbial realization. The analysis indicates that all these areas, such as lexicography, pedagogy, and the choice of language selection, are subject to the effects of these systematic differences.

Ammari & Al-Ahmad (2023) conducted corpus-based research that explores four English synonyms in the Corpus of Contemporary American English (COCA) and the Cambridge Learner Corpus (CLC) to examine frequency and collocation variations, as well as grammatical and genre distribution. The paper emphasizes that synonyms might seem interchangeable, but their application is influenced by contextual, collocational, and grammatical factors that make it difficult for English as a Foreign Language (EFL) learners to learn.

As for the Noble Qur'an, Ali & Nordin (2014) focused on Qur'anic polysemy (*al-wujūh wa-al-naza*) by examining the process of selecting the senses of polysemous lexemes by English translators. The authors

conclude that literal measures are not always successful in conveying context-dependent meanings and urge translators to employ context-sensitive processes, such as discourse analysis, exegetical support, and robust semantic tools, to resolve competing senses effectively.

Bin Hariz (2023) compared the English translations of Yusuf Ali and Ahmed Ali, highlighting the problems that arise when translators rely on literal equivalence to translate synonyms and homonyms. It states that a lack of exposure to exegetical tradition and poor contextualization result in semantic drift and loss of subtlety. The author proposes a pragmatic methodology grounded in local context and the contextual application of Qur'anic exegesis, cross-lingual sound/etymological associations, and the judicious use of semantics to mitigate lexical disparity in sacred text translation.

To sum up, as noted in the reviewed articles, lexical ambiguity remains an underlying issue in Arabic-English translation, and partial benefits may be achieved with the assistance of AI-powered solutions. Despite specific issues with homonyms and polysemy when using more traditional methods of linguistics, including Google Translate and SYSTRAN, the more powerful LLM has potential. However, it has not proven reliable under high-stakes conditions. In addition, experiments involving the classification of classroom experiments, the translation of the Qur'an, and cognitive linguistic theory show that computational power alone is insufficient to resolve homonyms; cultural, contextual, and cognitive sensitivity are also required. The interaction among the complexity of language, the level of technological development, and real-world practice shapes the core of the present study, which aims to close a gap in the research by systematically and intentionally evaluating the extent to which Arabic homonyms are recognized and translated by AI chatbots. In this study, we aim to illuminate the issue of semantic ambiguity in AI-based translators and thereby improve their accuracy in real-world intercultural communication.

3. Methodology

3.1. Data Analysis Approaches

Homonymy poses a challenge for translation, let alone for AI, since chatbots may not be able to comprehend the ST's purpose and context. The present study follows a mixed-methods approach. In qualitative terms, translations are analyzed according to one of the quality assessment models. In quantitative terms, percentages are provided to compare the chatbots' ability to convey the meanings of the homonyms in question. To gauge the success of these most popular AI chatbots in conveying the meaning of Arabic homonyms, the researchers took several steps. The qualitative analysis examined how each chatbot interpreted and rendered the contextual meanings of the homonyms. These findings informed the adequacy and fluency ratings, which were then summarized quantitatively using percentages. Thus, the qualitative findings formed the basis for the quantitative comparison among the three systems.

Adequacy and fluency were rated separately using the five-level scale. Adequacy assessed whether the contextually intended meaning of each homonym was accurately conveyed, while fluency assessed the grammaticality, naturalness, and readability of the English output. The homonym was the scoring unit; therefore, when a sentence contained multiple homonyms, each was evaluated separately. The item-level ratings were then used to calculate the frequencies and percentages reported in the summary tables.

3.2. Corpus Collection and Compilation

A search was conducted for common full homonyms in accordance with the rules of Arabic rhetoric. Focus was placed on those found in the Noble Qur'an and poetry, which were considered more reliable. Both enjoy unique features, mainly in terms of rhythm, cultural mediation, multiplicity of interpretation, prosody, intertextuality, and frequent use of rhetorical devices. Their choice of diction, structure, and style leads to a special relation between form and meaning, making it so hard for the translator to replicate. On the one hand, the Noble Qur'an is the sacred word of God, whose miraculous language is recognized as inimitable, even in its original language. On the other hand, poetry has always been considered the apex of eloquence in Arabia, bearing in mind much of the above as well as its metrical nature. Based on Arabic rhetoric, homonymous expressions may take the form of any of the Arabic parts of speech, namely nouns, verbs, and particles.

It is noteworthy that prose was excluded because it does not enjoy the same qualities of the Noble Qur'an and poetry, and it may include examples of nonsense pun expressions which may be made up only for

the sake of fun. In addition, some cases from the Noble Qur'an or poetry are left out since they include the same words found in the selected examples, so redundancy is avoided. Another reason associated with Qur'anic verses concerns the difference between scholars on the interpretation of their intended meaning. As a result, 23 examples were approved within 18 sentences.

All items were reviewed as full homonyms, i.e., identical lexical forms with distinct, unrelated meanings, and cases not meeting this definition were classified according to their appropriate linguistic category. The evaluation focused primarily on whether each system identified and conveyed the contextually intended meaning of the ambiguous form. Where wordplay involved the simultaneous activation of more than one meaning, preservation of this rhetorical effect was also considered.

The 23 examples were selected based on three criteria: their classification as full homonyms in Arabic rhetorical references, the presence of clearly distinguishable contextual meanings, and the availability of sufficient context to determine the intended meaning. Repeated examples and cases involving considerable interpretive disagreement were excluded. Examples were identified from Qur'anic and classical Arabic rhetorical and linguistic sources. Items were included when they contained identical lexical forms with distinct, contextually identifiable meanings. Repeated, disputed, or unclear cases were excluded.

These collected utterances were examined with respect to the meanings of homonyms, taking into account their purpose and the contexts in which they were used. To this end, exegesis books were consulted for the Noble Qur'an verses, whereas references to rhetoric were reviewed. The intended contextual meaning of each homonym, established through recognized exegetical and rhetorical sources, served as the reference standard for assessing translation adequacy.

It is worth noting that the sample is not intended to be fully representative of Arabic homonymy. Rather, it represents a focused selection of full homonyms from Qur'anic and classical poetic texts and therefore reflects the specific linguistic and rhetorical characteristics of these genres.

3.3. Investigated AI Tools

The three systems were accessed through their free consumer web interfaces in September 2025: ChatGPT (GPT-5, OpenAI), DeepSeek (V3.1), and Gemini 2.5 Flash (Google). No API was used, and all systems were accessed through free-tier accounts using their default interface settings. Browsing/search, memory, and other optional tools were not intentionally enabled during testing.

Each item was submitted individually and once to each chatbot in a fresh conversation using the same neutral zero-shot prompt, "Translate the following into English." No prior chat history or conversational context was provided. No regeneration or selective retention was used, and only the first response was recorded. The single-run design may be affected by stochastic variation; therefore, the findings represent performance under this specific testing protocol rather than stable differences between the systems.

3.4. Used Framework

The Translation Automation User Society (TAUS) model for translation evaluation was adopted to check the quality of the choices made. It is often used in machine translation and is centered on the roles of Adequacy and Fluency, emphasizing simplicity and speed. It is primarily aimed at achieving high customer satisfaction. It also seeks to determine whether post-editing (PE) of machine translation is more effective than immediate human translation, and to explore errors of different types (Görög, 2014). The first criterion concerns semantic correctness, and the second, linguistic quality (O'Brien, 2012). Each one is rated on a 1-5 scale. On the one hand, the more meaning is preserved, the closer to 5 the Adequacy score is. On the other hand, the more native-like the rendition, the closer it is to 5 on Fluency. The TAUS Dynamic Quality Framework (DQF) is known for its systematic, straightforward quality evaluation approach. Their flexibility assists in assessing diverse content and automated translations (Bououden, 2025). It is worth noting that all the researchers are native speakers of Arabic and English language specialists, which enables them to judge the renditions according to both criteria. The translations were evaluated by 3 researchers, all native Arabic speakers and English language specialists. Each output was assessed separately for adequacy and fluency using the stated five-level scale, with adequacy judgments guided by the contextual meanings established through recognized exegetical and rhetorical sources. Any disagreements were resolved through discussion and consensus.

4. Analysis and Findings

This section presents the analysis of 7 sample sentences that contain 10 of the 23 cited homonymy examples. This is achieved by first explaining the context of each example along with the intended meaning of the homonyms. After which, the translations provided by ChatGPT, DeepSeek, and Gemini are presented and analyzed based on the TAUS Adequacy/Fluency model above. The Adequacy/Fluency check will be rated as follows: 1/5: unacceptable, 2/5: below average; 3/5 satisfactory, 4/5 above average; and 5/5: outstanding.

Given the small sample size, SPSS or similar statistical software may not be necessary. Instead, the general percentages in the tables below are believed sufficient to demonstrate the distribution of the translation ratings achieved by the different chatbots. The qualitative analysis of the examples is presented below.

The seven examples discussed in detail represent the qualitative analysis, while the complete set of 23 homonyms was included in the evaluation.

Example 1

“(30:55). ”عاس ري غ او ثبل ام نومرجملا مسقي ةعاسلا موقت مويو“	
Wa-yawma taqūmu as-sāʿatu yuqsimu al-mujrimūna mā labithū ghayra sāʿah.	
1A. (ChatGPT)	And on the Day when the Hour is established, the criminals will swear they remained but an hour .
1B. (DeepSeek)	And on the Day the Hour appears, the criminals will swear they had remained but an hour .
1C. (Gemini)	And on the Day the Hour will be established, the criminals will swear that they had remained (in the world) for nothing but an hour .

Example 1 contains two homonyms of the term “ةعاس” which are used to express explicitly different meanings. The term can be translated literally as “hour”; however, the context shows that this meaning was only intended in one of its uses. In this verse, the first use of the word, which is presented with a definite article, refers to the Day of Judgment, and the latter reflects the general use of a time measurement unit of 60 minutes. In addition, the former use of the term is derived from the representation of the Day of Judgment as the final hour of worldly life. However, the latter is used to show how short they consider their lifetimes to be (Al-Zamakhshari, 1987). Both homonyms are nouns, one definite and one indefinite, which are related to time yet their meanings vastly differ.

All three systems use the term “hour” for the renditions of the two homonyms. However, the terms are distinguished through two elements that mark generality and specificity. The first use of the term, which refers to the Day of Judgment, is preceded by the definite article “the” while the second, which refers to time, is preceded by the indefinite article “an”. Moreover, the first use is capitalized, which, according to English grammar conventions, displays the term as a proper noun. This aligns with the use of the term “the Hour” to refer to the Day of Judgment in both Islam and Christianity. This reflection of actual language use can be attributed to the fact that large language models (LLMs) are composed of text from all over the web. This indicates that the systems adequately distinguished between the two terms and properly presented the meanings that suited their contexts. Each of the three is thus rated “outstanding” according to the above model.

However, it must be noted that upon searching for these renditions, it turned out that 1B is identical to that of Sahih International which is one of the prominent translations of the Noble Qur’an. This may have negative ethical implications in relation to plagiarism.

The adequacy and fluency ratings for Example 1 are presented in Table 1.

Table 1. Adequacy and Fluency Ratings for Example 1

AI System	Adequacy	Fluency
ChatGPT	Outstanding	Outstanding
DeepSeek	Outstanding	Outstanding
Gemini	Outstanding	Outstanding

As shown in Table 1, all three systems achieved outstanding adequacy and fluency, accurately conveying contextual meaning while producing natural, grammatically correct English.

Example 2

	”هَبْ يَنْتَرُمْ أَيْ إِلَا مَهْل تَلْقَ إِي” (5:117) Mā qultu lahum illā mā amartanī bihi
2A. (ChatGPT)	I said to them <u>only what</u> You commanded me to say.
2B. (DeepSeek)	I did <u>not</u> say to them except <u>what</u> You commanded me.
2C. (Gemini)	I did <u>not</u> say to them except <u>what</u> You commanded me to say.

In Example 2, the verse depicts a part of a conversation between Almighty Allah and Jesus Christ – peace be upon him. In this conversation, Allah asks Jesus whether he asked people to worship him and his mother, which Jesus denies. The verse contains Jesus’s denying reply in which he employs polite language to stress that he only told them what Allah had commanded him, which refers to being devoted only to Allah, his own Lord and their Lord (Ibn Ashour, 1984). The verse contains two homonyms of “إم”. Whereas the former homonym is a negative particle (lit. not), the latter functions as a relative pronoun (lit. what), together indicating the exclusion style.

To render Arabic limitation construction (negation + exception), 2A uses ‘only’, while 2B and 2C combine ‘not’ with ‘except’. On the other hand, ‘what’ is employed in all the translations as a relative pronoun. Although the Arabic particles have the same spelling, the chatbots manage to differentiate between them and achieve an outstanding rating. Moreover, the use of “only” in 2A highlights that the rendition was not constricted to the individual wordings of the ST but instead delivered a functional rendition of the limitation depicted in the ST. Again, the nature of LLMs is expected to produce acceptable renditions of verses, as there are dozens of translations of the Noble Qur’an online.

The adequacy and fluency ratings for Example 2 are presented in Table 2.

Table 2. Adequacy and Fluency Ratings for Example 2

AI System	Adequacy	Fluency
ChatGPT	Outstanding	Outstanding
DeepSeek	Outstanding	Outstanding
Gemini	Outstanding	Outstanding

As shown in Table 2, all three systems achieved outstanding adequacy and fluency, successfully distinguishing the two meanings of “إم” while producing clear and natural English translations.

Example 3

	هل لا دب ع نب يحيى يحيى يدل ايحي	من إف نامزلا مرك نم تام ام
	Mā māta min karami az-zamāni fa-innahu yahyā ladā Yahyā ibn ‘Abdillāh	
3A. (ChatGPT)	Whatever generosity time may take away, it <u>lives on</u> with Yahyā ibn ‘Abd Allāh.	
3B. (DeepSeek)	He who time’s generosity has not killed <u>lives on</u> with Yahya ibn Abdallah.	
3C. (Gemini)	The generosity of time does not die, for it <u>lives on</u> through Yahya ibn ‘Abdullah.	

Example 3 is a line by the famous poet Abu Tammam, praising one of the Abbasid-era officials, Yahya al-Barmaki (Ateeq, 2006). It says that if people cease to be kind, then you will find kindness in Yahya’s morals. In this verse, the praised person’s name “Yahya”, and the verb for “lives” are homonyms. As their spellings differ in the final letter, they are considered homophones.

The same TL choices which distinguish the verb from the name were made across all translations. Since homophones have different spellings, this could help the chatbots differentiate between the meanings of homonyms. The verb is rendered fairly literally however the translation is strengthened by presenting it as a phrasal verb. “Lives on” reflects the sense of “continues” that is presented in the ST while a direct rendition of “lives” would cause meaning loss as the term correlates with life. Additionally, 3C also improves on its rendition with the use of “through” rather than “with” as a translation of “يدل”. While this is not directly a part of the translation of the homonym itself, it still plays an important role in conveying its intended meaning. This rendition therefore emphasizes the relevance of context and the surroundings of a term when handling homonyms. In all the translations, the proper noun is transliterated and capitalized adequately presenting it as a given name. Therefore, translations 3A and 3B receive an above-average rating, while 3C receives the highest rating.

The adequacy and fluency ratings for Example 3 are presented in Table 3.

Table 3. Adequacy and Fluency Ratings for Example 3

AI System	Adequacy	Fluency
ChatGPT	Above Average	Outstanding
DeepSeek	Above Average	Outstanding
Gemini	Outstanding	Outstanding

As shown in Table 3, Gemini achieved outstanding adequacy, while ChatGPT and DeepSeek were rated above average. All three translations were rated outstanding for fluency, as they were clear and grammatically well-formed.

Example 4

مدهض رأ يف تدم ام مدهض رأو مدهراد يف تدم ام مدهرادف Fa-dārihim mā dumta fī dārihim wa-arḍihim mā dumta fī arḍihim.	
4A. (ChatGPT)	Treat their household kindly so long as you are among them, and please them so long as you are in their land.
4B. (DeepSeek)	Honor their home as long as you are in their home, and their land as long as you are in their land.
4C. (Gemini)	Treat them well as long as you are in their home, and please them as long as you are on their land.

In example 4, the line of poetry is attributed to Al-Qairawani and offers wisdom on dealing with people (Awni, 2021). It suggests that you need to accommodate yourself to those whom you cannot bear, by attempting to satisfy them as long as you are in their home or territory. In parallel terms, the former homonym in each pair is a command verb (lit. flatter and please, respectively), and the latter is a noun phrase (lit. their home and land, respectively).

Unlike the previous examples, the translations did not align across the systems. For the first pair of homonyms, 4A and 4B include the second homonym’s meaning in the rendition of the first. These renditions adequately convey the intended meaning of the first homonym while partially retaining the connection between the terms, a key stylistic element of the ST. However, 4A falls short in this regard as it employs a paraphrase when rendering the second homonym, meaning that this connection between the phrases is lost. Furthermore, this paraphrase weakened the equivalence, as “among them” suggests coexistence in a place that is not necessarily marked as belonging to the group in question. Additionally, both renditions slightly shift the meaning as well by ascribing the respect to the places rather than their people, this can be justified as the use of metonymy. 4B continues this pattern with the second homonym pair. Moreover, the sentence only contains a singular verb “honor” that functions with both clauses. 4C is the only rendition that completely separates the homonyms in both phrases. This results in 4C being the only rendition where both verbs are directly aimed at people.

The adequacy and fluency ratings for Example 4 are presented in Table 4.

Table 4. Adequacy and Fluency Ratings for Example 4

AI System	Adequacy	Fluency
ChatGPT	Above Average	Outstanding
DeepSeek	Above Average	Outstanding
Gemini	Outstanding	Outstanding

As shown in Table 4, Gemini achieved outstanding adequacy by clearly distinguishing the homonyms, while ChatGPT and DeepSeek showed some meaning shifts. All three systems achieved outstanding fluency.

Example 5

ينتملك ينتم لك املف		ينتم لك ىتح بابلا تقوط	
Ṭaraqṭul-bāba ḥattā kalla matnī		fāllammā kalla matnī kallamatnī.	
5A. (ChatGPT)	I knocked on the door until my two sides grew weary; when they grew weary, she spoke to me.		
5B. (DeepSeek)	I knocked on the door until my forearms grew weary, and when my forearms grew weary, you spoke to me.		
5C. (Gemini)	I knocked on the door until my back grew tired, and when my back grew tired, she spoke to me.		

In example 5, the line was composed by Isma'il Sabri and included in a narrative poem that depicts a short incident. Here, the poet describes how he continuously knocked on a woman's door without a response. The line indicates that she only answered after a long time by stating that she only spoke to him after he grew weary. In the ST, "she spoke to me" and "[my body] grew weary" are homonyms. The two are homophones, yet are distinguished orthographically by the space that separates the first homonym into two words. Yet both homonyms depict a verb phrase. Whereas the former homonym is a typical verbal sentence consisting of a verb and a subject in separate form (lit. My body grew weary), the latter also forms a verbal sentence as a connected sequence of verb, subject, and object (lit. She talked to me). Therefore, they are considered homophones.

Regarding the first homonym, all renditions involved a degree of paraphrase. 5A and 5B used the verbal phrase "grew weary" while 5A used "grew tired". The subjects of the verbs also varied and were rendered as "two sides", "forearms," and "back". The poem included repetition, and while 5B and 5C repeated the phrase exactly, 5A replaced the subject with a pronoun. Since the homonymy is not retained in the translations, this change does not influence the quality of the TT. Thus, all renditions of the phrase are acceptable as they retain the ST sense which presents tiredness or pain as the result of the efforts spent in knocking.

The renditions of the second homonym aligned in 5A and 5C, which accurately reflected the meaning of the ST. However, 5B contained a translation error in the rendition of the pronoun, changing the subject from "she" to "you". Ironically, this error can be attributed to homonymy between the ST and what the AI system mistakenly treated as the ST. The difference between "you spoke to me" and "she spoke to me" is presented in diacritics. When these diacritics are not written the two terms become homographs. However, if the subject was "you" the phrase would not be a homophone of "growing weary" which alters the ST.

This example highlights how the way homonyms are presented orthographically can influence how AI systems distinguish between them. It also shows that the difficulty of translating homonyms is not limited to those presented in the ST.

As a result, Translations 5A and 5C are rated as outstanding, whereas Translation 5B is rated below average because it contains a translation error resulting from the selection of an incorrect homonym.

The adequacy and fluency ratings for Example 5 are presented in Table 5.

Table 5. Adequacy and Fluency Ratings for Example 5

AI System	Adequacy	Fluency
ChatGPT	Outstanding	Outstanding
DeepSeek	Below Average	Outstanding
Gemini	Outstanding	Outstanding

As shown in Table 5, ChatGPT and Gemini achieved outstanding adequacy, while DeepSeek was rated below average due to its incorrect rendering of the pronoun. All three translations were fluent and grammatically acceptable.

Example 6

عيرق رغا إل اولصح نإ	مهنم ام قثالث لثولملا داس
عورف ميركلا سابعب تلعو	دعب لصف داسو عيبرلا داس
عيبر عيبرلاو لصف لصفلاو	يغولا مدتحا اذإ سابع سابع
Sāda al-mulūka thalāthatun mā minhum in ḥuṣilū illā agharru qarīʿ. Sāda ar-Rabīʿu wa sāda Faḍlun baʿdahu wa ʿalat bi-ʿAbbāsi al-karīmi furūʿ. ʿAbbās, ʿAbbās idhā ihtadama al-waghā, wa al-Faḍlu Faḍlun wa ar-Rabīʿu Rabīʿ.	
6A. (ChatGPT)	ʿAbbās is ʿAbbās when battle rages, and Faḍl is Faḍl, and al-Rabīʿ is al-Rabīʿ.
6B. (DeepSeek)	Abbas is Abbas when the battle rages, Al-Fadl is Al-Fadl, and Al-Rabeeʿ is Al-Rabeeʿ (i.e., a person’s true nature is revealed in crisis).
6C. (Gemini)	Abbas is Abbas when the battle intensifies, and Fadl is Fadl, and Rabiʿ is Rabiʿ.

In Example 6, the lines were said by Abu-Nuwas, a prominent Arab poet, complimenting three Abbasid officials. Each one is given an attribute identical to his name (Al-Meidani, 1996). Each of the three pairs is a typical nominal sentence consisting of a proper noun followed by a common one (lit. Abbas + lion; Fadhl + virtue/ advantage; Rabeeʿ + spring).

In each of the three cases, the latter homonym mimics the former’s proper name ‘Abbas’, ‘(Al)Fadl’ and ‘(Al-) Rabeeʿ’. Consequently, they miss the intended positive meanings of ‘lion’, ‘virtue’/‘advantage’ and ‘spring’, respectively. This leads to all the renditions to be rated unacceptable.

The adequacy and fluency ratings for Example 6 are presented in Table 6.

Table 6. Adequacy and Fluency Ratings for Example 6

AI System	Adequacy	Fluency
ChatGPT	Unacceptable	Outstanding
DeepSeek	Unacceptable	Outstanding
Gemini	Unacceptable	Outstanding

As shown in Table 6, all three systems were unacceptable in adequacy because they failed to convey the intended meanings of the homonyms. However, their translations remained fluent and grammatically acceptable.

Example 7:

اميهأ نأ بجع الو تدمهف	يديس اي كباتك تدمهف
أميظع أرّبو أميظن أرّدُ	نم ثلّمأت يّنأل لّاذو
fahimtu wa lā ʿajaba an ahīma. hu durran naẓīman wa birran ʿaẓīman.	Fahimtu kitābaka yā sayyidī, Wa dhālika li-annī taʿammaltu min
7A. (ChatGPT)	I understood your letter, my master; I understood—and it is no wonder that I should be enraptured.
7B. (DeepSeek)	I understood your letter, my lord, I understood—and no wonder I am enthralled.
7C. (Gemini)	I understood your letter, my master. I am lost in thought, and it is no surprise that I am lost.

In example 7, the lines are composed by Abul-Fat’h Al-Busti, also containing praise to the addressee. Here, the poet expresses understanding and adoration of the letter he received, due to its distinguished goodness and eloquence (Al-Meidani, 1996). In the first line, both homonyms are formed from past verbs with their subject

pronouns, the latter preceded by a coordinator (lit. I understood and so I got lost/so I fell in love, respectively). In each translation, the former homonym is accurately rendered. On the other hand, 7A and 7B repeat the same pattern for the latter ‘I understood’, referring to understanding the addressee’s message, most probably because the case represents homonymy. 7C opts for a different rendition, somewhat close to adoration, ‘lost in thought’, which might be linked to the fact that the presence of the same verb is mentioned at the very end of the line. Accordingly, 7A and 7B are rated as unacceptable, and 7C is above average.

The adequacy and fluency ratings for Example 7 are presented in Table 7.

Table 7. Adequacy and Fluency Ratings for Example 7

AI System	Adequacy	Fluency
ChatGPT	Unacceptable	Outstanding
DeepSeek	Unacceptable	Outstanding
Gemini	Above Average	Above Average

As shown in Table 7, ChatGPT and DeepSeek were unacceptable in adequacy, while Gemini was above average. Fluency remained high, although Gemini’s rendition was somewhat less natural.

5. Results and Discussion

The results show that the three chatbots vary in translation adequacy. When comparing all the cases, Gemini ranks first in the outstanding category with 56.52%, followed by ChatGPT with 52.17%, and then DeepSeek with 34.78%. The middle rates – above average, satisfactory, and below average have low shares. As for unacceptable renditions, DeepSeek ranks first at 47.28%, followed by ChatGPT at 34.78%, and then Gemini at 30.43%, as shown in Table 7.

Table 7: Ratings of Chatbot Translations for All Homonyms (n = 23 Homonym Items)

Rate	ChatGPT	DeepSeek	Gemini
Outstanding	12 (52.17%)	8 (34.78%)	13 (56.52%)
Above average	1 (4.34%)	1 (4.34%)	3 (13.04%)
Satisfactory	0 (0%)	2 (8.69%)	0 (0%)
Below average	2 (8.69%)	1 (4.34%)	0 (0%)
Unacceptable	8 (34.78%)	11 (47.28%)	7 (30.43%)
	23 (100%)	23 (100%)	23 (100%)

When the verses are excluded, the effectiveness of the renditions drops. The rate of outstanding adequacy becomes 44.44% for Gemini, 38.88% for ChatGPT, and 16.66% for DeepSeek. In terms of the middle categories, they also stay low. As for unacceptable translations, the highest share goes to DeepSeek by 61.11%, then 44.44% to ChatGPT, and 38.88% to Gemini, as indicated in Table 8.

Table 8: Ratings of Chatbot Translations for Homonyms Excluding Qur’anic Verses (n = 15 Homonym Items)

Rate	ChatGPT	DeepSeek	Gemini
Outstanding	7 (38.88%)	3 (16.66%)	8 (44.44%)
Above average	1 (5.55%)	1 (5.55%)	3 (16.66%)
Satisfactory	0 (0%)	2 (11.11%)	0 (0%)
Below average	2 (11.11%)	1 (5.55%)	0 (0%)
Unacceptable	8 (44.44%)	11 (61.11%)	7 (38.88%)
	18 (100%)	18 (100%)	18 (100%)

It is found that, aside from the Qur’anic verses, the translation choices vary in adequacy for several reasons: First, homophones tend to be easier to render since they have different spellings and, thus, could be comprehended as different words. Even when an outstanding rendition is not achieved, above-average or satisfactory choices are

found. Second, outstanding translations tend to be more closely related to contemporary Arabic, which may lack many homonyms found in traditional Arabic. While these chatbots are built on large language models (LLMs) for modern languages, older vocabulary and structures may not be well represented. Third, since people's names have specific meanings, they may lead chatbots to confuse proper nouns with common nouns. None of these cases was rendered correctly. Finally, if the SL homonyms do not have a single TL equivalent with both senses, which is the typical case, the TL choices are not expected to achieve homonymy of any type. This loss of pun applies to both human and AI chatbot translations.

6. Conclusion

The study concluded that AI chatbots differ considerably in their ability to distinguish Arabic homonyms, and the results match closely with the patterns elaborated (Aldawsari 2023, 2025; Al-Jarf, 2025). These previous studies indicated that LLMs attempt to produce fluent output but remain inconsistent in semantic precision, particularly when translation is context-sensitive. The results of this study confirm this tendency: while all three applications performed strongly on Qur'anic verses, mirroring the literature's assertion that widely translated texts rely on corpus familiarity, they showed reduced adequacy when translating poetic forms (Farghal & Haider, 2024, 2025), standard classical vocabulary, and context-dependent homonymy. This was addressed by Abdulmughni and Alrefae (2025) and Bin Hariz (2023), who emphasize that literary text, rhetorical texture, and culturally embedded meaning continue to fall outside the stable competence of current AI models.

Simultaneously, the findings strengthen the observations in the literature regarding the structural and linguistic nature of the challenge. The chatbots' relative success with homophones but difficulty with homonymy aligns with Wael et al. (2024), who presented the impact of missing diacritics and deep-rooted Arabic ambiguity on computational processing. Furthermore, the confusion between proper nouns and common-noun senses, as well as the inability to keep rhetorical parallelism, resonates with studies on human translators (Rabadi, 2015) that emphasize the need for explicit semantic guidance. Considering both, the results underscore that although LLMs represent a clear improvement over traditional MT systems, they remain limited in contexts that require standard classical lexical knowledge, rhetorical interpretation, or sensitivity to historical usage. Therefore, the previous studies and the present findings collectively suggest that AI chatbots can assist but cannot yet replace human expertise in translating Arabic homonymy, mainly in critical or stylistically rich discourse.

In this study, the findings should be interpreted as descriptive of the three systems' performance on this specific dataset at the time of testing. Observed differences cannot establish general model superiority or be attributed to differences in training, data exposure, or model architecture. Accordingly, no causal or broader comparative claims are made beyond the evaluated examples.

The results are merely applicable to the three systems, selected dataset, and testing procedure. They do not establish general superiority of the model or improvement over traditional machine translation, nor causal relationships between model training and translation errors. No paired inferential analyses, uncertainty intervals or effect sizes are provided for the study. The differences reported should not be taken as a genuine ranking of the systems. In other words, Comparisons between the three systems are descriptive. While this sample may have been rated higher for Gemini, it does not necessarily mean that it is better than ChatGPT or DeepSeek.

The differences in the time period, familiarity, language difficulty, and accessibility on the internet of the examples used in the Qur'anic and poetic texts may be a factor to consider during comparison. So, the differences observed cannot be traced back solely to the type of texts or the amount of training-data exposure. Potential training-data contamination is an important limitation. Success in Qur'anic examples might simply be due to the fact that previous experience with widely circulated translations has not been contextual disambiguation. The present design is unable to differentiate between these options; future research should add paraphrased examples, newly-constructed or less-circulated examples as controls. Consumer interfaces often change and are stochastic, so their single outputs are not reproducible. Hence, the results only reflect the performance of the systems at the time of the measurements and do not indicate a stable distinction between the model families. Future work should include several independent runs and use previously available model versions if they exist. One of the significant weaknesses is the possibility of contamination of the training data. Future studies could incorporate novel and paraphrased or low-exposure examples to further differentiate between contextual

disambiguation and memorized translation.

No inferential statistical comparisons were made between the systems in this study. The observations are paired because the same items were translated by all three systems and partly clustered because of some homonyms that were found in the same sentences. Therefore, the percentages reported should be taken as a description only and do not constitute an indication of statistically significant differences between the models. Moreover, the study did not use an independently validated human reference translation or a traditional machine translation system as a baseline, so its findings cannot be extended to the superior performance of LLM-based MT. Since only undiacritised inputs were used, it is not possible to separate the effect of the removal of diacritics from the effect of sentence context. Future studies should evaluate diacritizations and non-diacritizations under controlled conditions, with the same items.

It is recommended to investigate how AI chatbots process homonyms when they are introduced separately and how they process the translation of other rhetorical devices in Arabic, with emphasis on meaning and sound along with them. The role of AI chatbots in translating Arabic homonyms into English should be explored in the future, as it can impact the precision of the translation, language acquisition, and intercultural communication. Arabic words can have more than one meaning depending on context, morphology, diacritics, and cultural context. This can cause AI Chatbots to get the meaning wrong or literally, particularly when there is not enough context. However, users should know that the translations made by chatbots may be helpful but not necessarily accurate, especially in texts like academic papers, legal documents, religious texts, or literary works (Al-Abbas & Haider, 2020; Al-Salman & Haider, 2024; Farghal & Haider, 2024; Haider & Alkhatib, 2024), where words have multiple definitions. On the other hand, excellent AI performance can help users access Arabic information, facilitate bilingual conversations, and help translators by providing context-specific suggestions. In conclusion, AI chatbots for translating Arabic and English homonyms highlight the transformative power of technology in enabling quick translations and the need for human intervention to ensure accuracy, depth, and cultural resonance.

References

- Abdelmughni, H. M. A. (2025). Artificial intelligence in Arabic–English literary translation: Technological innovations, linguistic accuracy, and ethical implications. *International Journal of Language and Literary Studies*, 7(3), 107–129. <https://doi.org/10.36892/ijlls.v7i3.2120>
- Abul-Saud, M. (2010). *Irshad al-‘aql al-salim ila mazaya al-kitab al-karim* [Guiding the sound mind to the merits of the Noble Book]. Dar Alkutub Alilmiyya.
- Aburahme, M. (2021). The life of the Arabic language, its merit, and means of its development in the thought of Muhammad al-Khadr Hussein. *Jordan Journal of Applied Science–Humanities Series*, 27(2), 110–121. <https://doi.org/10.35192/jjoas-h.v27i2.8>
- Abu Romman, O. (2025). Conceptual metaphor theory in cognitive linguistics. *Jordan Journal of Applied Science–Humanities Series*, 43(2), 127–142. <https://doi.org/10.35192/jjoas-h.v43i2.1182>
- Al-Abbas, L. S., & Haider, A. S. (2020). Evaluating the accuracy and consistency in rendering Quranic terms with overlapping meanings into English. *AlBayan: Journal of Qur’an and Hadith Studies*, 18(2), 111–137.
- Al-Aloussi, M. (1995). *Rouh al-ma‘ani fi tafsir al-Qur’an al-‘atheem wa al-sab‘ al-mathani* [The spirit of meanings in the exegesis of the Great Qur’an and the seven oft-repeated verses]. Dar Alkutub Alilmiyya.
- Al-Darabee, M., Farghal, M., & Haider, A. S. (2025). Netflix versus Google Translate: A case study of the English–Arabic translation of scatological terms. *Studies in Linguistics, Culture and FLT*, 13(2), 77–95.
- Aldawsari, H. A. H. (2023). Comparing the performance of Google Translate and SYSTRAN on Arabic lexical ambiguity. *Arab World English Journal for Translation & Literary Studies*, 7(3), 19–34. <https://doi.org/10.24093/awejtls/vol7no3.2>
- Aldawsari, H. A. H. (2025). Evaluating the performance of large language models on Arabic lexical ambiguities: A comparative study with traditional machine translation systems. *World Journal of English Language*, 15(3), 354–363. <https://doi.org/10.5430/wjel.v15n3p354>
- Al-Ezzi, R., & Al-Qudah, I. (2024). English translation of verbal humour in Egyptian comedy films. *World Journal of English Language*, 14(1), 157–157. <https://doi.org/10.5430/wjel.v14n1p157>
- Ali, A. Q. (2022). *The challenges of translation of homonym* [Unpublished bachelor’s thesis]. University of Diyala.
- Ali, E. A. M. (2020). Lexico-semantic issues on translating homonymy and polysemy: A comparative study based on the Noble Qur’an. *QUR’ANICA: International Journal of Qur’anic Research*, 12(1), 45–68. <https://doi.org/10.22452/quranica.vol12no1.3>
- Ali, M. A., Brakhw, M. A., & Nordin, M. Z. F. (2014). Transferring polysemic words from Arabic into English: A comparative study of some samples from the Holy Qur’an. *Australian Journal of Basic and Applied Sciences*, 8(23), 38–43.
- Al-Jarf, R. (2025). AI translation of full-text Arabic research articles: The case of educational polysemes. *Journal of Computer Science and Technology Studies*, 7(1), 311–325. <https://doi.org/10.32996/jcsts.2025.7.1.23>
- Alkarazoun, G., & Riziq, D. (2025). A syntactic and pragmatic analysis of the colloquial expression ?inno ‘that’ in Jordanian Arabic: Evidence from social media conversation. *Languages*, 10(9), Article 205. <https://doi.org/10.3390/languages10090205>
- Al-Qudah, I. (2022). Censorship as enabling: Importing, distributing, and translating foreign films in the Arab Middle East. *Heliyon*, 8(8), Article e09977. <https://doi.org/10.1016/j.heliyon.2022.e09977>
- Alqudah, I. (2025). Who censors what and how? A linguistic inquiry into censorship in Arab audiovisual media. *Forum for Linguistic Studies*, 7(7), 402–413. <https://doi.org/10.30564/fls.v7i7.9990>
- Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science*, 8(4), 2483–2498. <https://doi.org/10.5267/j.ijdns.2024.5.009>
- Alsharif, B., & Khasawneh, R. (2025). Beyond the literal: Machine translation performance and strategies in rendering audiovisual political idioms. *Research Journal in Advanced Humanities*, 6(2). <https://doi.org/10.58256/4442fg06>
- Al-Zamakhshari, M. (1987). *Al-kashaf ‘an ghawamidh al-tanzeel* [The unveiler of the obscurities of revelation].

- Ammari, R., & Al-Ahmad, W. (2023). A corpus-based study of English synonyms: Small, little, tiny, and petite. *Jordan Journal of Applied Science–Humanities Series*, 36(2), 1–20. <https://doi.org/10.35192/jjoas-h.v36i2.542>
- Bin Hariz, M. A. (2023). Problems of translating synonyms and homonyms in the Holy Qur'an: A comparative study of two Qur'anic translations by Yusuf Ali and Ahmed Ali. *Electronic Journal of University of Aden for Humanities and Social Sciences*, 4(1), 129–136. <https://doi.org/10.47372/ejua-hs.2023.1.239>
- Bououden, R. (2025). Evaluating Arabic–English neural machine translation: Challenges across different text types. *Journal of Languages and Translation*, 5(Special Issue), 59–73.
- 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>
- 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.
- Fraihat, B. (2024). The effect of implementing a remedial plan based on reading strategies on improving the reading comprehension achievement of Jordanian EFL eleventh-grade students in the vocational stream. *Jordan Journal of Applied Science–Humanities Series*, 39(1), 18–26. <https://doi.org/10.35192/jjoas-h.v39i1.729>
- Garcia, M. (2021). Exploring the representation of word meanings in context: A case study on homonymy and synonymy. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-long.281>
- Haider, A. S. (2019a). The representation of Al-Megrahi's release in Arabic and English newspapers in 2009 and 2010: A corpus-assisted discourse study. *Dirasat: Human and Social Sciences*, 46(1, Suppl. 2), 297–317. <https://doi.org/10.35516/0103-046-986-019>
- Haider, A. S. (2019b). Syrian–Lebanese relations: A corpus-based critical discourse analysis of Bashar Al-Assad's speeches and interviews. *Dirasat: Human and Social Sciences*, 46(4), 551–570. <https://doi.org/10.35516/0103-046-004-040>
- Haider, A. S. (2019c). Using corpus linguistic techniques in (critical) discourse studies reduces but does not remove bias: Evidence from an Arabic corpus about refugees. *Poznan Studies in Contemporary Linguistics*, 55(1), 89–133. <https://doi.org/10.1515/psicl-2019-0004>
- Haider, A. S. (2026a). AI as a language tutor: Jordanian university students' perceptions of AI-generated feedback on their English. *Research Journal in Advanced Humanities*, 7(4), 1–17.
- Haider, A. S. (2026b). Trust in AI translation: Reception and adoption among Jordanian business professionals. *Research Journal in Advanced Humanities*, 7(4), 1–15.
- Haider, A. S., Akar, R. A., Saed, H., Tair, S. A., & Ismail, H. H. (2025). AI-powered Arabic translation of business and finance acronyms and abbreviations: Evidence from Gemini, ChatGPT, and Google Translate. In *Innovation, technologies, and business management* (pp. 519–528). Springer. https://doi.org/10.1007/978-3-031-84889-6_62
- Haider, A. S., & Alkhatib, R. (2024). Subtitling English legal acronyms into Arabic: Human vs machine. *Kutafin Law Review*, 11(4), 810–842.
- Haider, A. S., Alzghoul, O., & Hamadan, Y. (2023). Creating and experimenting a new parallel corpus of English–Arabic subtitles of culinary shows: A useful guide for translating food across cultures. *World Journal of English Language*, 13(8), 358–372. <https://doi.org/10.5430/wjel.v13n8p358>
- Haider, A. S., & Hussein, R. F. (2022). Modern Standard Arabic as a means of euphemism: A case study of the MSA intralingual subtitling of *Jinn* series. *Journal of Intercultural Communication Research*, 51(6), 628–643. <https://doi.org/10.1080/17475759.2022.2106289>
- Haider, A. S., Obeidat, M. M., Hamdan, Y., Abu-Rayyash, H., & AlZghoul, O. (2025). SDL Trados Live Cloud vs. traditional subtitling workflows: An empirical evaluation of performance. *Research Journal in Advanced Humanities*, 6(4). <https://doi.org/10.58256/fc624p07>
- Haider, A. S., & Shohaibar, R. S. (2024). Netflix English subtitling of idioms in Egyptian movies: Challenges and

- strategies. *Humanities and Social Sciences Communications*, 11, 1–13. <https://doi.org/10.1057/s41599-024-03327-4>
- Hammouri, D. M. (2025). The role of artificial intelligence in generating Arabic neologisms: A study of gaming localization. *Research Journal in Advanced Humanities*, 6(4), 1–15.
- Hauer, B., & Kondrak, G. (2020). One homonym per translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(5). <https://doi.org/10.1609/aaai.v34i05.6296>
- Holes, C. (2004). *Modern Arabic: Structures, functions, and varieties*. Georgetown University Press.
- Ibn Ashour, M. (1984). *Al-tahreer wa al-tanweer* [Verification and enlightenment]. Aldar Altunisiyya Linnashr.
- Ja'afreh, N. (2023). A pragmatic contrastive analysis of apology strategies in Jordanian Arabic and English language. *Jordan Journal of Applied Science–Humanities Series*, 37(2), 69–86. <https://doi.org/10.35192/jjoas-h.v37i2.601>
- Jahameh, S., & Al-Jazzar, F. (2025). The technological impacts of graphic design: Visual adaptations of English borrowed terminology in Arabic—A case study of Arabic typography. *Jordan Journal of Applied Science–Humanities Series*, 44(2), 154–173. <https://doi.org/10.35192/jjoas-h.v44i2.1853>
- Jaradat, A., Hammouri, D. M., Gwasmeh, M., & Haider, A. S. (2026). Reflexivity and reciprocity in two Arabic varieties: Evidence for REF-REC category. *Languages*, 11(2), Article 26. <https://doi.org/10.3390/languages11020026>
- Khasawneh, R., & Alsharif, B. (2025). Translating idiomatic expressions: A systematic review. *Forum for Linguistic Studies*, 7(10), 881–893. <https://doi.org/10.30564/fls.v7i10.11172>
- Khatatbah, D., & Ashour, M. (2024). The role of school administration in activating the use of information technology at schools in Irbid Governorate. *Jordan Journal of Applied Science–Humanities Series*, 38(2), 121–137. <https://doi.org/10.35192/jjoas-h.v38i2.718>
- Khudjaeva, R. N. M. (2023). Conceptual analysis of homonyms translation in English fiction literature. *Journal of Modern Educational Achievements*, 12(12), 318–322.
- Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, 81(1–3), 205–223. <https://doi.org/10.1006/brln.2001.2518>
- Li, S., Chen, J., Yuan, S., Wu, X., Yang, H., Tao, S., & Xiao, Y. (2024). Translate meanings, not just words: IdiomKB's role in optimizing idiomatic translation with language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38.
- Meqdad, S., Al-Bayyari, M., & Al-Taher, M. (2023). Semantic disparity between *Amr* (command) and *Nahy* (negative command) in the Noble Quran: An interpretive/contextual study. *Jordan Journal of Applied Science–Humanities Series*, 36(1), 1–10. <https://doi.org/10.35192/jjoas-h.v36i1.516>
- Naib, A. (2025). Algerian viewers' attitudes toward taboo language: A case study of Mourad Oudia's humorous YouTube content. *Jordan Journal of Applied Science–Humanities Series*, 44(1), 167–191. <https://doi.org/10.35192/jjoas-h.v44i1.1830>
- Obeidat, M. M., Haider, A. S., Tair, S. A., & Sahari, Y. (2024). Analyzing the performance of Gemini, ChatGPT, and Google Translate in rendering English idioms into Arabic. *FWU Journal of Social Sciences*, 18(4), 1–18. <https://doi.org/10.51709/19951272/Winter2024/1>
- O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *The Journal of Specialised Translation*, 17, 56–77.
- Rabadi, R. I. (2015). Homonymy and its effect on students of translation at Jordanian universities. *Arab World English Journal*, 6(4), 26–38. <https://doi.org/10.2139/ssrn.2843923>
- Ramadhan, H. S. (2023). Polysemy in English and Arabic: A contrastive study. *The Middle East International Journal for Social Sciences*, 5(1), 48–58. <https://meijss.org/wp-content/uploads/2023/08/MEIJSS-4858.pdf>
- Riziq, D., & Hamdan, J. (2023). The semantic structure of motion verbs in Jordanian Arabic: Talmy's typology revisited. *Italian Journal of Linguistics*, 35(2), 143–170. <https://doi.org/10.26346/1120-2726-196>
- Ryding, K. C. (2005). *A reference grammar of Modern Standard Arabic*. Cambridge University Press.
- Tannous, B., & Haider, A. S. (2025). Assessing the accuracy of AI tools (Google Translate and Gemini) in translating Arabic idiomatic expressions and proverbs into English. *International Journal on Artificial Intelligence Tools*, 34(4–5), Article 2550016.

- Versteegh, K. (1997). *The Arabic language*. Edinburgh University Press.
- Wael, T., Elrefai, E., Makram, M., Selim, S., & Khoriba, G. (2024). Pirates at ArabicNLU2024: Enhancing Arabic word sense disambiguation using transformer-based approaches. In *Proceedings of the Second Arabic Natural Language Processing Conference* (pp. 372–376). Association for Computational Linguistics.
- Wright, W. (2011). *A grammar of the Arabic language*. Cambridge University Press.
- Yang, C., & Zhang, Y. (2024). The application of Chinese–English machine translation in cultural terminology translation: A case study of homographs and polysemous words. In *2024 7th International Conference on Artificial Intelligence and Big Data (ICAIBD)* (pp. 226–231). IEEE. <https://doi.org/10.1109/ICAIBD62003.2024.10604629>