

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

Saleh Al-Salman^a and Ahmad S. Haider^{a,b*}

^aDepartment of English Language and Translation, Applied Science Private University, Amman, Jordan

^bMEU Research Unit, Middle East University, Amman, Jordan

CHRONICLE

Article history:

Received: January 19, 2024

Received in revised format: February 15, 2024

Accepted: May 12, 2024

Available online: May 12, 2024

Keywords:

AI translation

Machine translation

Research titles

Interdisciplinary research

Accuracy

Evaluation

ABSTRACT

Breakthroughs and advances in translation technology by virtue of AI-powered MT tools and techniques contributed significantly to providing near-perfect translation. This study aims to evaluate the accuracy of three translation technologies (Google Translate, Gemini, and ChatGPT) in translating multidisciplinary Arabic research titles in the Humanities and Social Sciences into English. A corpus of 163 titles of Arabic research articles from various disciplines, including media studies, literature, linguistics, education, and political science, was extracted from a Scopus-indexed journal, namely *Dirasat: Human and Social Sciences Series*. The research methodology in the present study lends itself largely to Koponen's (2010) translation error strategy framework. Based on the data analysis, the findings showed that the renditions provided by these programs were categorically marked with either sense or syntax errors, which often rendered the translations inaccurate. Many polysemous terms with multiple related senses were mistranslated. The results showed that the Gemini translations contained the least errors. In contrast, the human translations contained the least mistranslation and diction errors. Google Translate and ChatGPT, on the other hand, contained the highest number of equivalence-based errors. Unexpectedly, the human translations contained the highest number of syntactic errors, reflecting a lack of target language proficiency. The study's conclusions and findings would be beneficial to translators, students, and scholars who may consider translating their Arabic study research titles and abstracts through the most commonly used AI tools.

© 2024 by the authors; licensee Growing Science, Canada.

1. Introduction

The interplay of computers and human language is the focus of natural language processing (NLP). Machine translation (MT), which converts text between languages, is one of the most significant NLP tasks (Almahasees, 2021). As AI-powered machine translation (MT) technologies can translate texts rapidly and accurately, even in difficult fields like science and technology, they have grown in popularity in recent years (Bouguesmia, 2020). Nowadays, AI-powered machine translation tools are widely used. Concerning the translation of specialised language, it can be challenging to translate research titles accurately because they are frequently very technical and contain complicated concepts, specialised jargon, and terminology. Furthermore, since research titles are often used to convey and highlight the study focus to a broad audience, accurate and fluent translation is required (Bowman & Kinnan, 2018). With the world being more globalised, international consumers have been looking for quality services that meet their local needs and cultural norms (Alrousan & Haider, 2022). To this end, translation technology providers have consequently transformed their products from localisation to internationalisation and eventually to globalisation (Hartley, 2009). Such transformations have been made possible through breakthroughs and advances in

* Corresponding author.

E-mail address A_haidar@asu.edu.jo (A. S. Haider)

translation technology by virtue of AI-powered MT tools and techniques, which contributed significantly to providing near-perfect translation (Akasheh, Haider, Al-Saideen, & Sahari, 2024). Translation technology tools fall under three main categories: (1) computer-assisted tools (CAT), (2) machine translation (MT) software, and (3) translation management systems (TMS). Translation technology hinges on using special software tools to convert texts between languages (Yves, 2019). The quality of machine translation output usually depends on many factors, top among which are the content of the text, the machine translation tool used, and the language pairs (Hutchins, 2005). Considering this formula, companies can choose either low-level, raw MT without editing or opt for a machine translation post-editing task, which can be either light or full. In this case, a human translator is hired to produce a more refined output, depending on the goal (Jia, Carl, & Wang, 2019).

In the Middle East and North Africa (MENA) region, where Arabic is spoken widely, artificial intelligence (AI) techniques are now being used to translate different texts, including academic articles, from Arabic to English. Research titles may be difficult to translate as they contain many specialised terms. Research can be seen as a representation and documentation of human knowledge, the sciences, and their advancement (Hussein, Haider, & Al-Sayyed, 2021). Naturally, they will contain specialised areas of knowledge that are described through advanced language, specialised terminology, and jargon. Furthermore, titles must be a concise reflection of an entire work (Al-Salman & Haider, 2021; Haider & Hussein, 2020). This means that the translation of research titles requires an understanding of the language and the subject at hand.

In this study, the accuracy of translation of interdisciplinary Arabic research titles into English has been evaluated using three translation tools (Google Translate, ChatGPT, and Gemini). This study attempts to address the following two research questions:

- How do the translation accuracies of AI tools (Google Translate, ChatGPT, and Gemini) for Arabic research titles compare to human translation accuracy?
- What common mistakes do AI tools make when translating Humanities/ Social Sciences research titles from Arabic to English?

2. Review of Related Literature

With its beginnings more than five decades ago when computers were available, translation technology was only of modest and limited capacity (Almahasees, 2018). For example, only in 2006, Google launched Google Translate, and in 2016, Google introduced Neural Machine Translation (NMT) to surpass CAT (Bin Dahmash, 2020). But as the world has become more interconnected and globalised, translation technology has been growing steadily and forcefully to often provide near-perfect translation. The rapidly expanding machine translation market has led end-users and consumers to make well-informed decisions based on the accuracy, readability, and adequacy of the translation output before deciding on the choice of any given machine translation engine. Based on performance and quality standards, comparative performance assessment between AI-powered translation technology tools, including Google Translate, Microsoft Translator, and ChatGPT, among others, has become necessary (Khoshafah, 2023; Son & Kim, 2023). While acknowledging the impact of this remarkable AI-leveraged achievement, translation technology products have yet to undergo thorough monitoring and quality assessment. According to Yilmaz, Naumovska, and Aggarwal (2023, p. 1), “Although artificial intelligence (AI) has the potential to significantly disrupt businesses across a range of industries, we have limited empirical evidence for its substitution effect on human labour”.

Using Koponen’s (2010) translation error category, Rusadi and Setiajid (2023) evaluated the translation accuracy of Google Translate and ChatGPT by examining the errors they produced in translating the installation GUI texts of Windows 11 Education into Indonesian. Their findings identified five of Koponen’s six error categories generated by Google Translate and Chat GPT, totalling 29 errors. Their distribution showed: “omitted concepts (17.24%), added concepts (24.13%), untranslated concepts (20.68%), mistranslated concepts (3.44%), and substituted concepts (34.48%)” (p. 1). These findings suggest that further improvements are needed to upgrade the translation output of these two popular MT tools. Another manifestation of AI-based translation tools is Instagram Translate (IG), which, in 2016, succeeded in translating users’ photo captions in different language pairs. In a study aimed at measuring the IG Translate translation performance compared to human performance, Putri and Setiajid (2021) used Koponen’s translation error strategy framework to test the effectiveness of IG Translate in translating the photo captions on Jokowi’s official Instagram account. Their findings showed differences between IG translation and human translation. Human translation was preferred in handling special terms, with a larger linguistic repertoire of grammar and vocabulary. In addition, human translation provides a higher level of accuracy, readability, and acceptability than machine translation, both in translating formal and informal language. Conversely, machine translation is a better tool for translating a source language text with common words, good grammar, and formal language. Machine translation tools continued to produce quick translation content that needed further post-editing (Almahasees & Jaccopard, 2020). On the correctness of machine translation in conveying the intended meaning, Koponen and Salmi (2015, p. 123) conducted a machine translation post-editing task of two English newspaper articles (673 words, with 32 sentences) from English into Finnish. The results showed that it was possible to grasp the meaning of half of the machine-translated sentences without having access to the source text. The results also suggested that “errors in word forms and mangled relations are the kind of machine

translation errors that are easier to recover from context, while mistranslated idioms and missing content seem to be more critical to understand the meaning.”

To identify types of errors in machine translation of news texts from Arabic into English, Abdelaal and Alazzawie (2020, p. 408) launched a study to assess the quality and semantic adequacy of the Google Translate output. Through a mixture of qualitative and quantitative approaches, they also aimed to find out how much human translation is needed to fix the emerging errors. The results showed that “omission, which is a lexical error, and inappropriate lexical choice, which is a semantic error, are the most common errors”. Consequently, whereas machine translation can be helpful in the translation process for ease and speed, this happens at the expense of accuracy, which requires the work of a human translator, at least for editing and proofreading. The increasing growth and spread of AI and GPT worldwide have led stakeholders to draw a comparison between them based on their accuracy, adequacy, and overall performance according to the consumers’ specific needs. Seeking to evaluate ChatGPT as a machine translation tool, Jiao, Wang, Huang, Wang, and Tu (2023, p. 1) used the criteria of translation prompts, multilingual translation, and translation robustness to evaluate ChatGPT performance. The results revealed that “ChatGPT performs competitively with commercial translation products (e.g., Google Translate) on high-resource European languages but lags behind significantly on low-resource or distant languages.” However, following the emergence of GPT-4, its performance has improved significantly even with distant languages, competing well with other commercial translation tools. George and George (2023) argue that as a natural language processing (NLP) model, ChatGPT combines GPT-2 and GPT-3 models to provide varied services, including voice and conversation. It is a powerful tool that will boost the business sector through promoting e-commerce and marketing, international travel, health, education, and other supporting services. Khoshafah (2023) reports that ChatGPT has contributed significantly to enhancing cultural communication through translating various language pairs, including English and Arabic. However, the author stresses that the ChatGPT translation output needs continuous evaluation and assessment to ensure accuracy and adequacy through better lexical choices and idiomatic expression based on linguistic and cultural context. This is particularly important in translating highly specialised medical, legal, scientific, or literary texts.

Although several studies examined the translation accuracy of MT tools across languages, little attention has been paid to similar studies in the context of translating academic research titles. Therefore, this study fills this gap by evaluating the accuracy of three AI-powered translation technologies (Google Translate, Gemini, and ChatGPT) in translating multidisciplinary Arabic research titles in the Humanities and Social Sciences into English.

3. Research Methodology

In this study, the three AI tools (Google Translate, ChatGPT, and Gemini) have been used to translate each title in the corpus into English. The following configurations are used to generate the translations. The source language is Arabic, the target language is English, and the translation mode is "idiomatic."

3.1. Data Selection and Corpus Compilation

A collection of 163 Arabic research titles covering a range of Humanities/Social Sciences is created. The following criteria are applied when choosing the titles. (1) published in *Dirasat: Human and Social Sciences Series*; (2) appear in Volume 50 and Issues 1,2,3,4,5 and 6 (2023); and (3) written in Arabic as a source language. To elaborate, the *Dirasat: Human and Social Sciences Series* was established in 1974 by the Deanship of Scientific Research at the University of Jordan. It is an academic journal that is peer-reviewed and double-blind. The Journal is published in print and online versions. It publishes high-caliber articles across a wide range of social science and humanities topics. The articles are indexed in SCOPUS (Elsevier), which is the largest abstract and citation database of peer-reviewed literature, and it contains indexes to journal articles in the human and social sciences. Table 1 shows the number of investigated titles.

Table 1

Number of investigated titles extracted from *Dirasat: Human and Social Sciences*

Volume and Issue	Number of Articles
Vol. 50 No. 1 (2023)	33
Vol. 50 No. 2 (2023)	27
Vol. 50 No. 3 (2023)	25
Vol. 50 No. 4 (2023)	21
Vol. 50 No. 5 (2023)	26
Vol. 50 No. 6 (2023)	31
Total	163

3.2. Machine Vs. Human Translations

In this study, the accuracy of the translation of interdisciplinary Arabic research titles into English has been evaluated using three translation tools (Google Translate, ChatGPT, and Gemini). Google Translate is known as a free online language

translation service which supports many languages through translating texts, documents, websites, and spoken words and using the camera to translate texts in real time (ElShiekh, 2012). Unlike Google Translate, Gemini is a multi-task tool that -- in addition to providing translation services and language-related tasks-- answers questions and gives information according to users' needs (Aydin, 2023). On the other hand, ChatGPT is basically a conversation and dialogue language model based on Open AI's ChatGPT. It has virtual assistants and provides coherent and detailed responses to users' questions (Henrickson & Meroño-Peñuela, 2023). The human translation rendered in all the examples included in this study represents the translation provided by the authors of the articles extracted from the research journal *Dirasat: Human & Social Sciences Series*, Volume 50, numbers 1-6 (2023). The product does not reflect a professional translation provided by top-notch translation experts. In fact, the authors of the said articles are neither specialists in English nor in Translation. This is not to be confused with the human translators' team, who evaluated the product of the human and machine translators represented in this study.

3.3. Framework

A taxonomy of error categories, specifically designed for analysing machine as well as human translation outputs and proposed by De La Cruz-Cabanillas and Tejedor-Martínez (2016), was employed. Their framework builds upon established error classifications from previous studies, including those by Ledesma (2001), Koponen (2010), Popović and Ney (2011), Santos Gargallo (1993), and Vázquez (1999).

4. Research Findings and Discussion

The current research comprises two sections. First, a quantitative section that measures the frequency of errors made by each translator. The errors in the translations detected in the corpus were counted and categorised based on type. The second section gives a more detailed account of the types of errors, in addition to providing examples.

4.1. Quantitative Analysis

This section quantifies the errors made by the four examined translators. The errors are categorised into six main sections, which are further divided into various subsections. The main error types are Syntactic errors, diction errors, mistranslation errors, addition, omission, and untranslated concepts (Table 2).

Table 2
Categorisation of errors by the investigated human and machine translators

Type of Errors	Human*	Google Translate	ChatGPT	Gemini
Syntactic errors	28	15	14	7
Diction errors	22	29	22	20
Mistranslation	7	15	18	12
Addition	3	0	1	2
Omission	9	3	7	7
Untranslated concepts	23	21	20	18
Total	92= 28.5%	83= 35.9%	82= 35.5%	66= 28.6%

Syntactic errors are errors related to language and grammar rather than issues with equivalence. Syntactic errors lead to unnatural language and decreased readability. Furthermore, some syntactic errors may affect the sense of the text. The human translations contained the highest number of syntactic errors, with a total of twenty-eight errors, while Gemini committed the least errors, with only seven. The syntactic errors found in the translations are related to word order, missing article, extra article, wrong preposition, extra preposition, pluralisation, derivation, conjunction, and incorrect attribution.

Like mistranslation errors, **diction errors** relate to word choice and sense. In these errors, the renditions are equivalent but not entirely exact, resulting in shifts in nuance. These errors are related to imprecise meanings, uncommon terms, and incorrect connotations.

Mistranslation errors involve rendering terms with non-equivalents, which shifts the meaning of the text. In this work, which focuses on research papers, the mistranslated terms were categorised according to specialisation. These are related to keyword (specialised terms and jargon) mistranslation and content word mistranslation.

Addition errors involve adding a word to the target text that did not exist in the source text. Most of these additions were related to the sense and only resulted in redundancy. Two such additions were made in the human and Gemini translations, one in the ChatGPT translations and none in the Google translation. The human translation also contained another additional error that deviated from the sense of the text.

Omission errors involve leaving out elements of the source text in the target texts. These omissions were categorised into key and content words like mistranslation errors, in addition to function words. These are related to keyword omission, content word omission, and function word omission.

Untranslated Concepts are those retained in their source text form. These include titles, cultural expressions, linguistic expressions, untranslated terms, unrelated transliteration, and unofficial transliteration.

Regarding the five investigated fields, the field of linguistics was the most problematic, reflecting the largest number of translation errors, followed by Literature, Media studies, and Education. Political science, on the other hand, consistently contained the least errors among the translations. This shows that the nature of the field affects the quality of translation and emerging errors.

As Table 2 shows, the categorisation of errors of all five types combined totalled 323. Of this total, the human errors made 92, with 28.5%. The three AI-powered tools combined made 231, with 71.5%. These findings, supported by numbers and percentages, provide clear-cut answers to research question 1 in this study, namely, "How do the translation accuracies of AI tools (Google Translate, ChatGPT, and Gemini) for Arabic research titles compare to human translation accuracy? It is quite obvious that human translation is way more advanced and accurate compared to the outcome of machine translation tools, as the percentages against each one show. Interestingly, our findings, which support the product of human translators, are consistent with the research findings reached by Putri and Setiajid (2021), who reported that human translation was preferred in handling special terms with a larger linguistic repertoire of grammar and vocabulary. In addition, human translation provides a higher level of accuracy, readability, and acceptability than machine translation, both in translating formal and informal language. Conversely, machine translation is a better tool for translating a source language text with common words, good grammar, and formal language. On the other hand, the results of Table 2, based on the figures and percentages presented, provide clear answers to research question 2, namely, "What common mistakes do AI tools make when translating Humanities/ Social Sciences research titles from Arabic to English? The research findings identified the major types of errors, which were defined as common errors in all five fields, both by human translators and AI-powered tools alike, but to varying degrees. These errors covered the major types of errors listed in Table 2 above and the subtypes of each to include both content/meaning and form. These problems included improper diction, connotations, mistranslation due to omission or addition, cultural expressions, and metaphors, among many others. The language-related problems covered various aspects, including morphology, word order, missing articles, wrong prepositions, conjunctions, derivation, function words, and specialised jargon, among others. This is consistent with the research findings of Abdelaal and Alazzawie (2020, p. 408), who reported that "omission, which is a lexical error, and inappropriate lexical choice, which is a semantic error, are the most common errors". Consequently, whereas machine translation can be helpful in the translation process for ease and speed, this happens at the expense of accuracy, which requires the work of a human translator, at least for editing and proofreading.

4.2. Qualitative Analysis

This section provides examples of the various error types and analyzes how they may affect the quality of the renditions both in form and function.

4.2.1. Media Studies

This section examines titles of Arabic research articles within the realm of media studies, as Table 3 shows.

Table 3

Examples of untranslated concepts, syntactic errors, and various error types identified in the titles of articles in the field of Media Studies

No	Source Text	Human	Google Translate	ChatGPT	Gemini
1	توظيف الإثارة في الحوارية التلفزيونية العربية تحليل مضمون برنامجي "أحمر بالخط العريض" و"أجرأ الكلام"	The Use of <i>Excitement</i> in Arab Talk Shows: An Analytical Study of the Two Programs 'Ahmar Belkat Al Areed' and 'Ajra Al Kalam'	Employing <i>excitement</i> in Arab television talk shows, analyzing the content of the programs "Ahmar Bi Khat Al Arid" and "Ajra Al Kalam"	Employing <i>Excitement</i> in Arabic Television Talk Shows: A Content Analysis of "Ahmar Bil Khatt Al Aareed" and "Ajras Al Kalam"	The Use of <i>Excitement</i> in Arabic Television Talk Shows: A Content Analysis of the Programs "Red with the Bold Line" and "The Boldest Speech"
2	حضور "العربي" في وسائل الإعلام الألمانية: الواقع والتحديات دراسة وصفية تحليلية في تلفزيون "دويتشي فيللي" ومجلة "دير شبيغل"	The Arabs in the German media: Reality & representation "DW" TV & "Der Spiegel" magazine as a case study	The presence of "Al-Arabi" in the German media: reality and representations, a descriptive and analytical study on "Deutsche Welle" TV and "Der Spiegel" magazine	The Presence of 'Arabs' in German Media: Reality and Representations - A Descriptive Analytical Study on 'Deutsche Welle' Television and 'Der Spiegel' Magazine	The Presence of "the Arab" in German Media: Reality and Representations: A Descriptive and Analytical Study in Deutsche Welle Television and Der Spiegel Magazine

Examples 1 and 2 both contain phrases within quotations as part of the source title. In example 1, the quoted phrases represent the titles of television programs. These titles consist of meaningful phrases. The first title can be translated as "Red in bold (font)", and the second title can be translated as "the boldest of speech". Despite the translatability of these phrases, the human translators (i.e., the authors of the articles published in *Dirasat: Human & Social Sciences Series*, who are not usually professional translation experts), Google Translate, and ChatGPT resorted to transliteration, thus leaving the titles as untranslated segments. Moreover, ChatGPT's transliteration contained another error by the addition of the letter s to the word "ajar," which does not exist either phonetically or orthographically in the original Arabic word. Gemini, on the other hand, managed to

properly convey the meaning of the quoted titles without any significant errors. In light of the above, it may be argued that the human translation is the most appropriate since it employs the official transliteration provided as the English title of the show reads. Thus, depending on the assessment criteria, the human translation or the Gemini translation can be deemed acceptable as one represents the official title while the other represents the sense of the title. Conversely, although Google and ChatGPT technically employed the same strategy as the human translator, the inconsistency with preexisting criteria results in unacceptable renditions. This example also contains syntactic errors as both the human translator and Google use the term “Arab,” which describes people or places, instead of the term “Arabic,” which describes the language.

While quotation marks are used in example 1 to indicate a title, in example 2, they are employed to mark the term between them, altering its implications. The purpose of such punctuation is to highlight the discourse and not the form. Regardless, Google resorted to transliteration, once again failing to understand the function of the quotation marks. This resulted in another untranslated concept error. While the other translations did not commit the same mistake, yet again, it was only the translation of Gemini that properly conveyed the sense of the source text. The rendition as “the Arab” depicts the grouping and generalizing, or stereotyping conveyed in the source text. The use of the plural s in the human and ChatGPT translations removes the intentional hominisation and thus rules out any possible interpretations or implications. The removal of the definite article by ChatGPT has a similar effect as well.

Example 2 also contains titles between quotations. However, unlike example 1, the titles in the source text do not consist of Arabic words but are transliterations of the names of foreign channels. Here, the three translation programs provided the English version of these channel names, while the human translator did the same for one channel but provided the acronym of the other. In this case, the English versions were transliterations in themselves as well. However, since the transliterations serve as the official English branding and work from the original programs and not from the Arabic text, this constitutes a proper translation and not an error of untranslated concepts. The two examples show that what is between the quotation marks informs which strategy is suitable for their rendition.

Aside from retention that results in untranslated concept errors, sense can be left out of the rendition through omission errors as well (Table 4).

Table 4

Examples of omission errors and various error types identified in the titles of articles in the field of Media Studies

No	Source Text	Human	Google Translate	ChatGPT	Gemini
3	اتجاه الخطاب الإعلامي الأردني نحو إسرائيل ٢٠٢١-٢٠١٧	The Attitude of the Jordanian Media Discourse towards Israel 2017-2021	The direction of Jordanian media discourse towards Israel 2017-2021	The Jordanian Media Discourse towards Israel 2017-2021	The Direction of Jordanian Media Discourse Towards Israel 2017-2021
4	سيمانية الصورة في المواقع الإلكترونية السياحية الرسمية دراسة تحليلية سيميولوجية لموقع "روح السعودية"	The semiotics of the Image in the Official Tourism Websites, A Semiological Analytical Study for the “Visit Saudi” Website	Image semiotics on official tourism websites: A semiological analytical study of the “Spirit of Saudi Arabia” website	Semiotics of Image in Official Tourism Websites: A Semiological Analytical Study of “Spirit of Saudi Arabia”	Semiotics of Images in Official Tourism Websites: A Semiologic Analytical Study of the “Spirit of Saudi Arabia” Website
5	حضور الفنانة التشكيلية الفلسطينية في فنون الميديا الجديدة: أيقونة الخريطة وذاكرة الشتات وجمالية الهوية	The Presence of Palestinian Female Visual Artist in New Media Arts: Map's Iconography, Diaspora's Memory, and Aesthetics of Identity	The presence of the Palestinian visual artist in new media arts: the icon of the map, the memory of the diaspora , and the aesthetics of identity	Presence of Palestinian Visual Artists in New Media Arts: Icon of the Map, Memory of Displacement, and Aesthetics of Identity	The Presence of the Palestinian Visual Artist in New Media Arts: The Icon of the Map, the Memory of the Di- asporea, and the Aesthetics of Identity

Examples 3 and 4 show errors in the form of ‘omission’ made by ChatGPT. In example 3, the term “attitude of” was not rendered, resulting in a partial loss of meaning. In example 4, the word “website” was omitted. Yet, since it was a repeated term and its use could have been inferred from the context, the effect on meaning was insignificant. The redundancy of the source text allowed for omission in this case without resulting in a loss of meaning. This shows that the prominence of a term impacts the acceptability of its omission. The words that make up a sentence carry different relevance to its sense. Thus, words that contribute more to the meaning cause larger issues when omitted.

Example 5 shows that the omission of complete words and phrases is not the only way information can be left out of the target text. The word “artist” in the Arabic source text is inflected with the feminine marker. English grammar does not gender nouns in the same way as Arabic, and the norm is for nouns to be gender-neutral. Therefore, translating the source feminine artist as the neutral English artist results in translation loss that requires compensation. Hence, the human translator employs the adjective female. This compensation is not a form of addition but a translation of the feminizing suffix. A rendition without this adjective is one that omits this source text suffix and the meaning it carries.

Aside from omission, examples 3 and 4 show errors in the form of mistranslation. Example 3 contains the word “اتجاه”, which could represent the meaning of “attitude” or “direction”, depending on the context. In this example, “attitude” is the accurate sense. As previously stated, ChatGPT omitted this word. Regarding other translations, only the human translation used the word “attitude”. Google and Gemini both use the word “direction”. Although this rendition is literal and is generally an equivalent of the Arabic word, its meaning is related to location rather than feeling and is thus considered nonequivalent in this context.

While example 3 contains a mistranslation error stemming from polysemy, the error in example 5 stems from homophony. The word “spirit” is a homophone for the colloquial “go/visit” in Arabic. Since machine translation works mostly with standard Arabic, it translates the word as “spirit”. The human translation was able to pull from wider language pools and pick the desired sense, translating the term as “visit”, according to context. Furthermore, the phrase ‘روح السعودية’ in this example refers to another website which has an official English name. Therefore, the correct sense can be confirmed as “visit” by examining the website.

Table 5 shows examples of diction and syntactic errors highlighting the relevance of word choice and order in the reflection of sense.

Table 5

Examples of various error types identified in the titles of articles in the field of Media Studies

No.	Source Text	Human	Google Translate	ChatGPT	Gemini
6	انعكاسات التوتير على قيم الأسرة الأردنية من وجهة نظر الوالدين: دراسة ميدانية من منظور تربوي إسلامي	Parents' Viewpoints on the Effects of the YouTube Phenomenon on Jordanian Family Values from an Islamic Educational Perspective: A Field Study	The repercussions of the YouTube phenomenon on Jordanian family values from the parents' point of view: a field study from an Islamic educational perspective	The Impact of the YouTube Phenomenon on Jordanian Family Values: A Field Study from an Islamic Educational Perspective	Reflections of the YouTube Phenomenon on the Values of the Jordanian Family from the Parents' Perspective: A Field Study from an Educational Islamic Perspective
7	انعكاسات التوتر الناتج عن تغطية الصحفيين الفلسطينيين للمواجهات بين المتظاهرين وجنود الاحتلال الإسرائيلي على أدائهم المهني	The Implications of the Tension Resulting from Palestinian Journalists' Coverage of the Confrontations between Palestinian Demonstrators and Israeli Occupation Soldiers on their Professional Performance	The repercussions of the tension resulting from Palestinian journalists' coverage of confrontations between Palestinian demonstrators and Israeli occupation soldiers on their professional performance	Reflections of Stress Resulting from the Coverage of Palestinian Journalists on Confrontations between Palestinian Protesters and Israeli Occupation Soldiers on Their Professional Performance	The Reflections of the Stress Resulting from the Coverage of Palestinian Journalists of the Clashes between Palestinian Demonstrators and Israeli Occupation Soldiers on their Professional Performance

Example 6 shows more instances of mistranslation. While the errors were all found in the same example, they were in the renditions of different phrases for each translation. This suggests that problematic terms are not necessarily the same for each translator. As can be seen, the human translation did not contain any mistranslation errors. ChatGPT contained an error by mistranslating the word “Youtuber”, which describes someone who uploads videos on the website YouTube. Gemini mistranslated “family values” as the “values of the family”. While this syntactic rearrangement usually results only in a paraphrase that does not affect meaning, this is not the case here. “Family values” is a meaningful collocation that acts as one unit. Therefore, its separation changes its sense. A change in family values shows a concept being affected, while changes in the values of a family show people changing their beliefs.

While mistranslation errors cause a complete shift in the sense of the source text or some of its elements, diction and word choice may also affect the sense, but to a lesser extent. For instance, the word ‘انعكاسات’ “reflections” was translated into “repercussions” and “impact” by Google and ChatGPT, respectively. Such renditions are nearly synonymous with the word “reflections” but have stronger negative connotations. Depending on the context, the acceptability of these renditions will vary.

Example 7 shows more instances of diction errors. Again, the word “reflections” was translated differently by the human and Google translators as “implications” and “repercussions”, respectively. In this context, the word “implications” is too weak, while “repercussions” implies accountability. Thus, neither rendition is ideal in this context.

ChatGPT uses the word “protestors” while the other translators use the term “demonstrators” instead. Although the two words are close in meaning, a protest’s call for change is stronger than that of a demonstration. This indicates that “protester” makes a better representation of the described group.

To sum up, the types of errors which prevailed in Tables 3, 4, and 5, representing the field of “Media Studies” varied in nature as they included problems at the two levels of (1) content (i.e., sense and meaning), and (2) form (i.e., language structure and mechanics of writing). Content errors were detected through keyword mistranslation and untranslated words, omission, diction, and transliteration. This was evidenced in translating titles and quotations which retained the original form. Other aspects of mistranslations included problems in form and language structure, which included wrong use of prepositions, articles, gender suffixes, function word omission and addition. As indicated in our analysis of the three tables above, the translation errors covered all six types of errors outlined earlier, namely syntactic, diction, mistranslation, addition, omission, and untranslated concepts. These findings show that all three AI translation tools fall short of providing flawless rendering of the source language texts under investigation. Upon tallying the number of errors for each of the three machine translation tools in the Media Studies field, the results showed Google Translate with 4 errors, ChatGPT with 6, and Gemini with 2. With only one error for the human translation, we conclude that human translation of Media Studies titles, despite some minor glitches, remains a more reliable source to contend with.

4.2.2. Literature

This section examines titles of Arabic research articles within the field of Literature, as Table 6 shows. Papers in literature often contain titles of literary works which serve as the subject of analysis in this field. In addition to titles, authors are often mentioned in research as well. This may complicate the translation of the research title. Examples 1 and 2 show how these titles may stand as untranslated concept errors. In example 1, the title was transliterated as “Bint al-Haram” by all three machine translation programs. Only the human translator translated the title as “the illegitimate girl”.

Table 6
Examples of various error types identified in the titles of articles in the field of “Literature.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
1	مظاهر فلسفة السُّخط في قصص "بنت الحرام" لجمعة شنب وأثرها في البعد الفني	The Aspects of Indignation's Philosophy in "The Illegitimate Girl" by Jum'ah Shanab and Its Effect on the Artistic Dimension	Manifestations of the philosophy of discontent in the stories of " Bint al-Haram " by Jumma Shanab and its impact on the artistic dimension	Philosophical Aspects of Anger in Jumaa Shnab's " Bint al-Haram " Stories and their Impact on the Artistic Dimension	Manifestations of the Philosophy of Anger in the Stories of " Bint al-Haram " by Jumaa Shanab and Its Impact on the Artistic Dimension
2	بنيّة الخطاب في النصّ الرّحليّ الموريسكيّ " رحلة أفوقاي (مختصر الشهاب إلى لقاء الأحباب نموذجًا)"	The Discourse Structure of Morisco Journey Narrative Avoqai Journey – A Brief Narrative of Al-Shihab ila Liqaa' Al-Ahbab) as a Model	The intention of the discourse in the Morisco travel text "The Journey of Afokai (Mukhtasar Al-Shihab to Meet the Loved Ones as an Example)"	"The Intent of Discourse in the Moorish Travel Narrative: A Case Study of 'The Journey of Afwāqāy (Mukhtasar al-Shihāb to the Meeting of Loved Ones)"	The Intention of Discourse in the Moorish Travelogue "Journey of Afuqa (Mukhtasar al-Shahab ila Liqa al-Ahbab as a Model)"

In example 2, the examined work has alternative titles, two of which are provided in the research title. The first title is Journey of Afokai, with Afoukay being an alternative spelling. The second title is “the comet to meeting loved ones”. When rendering the first title, no translation relied entirely on transliteration, and the word “journey” was translated directly. The problematic issue with the translations was the name “Afokai”. While transliteration is generally an adequate strategy for dealing with proper nouns and names, it can be classified as an untranslated concept error as the transliterations do not align with the official anglicization of the name. Among the four translations, only Google used the official spelling. Moreover, the translations of ChatGPT and Gemini show a greater degree of error as their transliterations do not resemble the source text and its pronunciation.

The second title shows errors beyond the transliteration of names. The source text indicates that the version of the story is the abridged or summarized one by including the word “مختصر” “summary”. This word indicates that the version of the examined text is a summary and not the full version and is not part of the original title. Yet it was transliterated by the three machine translators. Gemini transliterated the remainder of the title as well. Google and ChatGPT, however, resorted to actual translation in their renditions. Both programs rendered the sense of “to the meeting of loved ones”, but both transliterated the first word of the title, i.e., the word “مختصر”. It is possible that the programs misinterpreted the descriptor مختصر “summary” and the first word in the actual title “الشهاب” “comet”, considering them a given name of the author. This would explain why only one part of the words in parenthesis was translated while the other part was not.

A lack of understanding of which elements represent groups and units resulted in other translation errors. As shown in example 1, ChatGPT misassigned the units, i.e., word group, by linking the word “philosophy” with the word “aspects” instead of “anger.” This is a syntactic error that changes the meaning of the text. In example 2, the word “بنيّة” was mistranslated by all three machine translation tools as “intention” instead of the correct equivalent “structure,” which was rendered only by the human translator. Moreover, all the errors in the machine translations stemmed from the same mistake.

To clarify, the word “بنيّة” begins with the letter “b,” which is an integral part of the word “بنيّة” and does not function as a preposition as it often does in Arabic, but not in this context. In other words, the word “بنيّة” “structure,” is a single free morpheme that was misinterpreted by all three MT tools as a combination of the bound morpheme {b} which often functions as a preposition in Arabic and aligns well with the letters the free morpheme “نية” “intention,” leading consequently to a mistranslation.

The two examples also show the role of diction in the transfer of meaning. In example 1, the Arabic word “السُّخط” is used to express an intense sense of anger and injustice. Thus, the renditions of “anger” provided by ChatGPT and Gemini are acceptable but not ideal. Google’s translation of “discontent” is less suitable.

Example 2 shows that the word choice of human translators is not always ideal. In this example, the translator used the word journey to describe the type of text. The word “travel”, which was used by machine translations, is a better equivalent as it aligns with the name of the literary genre.

Table 7 shows examples of diction and mistranslation error types identified in the titles of articles in the field of “Literature.” Examples 3 and 4 show further mishandling of literary terminology. Example 3 refers to historical allusions, yet the human translators, together with Google and ChatGPT, render this literary device as sense, revelations, and implications, respectively. Example 4 references rhetoric ambiguity, but the machine translations refer to problems instead. Ambiguity reflects unclarity

that may be unintentional or intentional for artistic purposes. The word “problem” suggests a problematic issue that requires a solution. In example 5, “poetic imagery”, which refers to another literary device, is translated as “poetic image” by Google and Gemini. This shifts the meaning from a poetic device used to add aesthetics to a work to a picture or generalization of the poem. Example 6 shows that mistranslation errors are not limited to literary expressions. For instance, the word “polarities” was rendered by Google and Gemini as “intersections”. The meaning provided are antonyms, as the word “polarities” reflects distance and divergence while “intersections” reflects convergence. In example 7, the human and Google translations use the inaccurate sense for the polyseme in the source text and employ “legendary” in place of “mythical”. The human translation and ChatGPT’s renditions are also erroneous as they contain omissions where the word “international” that appeared in the source text is missing in their translations.

Table 7

Examples of **diction and mistranslation** error types identified in the titles of articles in the field of “Literature.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
3	إحياءات التاريخ في المجموعة الشعرية كوتشيرتو - لأدونيس التناص - التاريخ - الاحتلال	A Sense of History Emanating from Adonis’ Poetry Collection, “Concerto for Jerusalem” Intertextuality - History – Occupation	Revelations of history in the poetry collection Adonis’ Jerusalem Concerto. Intertextuality - History - Occupation	Historical Implications in Adonis’ Poetic Collection “Concerto of Jerusalem” - Intertextuality, History, and Occupation	The Allusions of History in the Poetry Collection “Concerto of Jerusalem” by Adonis: Intertextuality - History – Occupation
4	مصطلح المشاكلة البلاغية دراسة تداولية لسانية على نماذج شعرية مختارة	Rhetoric Ambiguity: A Pragmatic Study on Selected Poetic Excerpts	The term rhetorical problem : a linguistic pragmatic study on selected poetic models	Rhetorical Problematicization : A Discourse-Linguistic // Study on Selected Poetic Models .	The term “ problematics ” in rhetoric: A linguistic pragmatic study on selected poetic models
5	الصورة الشعرية ومصادرها الفعلية من منظور فينومينولوجي: مقاربة نظرية وتطبيقات على نصوص لمحمود درويش	The Poetic Image and its Actual Sources from a Phenomenological Perspective: Theoretical Approach and Applications on Texts by Mahmoud Darwish	The poetic image and its actual sources from a phenomenological perspective: a theoretical approach and applications to texts by Mahmoud Darwish	Poetic Imagery and its Actual Sources from a Phenomenological Perspective: A Theoretical Approach with Applications on Mahmoud Darwish’s Texts.	The poetic image and its actual sources from a phenomenological perspective: A theoretical approach and applications on the texts of Mahmoud Darwish
6	أنماط المكان وتقاطباته في روايات جمال أبو حمدان	Models of Spatial Polarities in Jamal Abu Hamdan’s Novels	Patterns of place and its intersections in Jamal Abu Hamdan’s novels	Place Patterns and Convergences in the Novels of Jamal Abu Hamdan	Spatial Patterns and Intersections in the Novels of Jamal Abu Hamdan
7	الرؤيا الأسطورية في رواية “الخيميائي” للروائي العالمي بولو كويلو	The Legendary Vision in the Novel “The Alchemist” written by Paulo Coelho	The legendary vision in the novel “The Alchemist” by the international novelist Paulo Coelho	Mythical Vision in Paulo Coelho’s Novel “The Alchemist.”	The mythical vision in the novel “The Alchemist” by the world-renowned novelist Paulo Coelho

To summarize, the translation of literary terms in the examples cited in Table 7 shows a clear disparity in the renditions provided by both the human translation and that of the AI tools. Coming up with the right target language equivalent in the translation of literary texts and titles is not easily attainable. This is due to the figurative, symbolic, and poetic nature of the language of literature, where sense and meaning may be mystified, even with human translators. In other words, cases of mistranslation, omission, erroneous transliteration, morpho-syntactic errors, diction, and faulty assignment of units and word groups were a common and recurrent feature in machine-translation output. Upon tallying the number of errors, the results showed Google Translate with 4 errors, ChatGPT with 7, and Gemini with 3. With only two errors for human translation, the results lend support to human translators in handling literary texts.

4.2.3. Linguistics

This section examines titles of Arabic research articles within the field of Linguistics, as Table 8 shows.

Table 8

Examples of various error types identified in the titles of articles in the field of “Linguistics.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
1	النظائر المخادعة الثامنة في اللغتين العربية والتركية	Complete False Cognates in Arabic and Turkish Languages	Completely deceptive counterparts in Arabic and Turkish	Perfect Deceptive Equivalents in Arabic and Turkish Languages.	Full-fledged homonyms in Arabic and Turkish
2	دلالات العطف في سورة الكهف: (أو، الواو، الفاء نموذجاً)	Conjunctions letters on Soura (AL Kahf) of the Holy Quran †the conjunction letters was: (Wa †Fa †Aw)	Connotations of conjunction in Surat Al-Kahf: (or, waw, fa’ as an example)	Implications of Affection in Surah Al-Kahf: (Or, Waw, Fa as Examples)	The Meanings of Conjunction in Surat Al-Kahf: (Or, Waw, Fa’ as a Model)

Example 1 again highlights the shortcomings of machine translation when dealing with specialized terminology. This part of the current study focuses on linguistics and specifically discusses false cognates. Cognate words are words in different languages that are similar in both sense and sound. This similarity is a direct result of the words sharing their source. False

cognates are word pairs which appear to be cognates due to phonetic and meaning similarity but do not share their etymology. The similarity in these cases is incidental.

Only the human translation rendered the first example in Table 3 correctly, while the machine translations committed errors in the rendition of its two elements. Regarding the word “false”, Google and ChatGPT provided a literal equivalent of the term in the source text. Thus, the translations include the word ‘deceptive’. This translation does not take the collocations and linguistic studies background into consideration. Therefore, this rendition, which is typically an exact literal equivalent of the Arabic word, becomes a less suitable option. However, Gemini’s rendition is the most erroneous as it omits the word entirely. The machine translations also failed to render the word cognates. Gemini also omitted the word “languages”, resulting in two omission errors in the rendition of a single title. The term “counterparts” provided by Google suggests equivalence and agreement. Typically, when discussing languages, “counterpart” can be used to describe the other language’s version of the same word, i.e., its translation. While cognates fall under counterparts, the two words are interchangeable since not all counterparts are cognates. ChatGPT uses the term equivalents, which is more limiting than the term “counterparts” as it can only imply similarity in meaning, while “counterpart” may suggest similarity in either form or sense. The term “homonym,” which was provided by Gemini, fails to invoke a relation between two languages as it only describes similar-sounding words or similarly spelled words within one language. Furthermore, a defining characteristic of homonyms is that they have different meanings.

In example 2, the research examines conjunctions. ChatGPT’s rendition does not convey this, as it mistranslates the term based on its homonym. The listing of the Arabic conjunctions showcases linguistic gaps that lead to untranslatability. As a result, the conjunctions were mostly left as untranslated concepts. Although the conjunctions may not have been translated accurately, it was possible for their sense to be transferred at least partially as conjunctions that serve the same functions do exist in English. Instead, only the conjunction “or” was translated. The conjunction “and” could have been translated directly, but since it is a letter that was spelled out in Arabic, it was transliterated. “F” is the only conjunction that does not have an exact equivalent, as it contains some nuance that expresses immediate succession and differs from what is used to describe longer intervals. The human translation, however, handled all these terms successfully. This could have been motivated by a desire to highlight Arabic grammar, which is the issue at hand.

Table 9 shows examples of untranslated concepts and mistranslation error types identified in the titles of articles in the field of “Linguistics.”

Table 9

Examples of untranslated concepts and mistranslation error types identified in the titles of articles in the field of “Linguistics.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
3	علة حذف فاء الفعل المضارع المثال الواوي	The Reason for Omission of the First Letter of Verb (mitāl alwaw ī)	The reason for deleting the present tense verb fa is the waw example	The Omission of the "Fa" in the Present Tense Verbs: The Example of the Waawi Verbs	The Reason for Deleting the "fa" of the Present Tense of the Waw-Lettered Perfect Verb
4	الفعل كان مؤكداً	Verb “kana	The action was certain	The Verb Was Affirmed	The verb “kana” as a confirmative
5	اسم النسب في القرآن الكريم دراسة نحوية دلالية	The Attributed-noun in The Holy Quran A semantic and Grammatical Study	The name of lineage in the Holy Qur’an: a grammatical and semantic study	Surnames in the Quran: A Semantic Syntactic Study	Names in the Quran: A Grammatical and Semantic Study

Examples 3 and 4 are two more research studies that focus on specific elements of Arabic grammar. Example 3 focuses on the linguistics of the Arabic language. It examines an area of Arabic morphology unique to the language. Moreover, the subject falls on a level of morphology that requires scholarly linguistic knowledge and not just fluency in the language. To understand this title, first, one must understand how word forms are represented in Arabic studies. Arabic is a derivational language based on “roots” or stem words. Most roots are three-lettered. To simplify the study of linguistics, the three letters ف ع ل (fa ‘a la) which are the root for the verb “did” are used as representatives. This root is used as a placeholder when describing morphological processes. These letters can be substituted with any three-letter root to which the same rules and processes are applied. Thus, the “fa” in the title refers to the first letter of a root. This letter may be any letter from the Arabic alphabet, and the word does not need to contain a “fa” sound. The research focuses on “waw” words, or words that begin with the letter “waw” و. Thus, the “fa”, which symbolizes the first letter of the root, represents the “waw” in this case. The paper thus examines the morphological process that turns a root into a present tense verb, which, in the case of some words, results in the omission of the “fa” or first root sound with a focus on the cases where this first letter is the vowel “waw.”

Example 4 studies one of the functions of the Arabic auxiliary verb “kan” (was), which specifically functions as a confirmative device. Gemini leaves the verb as an untranslated concept to highlight the word at hand. This provides a rendition that reflects the sense of the original title. The human translation transliterates the verb in a similar manner. However, this rendition does not represent the title as it omits other key elements. The function that is being examined was not mentioned. Google and ChatGPT translate the Arabic verb into the English “was”. Yet the renditions are unacceptable due to how the descriptor was handled. Instead of translating “was “as an affirmative”, it was translated as “affirmed”. In Arabic, these terms consist of the same letters but differ in diacritics (inflectional case endings). The word with a “kasrah” on the /k/ sound would describe the noun as a subject, while a “fathah” on the same letter would describe it as an object. Aside from diacritics, the two can be

distinguished by context. These translations mistakenly render the term as if “kan” was an object. Additionally, these translations treat the word “was” as if it is a functional verb within a title and not a concept which is being referred to. Furthermore, Google failed to recognize that the word “الفعل” referred to “verb” as a part of speech and not as an action, resulting in another mistranslation error.

Example 5 contains another area of Arabic linguistics that the machine translators failed to decode and recode. In Arabic grammar, there are ways to attribute nouns to a larger concept, such as a place or family. These words now describe someone or something that belongs in or to the original word. Such words are referred to as “اسم النسب”, which can be translated as nouns of attribution. The human translator rendered this term as “attributed nouns”, a phrase that can describe the concept at hand. This indicates that even in cases where the linguistic concept does not exist in the studies of the source language, a meaningful rendition is possible. The machine translations failed to transfer this sense by mistranslating both components of the source text phrase. The first error, which has most likely led to the second, was mistranslating the word اسم as the name instead of a noun. While the two are homonyms and share a root, only one is connected to the examined grammar, while the other is used to refer to persons. The term “نسب” can also be used to describe lineage, as it attributes a person to a family. Therefore, the phrase “اسم النسب” may mean family name. Accordingly, Google and ChatGPT render the phrase as “name of lineage” and “surname”. Gemini, on the other hand, misinterprets the phrase in the same manner but omits the descriptor resulting in the rendition “names”. This rendition widens the concept and thus becomes an unacceptable translation even in the case where “lineage name” was the actual source text.

To recap, the linguistic jargon, with its specialised terms in morphology and syntax, posed a challenge to AI-powered translation tools. Errors were detected at the two levels of form and content. The AI tools yielded erroneous renditions such as omission, diction, and mistranslation. Language problems included the wrong use of cognates, conjunctions, morphemes, and auxiliary verbs. In numbers, the results showed the following tally: Google Translate with 4 errors, Chat GPT with 3, and Gemini with 7. On the contrary, human translation reflected more accuracy and idiomaticity, without errors.

4.2.4. Education

This section examines titles of Arabic research articles within the field of Education, as Table 10 shows.

Table 10
Examples of various error types identified in the titles of articles in the field of “Education.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
1	أثر استخدام الألعاب الرقمية في تنمية مهارات التفكير الإبداعي وحل المشكلات لدى ذوي صعوبات الرياضيات بدولة الكويت	The Impact of the Use of Digital Games on the Development of Creative Thinking and Problem-solving Skills with Mathematics Disabilities in the State of Kuwait	The impact of using digital games on developing creative thinking and problem-solving skills among people with mathematics difficulties in the State of Kuwait	The Impact of Using Digital Games on Developing Creative Thinking Skills and Problem-Solving Abilities in Students with Mathematical Difficulties in Kuwait	The Effect of Using Digital Games in Developing Creative Thinking and Problem-Solving Skills among Children with Mathematical Difficulties in Kuwait
2	أثر استخدام استراتيجية التعلم المعكوس في تنمية مهارات التفكير التأملي	The Effect of Using Flipped Learning Strategy in Developing Reflective Thinking Skills	The effect of using the flipped learning strategy in developing reflective thinking skills	The Impact of Using the Inverted Learning Strategy on the Development of Contemplative Thinking Skills	The Effect of Using the Flipped Learning Strategy on Developing Reflective Thinking Skills
3	علاقة العزيمة والتصوّر العقلي بالتحصيل الأكاديمي لدى الطلبة ذوي الإعاقة وأقرانهم في الجامعات الأردنية	The Relationship between Grit, Mindset and Academic Achievement among Students with Disabilities and their Peers in the Jordanian Universities	The relationship of determination and mental imagery to academic achievement among students with disabilities and their peers in Jordanian universities	The Relationship between Determination, Mental Imagery, and Academic Achievement among Students with Disabilities and their Peers in Jordanian Universities	The Relationship between Determination and Mental Imagery with Achievement among Disabled
4	قيم الانتماء الوطني في كتب التربية الموسيقية لمرحلة التعليم الأساسي العليا في الأردن "دراسة تحليلية"	Values of National Belongingness in the Music Education Textbooks for the Higher Basic Education Stage in Jordan: An Analytical Study	Values of national belonging in music education books for the upper basic education stage in Jordan, “An Analytical Study”	The Values of National Belonging in Music Education Textbooks for Upper Basic Education in Jordan: An Analytical Study	Values of National Belonging in Music Education Textbooks for the Upper Primary Education Stage in Jordan: An Analytical Study

Example 1 demonstrates how context can affect the acceptability of a rendition. The area of this study is education, with a focus on those who face difficulty with mathematics. In Arabic, the term “learning difficulties” can be used to refer to “learning disabilities”. However, in English, the two terms describe different concepts. A learning disability relates to cognitive disabilities and is diagnosable, but learning difficulties are related to difficulties in certain skills and can be referred to as learning disorders. The focus of the paper in example 1 above is students who face challenges with mathematics; there is no confirmation whether the students have either disorders or disabilities. Therefore, to describe them as people challenged by math is the best equivalent. “Mathematics difficulties” is an acceptable rendition despite the more limited connotations. This was the rendition provided by the three machine translation tools. The human translator rendered the term as disabilities even though

that does not describe the study group. Since the title does not reflect the real nature of the study, this rendition is a definite case of mistranslation.

The human translator also omitted the phrase “those with” (people with). This is a key element of the title, and its removal results in an incoherent sentence as the subject is missing. Google renders this phrase as “people with”, while the two other programs resort to context-based addition. ChatGPT renders the terms as “students with”, and Gemini renders it as “children with”. The terms students and children do not exist in the source text but are adequate descriptors of the study group and are, therefore, acceptable.

Gemini also omits a word in its rendition; however, unlike human translation, the omitted term was not necessary to convey the meaning. In this case, it was the word ‘State’ that was used to describe ‘Kuwait’. But since ‘Kuwait’ is the name of a state, this inclusion may be redundant.

Example 2 also showcases the role of collocations and specialized terminology in transferring meaning. This can be seen in the differences in ChatGPT’s rendition. ChatGPT handled two terms differently. The first is the use of “inverse learning” in place of “flipped learning”, and the second was the use of “contemplative thinking” in place of “reflective thinking”. The first term is an acceptable alternative, but it is less commonly used. For the second term, although the two seem like synonyms, the two collocations have different senses. Contemplative thinking involves building on ideas, while reflective thinking involves revisiting and reanalyzing ideas. Hence, it is only “reflective thinking” that is an equivalent of the source text and the research paper.

In example 3, the human translation contains a similar error where a concept specific to the field is misinterpreted as a generic term. The paper focuses on mental imagery that involves creating mental representations without visual stimulation. What is intended here is not a metaphorical picturing of results to create motivation. Therefore, the rendition of “mindset” is nonequivalent.

Example 4 shows that using a wordy translation does not always result in mistranslation errors. For instance, the term “national belonging” is an acceptable rendition of the source text. However, the alternative term “nationalism” may be considered a more common jargon in this context.

To sum up, the examples given in Table 10 show that cases of mistranslation and omission were present in the different modes of translation, both human and machine alike. A case in point is the human translation of "صعوبات الرياضيات" as “mathematics disabilities”, and the omission of the phrase “لدى ذوي” and the wrong use of “mindset” as “التصور العقلي”. ChatGPT used “inverted learning” instead of “flipped learning” and mistranslated “التفكير التأملي” as “contemplative thinking” instead of “reflective thinking”. Also, Gemini’s omission of the word “State” in translating “the State of Kuwait” was noted. The errors included mistranslations due to the wrong use of diction, omissions which affected meaning, collocational usage, metaphorical usage, and syntactic problems. Upon tallying the number of errors for each tool, the results showed Google Translate without errors, ChatGPT with 2, and Gemini with 1. The human translation was spotted with two errors.

4.2.5. Political Science

This section examines titles of Arabic research articles within the field of Political Science, as Table 11 shows.

Table 11

Examples of various error types identified in the titles of articles in the field of “Political Science.”

No	Source Text	Human	Google Translate	ChatGPT	Gemini
1	أضواء على تطور الفكر السياسي عند الإمامية في العصر المملوكي: الفقيه محمد بن مكّي الجزيني (نموذجاً ٥٧٨٦هـ/١٣٨٤م)	Exploring the Evolution of the Political Thought of Imamate Shiism during the Mamluk Age: Faqih Muhammad Jam- aluddin al-Makki al-Amili in 786 AH (1384 AD) as an Example	Highlights on the development of political thought among the Imamis in the Mamluk era: the jurist Muhammad bin Makki al-Juzaini as a model (786 AH/1384 AD)	Insights into the Development of Political Thought among the Imamis in the Mamluk Era: The Scholar Mohammed bin Maki Al-Jazini as a Model (786 AH/1384 CE)	Lights on the Development of Political Thought among the Imamiyyah in the Mamluk Era: The Jurist Muhammad ibn Makki al-Jazini as a Model (786 AH/1384 AD)
2	البراغماتية الإيرانية في ظل ولاية الفقيه	Iranian Pragmatism under the Rule of Al-Faqih	Iranian pragmatism under the Guardianship of the Jurist	Iranian Pragmatism in the Shadow of the Guardianship of the Jurist	Iranian pragmatism under the rule of the jurist
3	"الموازنة الهجينة" في منطقة الهندو-باسفيك: "أوكوس" في مواجهة "الحزام والطريق"	Hybrid Balancing” in the Indo-Pacific Region: “AUKUS” vs. “Belt and Road”	“Hybrid budgeting” in the Indo-Pacific region: “OKOS” versus “Belt and Road”	"Hybrid Diplomacy" in the Indo-Pacific Region: " Or-cus " in the Face of "Belt and Road"	"Hybrid balancing" in the Indo-Pacific region: "AU-KUS" versus "Belt and Road"

Example 1 contains elements related to history, culture, and religion as it focuses on the politics of a specific area in a certain period where specialized technical jargon is used. These elements proved to be problematic as all translations contained errors. The first example in Table 11 is the title of a paper that studies Twelvers, who are a branch of Shia Muslims. When rendering the term “الإمامية” the human translation provided an acceptable rendition, but it is still not specific enough to match with the widely acknowledged title, “Twelver”. This rendition followed the wording of the source title, a strategy that led to deeper

errors in the machine translation rendering. Gemini, for instance, left the term as an untranslated concept by employing transliteration. Google and ChatGPT also retained the makeup of the source text but adapted the derivation to follow the norms of the target language. This can be classified as lexical creation. Since the source text term had a target language equivalent, this is a clear translation error.

Example 2 contains the word “ولاية” which means guardianship. This was the term used by Google and ChatGPT, while Gemini and the human translation used the term “rule”. While the term rule is acceptable as it generally conveys the sense at hand, guardianship is a more suited equivalent. Not only is this due to the literalness of the rendition, but also because the term was used in the title of an era known as “the guardian of the Islamic jurist.” This exact wording did not occur in any of the translations. When rendering the term “الفتية,” the human translator left it as an untranslated concept, but the machine translation followed a direct approach, rendering the term “jurist”. While this rendition is technically correct, it results in the loss of some meaning as it becomes too generic. These renditions show that the source text phrase was dealt with as single words rather than a whole unit.

Example 3 highlights two difficulties faced in the translation of specialized terminology as the title involves geopolitics. The first is the issue of polysemy and the employment of the correct equivalent. The Arabic term “موازنة” derived from وزن (weigh), can refer to balancing or budgeting based on context. Since the subject of the research is not financial in nature, it is clear that the word “budgeting” is not the favoured option in. Despite this, only the human translator and Gemini used the term “balancing,” while Google Translate mistranslated the term as “budgeting”, and ChatGPT ignored the term altogether. This indicates that machines have lower capabilities than humans in understanding context. Furthermore, machine translation tools struggled with providing the right equivalent of the specialized acronym AUKUS, which was rendered as “أوكوس.” This is a transliterated version of the English acronym AUKUS (Australia, United Kingdom, United States), with each English letter being represented by an Arabic one. The Arabic transliteration maintains the components of the English acronym instead of functioning as an acronym of the words translated forms. Despite the borrowed nature of the term, two of the three AI-powered programs failed to return it to its original form. While Gemini correctly converted the acronym back into “AUKUS,” Google and ChatGPT took a phonetic approach, thus creating a meaningless phrase. Additionally, since there is no standardized approach to the English transliteration of Arabic words, each program provided a different set of letters.

To recap, in all the examples of Political Science in Table 11, there were discrepancies in the translation, which varied from the inaccurate rendering of the term “الإمامية” through mistranslation, omission, diction, wording, polysemy, and transliteration. Add to this attempting lexical creation through derivational morphology in the case of Google Translate and Chat GPT. As for the human translation, it provided a relatively more acceptable rendering of the term. A similar discrepancy was reflected in handling the term “ولاية الفقيه” where two translations were rendered representing the human translation and Gemini’s as “Rule of Al-Faqih” and “Rule of the jurist” on the one hand, and Google Translate and ChatGPT’s use of “Guardianship of the Jurist”, on the other. In addition, both Google Translate and ChatGPT failed to render the acronym “AUKUS” correctly. A tally of the errors by number revealed the following: Google Translate 5, ChatGPT 5, and Gemini 3. As for human translation attempts, they were relatively more acceptable with only two errors.

5. Conclusions and Recommendations

The research findings of the current study have clearly shown that the meaning of a research title can be understood through the sense derived from its words and the accuracy of its syntactic structures. When the title is transferred to the target language, the rendition must show accuracy on the sense/meaning level and on the morpho-syntactic level as well. Most provided renditions had issues with either the sense or syntax which rendered the translations incorrect. This incorrectness can be attributed to the existence of translation errors. Although the errors were common, it was rare for a rendition to be incorrect on both the semantic and syntactic scales concurrently.

The most common error was that of mistranslation. Many terms with multiple senses were translated through an inaccurate equivalent. In all the cases where the mistranslated words were key sense words, the translations reflected an incorrect meaning. With less significant words, the meaning could sometimes remain intact. The use of the wrong sense reflects a lack of contextual adequacy. Human translators can be more knowledgeable on scientific matters or may resort to other sources including the research paper itself for clarification. The context utilized by machine translation is limited to the elements of the title, reducing access to a wider range of information. Therefore, while the use of corpora may improve the translations through collocations in some cases, the lack of external context may negatively affect results in others.

The morpho-syntactic errors committed by the three AI programs can be classified into two types. The first is strictly linguistic overt errors that represent language errors regardless of connection to the source text. The second type is linked directly to the source text in the form of mismatches. This can be further categorized into differences in arrangement and order as well as semantic and syntactic differences. The first involves a misalignment of elements and does not affect meaning. The second, however, results in a misrepresentation of the title’s sense. Like the issue with word equivalents, this is another indication of machine translation’s weak access to context.

Since machine translation works strictly with the elements of the title, errors such as addition and omission were rare and limited. The final type of error was “untranslated concepts” which was found in some acronyms, neologisms, and newer terms.

The above conclusions have most clearly shown that the current study has most adequately addressed and responded to the two research questions posed earlier in the study, namely (1) How do the translation accuracies of AI tools (Google Translate, ChatGPT, and Gemini) for Arabic research titles compare to human translation accuracy? And (2) What common mistakes do AI tools make when translating Humanities/Social Sciences research titles from Arabic to English? In addition, answers to these questions were linked to the results of previous research and literature on the effectiveness of AI-powered machine translation tools.

The results of this study will prove significant for a number of reasons. First, it will have noticeable effects on the application of AI technologies for the English translation of multidisciplinary Arabic research titles. Second, the results will assist researchers in determining which AI tools are most accurate for translating their research titles and shaping the guidelines they should follow when utilizing AI tools to translate their research titles. This will contribute to the accuracy, clarity, and fluency of research titles translated by AI. Third, the results shed light on the typical mistakes AI systems make when translating research titles. By using this data, MT systems driven by AI can perform better and become more dependable in translating research titles. The paper will also help stakeholders in the translation field to decide whether AI translation tools are helpful or simply an act of plagiarism in disguise.

The analysis also indicated that the type of errors affects the degree of understandability. For instance, mistranslation and diction errors result in the most significant shifts in sense. Thus, although the human translations contained the highest number of errors in general, the sense of most of their renditions was acceptable. Syntactic errors had a significantly smaller effect on sense but instead influenced the naturalness of the language and reduced its idiomaticity. Moreover, addition errors mostly resulted in redundant but otherwise acceptable renditions. Untranslated concepts also resulted in unnaturalness and loss of sense. However, unlike mistranslation errors, they generally do not cause shifts into new senses. Furthermore, the relevance and type of term where the error occurred also affected the level of acceptability. For instance, the mistranslation or omission of keywords was far less acceptable than the same errors committed on sense and function words. This is due to the meaning each word carries and its relation to the topic of the study. The study will be useful for scholars from the Middle East and North Africa who must translate their research titles into English. The same applies to MENA funding organizations and policymakers who are keen to encourage the use of AI technologies for translating research titles and for developers of MT tools driven by AI. It is worth noting, however that in spite of the weaknesses and drawbacks evidenced in the performance of AI-powered translation tools assessed in the current study, the overall product is a developed version of what it looked like two decades ago when it was lacking at the syntactic, semantic, and contextual level to bluntly jeopardize meaning, acceptability, and idiomaticity. The fact that the current study is limited to investigating a specific domain of titles, namely humanities and social sciences, further research is needed to include other fields and disciplines in science and technology, as well as medical and allied health sciences.

To conclude, the findings of the current study show that AI translation tools are not advanced enough to fully replace human translators in all fields and disciplines. This is in light of the fact that machine translation technology has been around decades, without reaching the level of accuracy achieved by humans. However, it is important to note that the type and purpose of the task under investigation are key factors in determining which translation tool to choose. For example, whereas AI-powered tools can be fast and effective in translating every day quick messages or emails, the same cannot apply in handling highly specialized technical content in the field of law, medicine, or science and technology. Such content may pose a challenge to AI-powered engines. Consequently, only highly qualified human translators can provide the optimum degree of accuracy and quality translation. An alternative procedure we recommend here is to subject the AI translation product to a process of robust and rigorous scrutiny and meticulous revision by well-versed human translation experts who will be entrusted with revising the machine translation product thoroughly, simulating the procedures followed in carrying out large-scale translation projects. Such 'human-assisted machine translation' (HAMT) can ensure a quality translation based on informed human decisions where cultural connotations, specialized technical jargon, lexical choices, and language nuances in form and content are maintained.

Funding Statement

This research received grant no. (94/2023) from the Arab Observatory for Translation (an affiliate of ALECSO), which is supported by the Literature, Publishing & Translation Commission in Saudi Arabia.

References

- Abdelaal, N. M., & Alazzawie, A. (2020). Machine Translation: The Case of Arabic-English Translation of News Texts. *Theory & Practice in Language Studies*, 10(4). doi:<http://dx.doi.org/10.17507/tpls.1004.09>
- AKASHEH, W. M., HAIDER, A. S., AL-SAIDEEN, B., & SAHARI, Y. (2024). Artificial intelligence-generated Arabic subtitles: insights from Veed. io's automatic speech recognition system of Jordanian Arabic. *Texts Livre*, 17, e46952-e46952. doi:10.1590/1983-3652.2024.46952.

- Al-Salman, S., & Haider, A. S. (2021). The Representation of Covid-19 and China in Reuters' and Xinhua's Headlines. *Search (Malaysia)*, 13(1), 93-110. doi:<https://fslmjournals.taylors.edu.my/wp-content/uploads/SEARCH/SEARCH-2021-13-1/SEARCH-2021-P8-13-1.pdf>
- Almahasees, Z. (2018). Assessment of Google and Microsoft Bing translation of journalistic texts. *International Journal of Languages, Literature and Linguistics*, 4(3), 231-235.
- Almahasees, Z. (2021). *Analysing English-Arabic Machine Translation: Google Translate, Microsoft Translator and Sakhr*. London: Routledge.
- Almahasees, Z., & Jaccomard, H. (2020). Facebook Translation Service (FTS) Usage among Jordanians during COVID-19 Lockdown. *Advances in Science, Technology and Engineering Systems Journal*, 5(6), 514-519. doi:10.25046/aj050661
- Alrousan, F., & Haider, A. S. (2022). Dubbing television advertisements across cultures and languages: A case study of English and Arabic. *Language Value*, 15(2), 54-80. doi:<https://doi.org/10.6035/languagev.6922>
- Aydin, Ö. (2023). Google Bard generated literature review: metaverse. *Journal of AI*, 7(1), 1-14.
- Bin Dahmash, N. (2020). 'I Can't Live Without Google Translate': A Close Look at the Use of Google Translate App by Second Language Learners in Saudi Arabia. *Arab World English Journal (AWEJ) Volume*, 11(3), 226-240
- Bouguesmia, M. T. (2020). Using AI in Translation, a Technological Leap, or a Translator's Nightmare. *ALTRALANG Journal*, 2(02), 78-102.
- Bowman, D., & Kinnan, S. (2018). Creating effective titles for your scientific publications. *VideoGIE*, 3(9), 260-261.
- De La Cruz-Cabanillas, I., & Tejedor-Martínez, C. (2016). The Error Analysis Approach for the Assessment of Automatic Translation. *Lingwistyka Stosowana/ Applied Linguistics/ Angewandte Linguistik*(16), 1-9. doi:https://www.researchgate.net/publication/287198349_The_Error_Analysis_Approach_for_the_Assessment_of_Automatic_Translation
- ElShiekh, A. A. A. (2012). Google translate service: transfer of meaning, distortion or simply a new creation? An investigation into the translation process & problems at google. *English Language and Literature Studies*, 2(1), 56.
- George, A. S., & George, A. H. (2023). A review of ChatGPT AI's impact on several business sectors. *Partners Universal International Innovation Journal*, 1(1), 9-23. doi:DOI:10.5281/zenodo.7644359
- Haider, A. S., & Hussein, R. F. (2020). Analysing headlines as a way of downsizing news corpora: Evidence from an Arabic-English comparable corpus of newspaper articles. *Digital Scholarship in the Humanities*, 35(4), 826-844. doi:<https://doi.org/10.1093/lc/fqz074>
- Hartley, T. (2009). Technology and translation. In J. Munday (Ed.), *The Routledge companion to translation studies* (pp. 120-141). London and New York: Routledge.
- Henrickson, L., & Meroño-Peñuela, A. (2023). Prompting meaning: a hermeneutic approach to optimising prompt engineering with ChatGPT. *AI SOCIETY*, 1-16.
- Hussein, R. F., Haider, A. S., & Al-Sayyed, S. (2021). A Corpus-Driven Study of Terms Used to Refer to Articles and Methods in Research Abstracts in the Fields of Economics, Education, English Literature, Nursing, and Political Science. *Journal of Educational Social Research*, 11(3), 119-131. doi:<https://doi.org/10.36941/jesr-2021-0056>
- Hutchins, J. (2005). Current commercial machine translation systems and computer-based translation tools: system types and their uses. *International journal of translation*, 17(1-2), 5-38.
- Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, 31(1), 60-86.
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv preprint arXiv:2301.08745*.
- Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy.
- Koponen, M. (2010). *Assessing machine translation quality with error analysis*. Paper presented at the Electronic proceeding of the KaTu symposium on translation and interpreting studies.
- Koponen, M., & Salmi, L. (2015). On the correctness of machine translation: A machine translation post-editing task. *The Journal of Specialised Translation*, 23(23), 118-136.
- Ledesma, I. B. (2001). *Errores y aprendizaje*. Paper presented at the Interferencias, cruces y errores.
- Popović, M., & Ney, H. (2011). Towards automatic error analysis of machine translation output. *Computational Linguistics*, 37(4), 657-688 doi:https://doi.org/10.1162/COLI_a_00072
- Putri, A. T., & Setiajidi, H. H. (2021). *Instagram Translate and Human Translation In The English Captions Of Jokowi's Account: An Analysis Of Koponen's Error Category*. Paper presented at the English Language and Literature International Conference (ELLiC) Proceedings.
- Rusadi, A. M., & Setiajidi, H. H. (2023). *Evaluating the Accuracy of Google Translate and Chatgpt In Translating Windows 11 Education Installation Gui Texts to Indonesian: An Application of Koponen's Error Category*. Paper presented at the English Language and Literature International Conference (ELLiC) Proceedings.
- Santos Gargallo, I. (1993). *Análisis contrastivo, análisis de errores e interlengua en el marco de la lingüística contrastiva*. Madrid.
- Son, J., & Kim, B. (2023). Translation Performance from the User's Perspective of Large Language Models and Neural Machine Translation Systems. *Information*, 14(10), 574.
- Vázquez, G. (1999). *Errores?, sin falta!* Madrid.
- Yilmaz, E. D., Naumovska, I., & Aggarwal, V. A. (2023). AI-Driven Labor Substitution: Evidence from Google Translate and ChatGPT. *SSRN*. doi:<https://ssrn.com/abstract=4400516>

Yves, G. (2019). Impact of technology on Translation and Translation Studies. *Russian Journal of Linguistics*, 23(2), 344-361.



© 2024 by the authors; licensee Growing Science, Canada. This is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).