A Multidimensional Analysis of Human and ChatGPT-Generated English Translations of Arabic Film

Sadia Ali[1], & Naeem Afzal[1]

[1] Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia

Correspondence: Sadia Ali, Department of English Language and Literature, College of Science and Humanities, Prince Sattam bin Abdulaziz University, Al-Kharj, Kingdom of Saudi Arabia. E-mail: sad.ali@psau.edu.sa

Received: March 15, 2024 Accepted: June 7, 2024 Online Published: June 19, 2024

doi:10.5430/wjel.v14n4p632 URL: https://doi.org/10.5430/wjel.v14n4p632

Abstract

This study analyzed lexico-grammatical variations between two text types: human-written and machine-generated, using Biber's multidimensional analysis. It explores the effectiveness and limitations of AI-driven translation systems in maintaining the quality of film translations. It aims to add to the current discussion on the impact of AI in the field of translation. The research methodology involves selecting films from the Middle East and collecting their translations, both human-written and generated by ChatGPT. Biber's multidimensional analysis framework analyses the translations across dimensions such as involved versus informational discourse, narrative versus non-narrative concerns, explicit versus situation-dependent, overt expression of argumentation/ persuasion, and abstract versus non-abstract discourse. The findings of the analysis reveal similarities and differences between human and ChatGPT translations. Human translations are more involved, situation-dependent, argumentative, non-abstract, and less non-narrative than the translations generated by AI. However, further improvements and refinements in AI translation models could help bridge the gap between human and AI translations. The results gained from this comparative analysis offer insight into improving AI-driven translation systems, leading to more effective cross-cultural communication through film. This research will potentially contribute to the advancement of the field of translation studies by bridging the gap between human and AI translations. It provides valuable implications for the future development of AI technologies in film translation.

Keywords: AI-driven translation; comparative analysis; film English translations; human translation; multidimensional analysis

1. Introduction

Film translation helps the global audience to access movies from diverse linguistic backgrounds. Previously, it was the role of a human translator to provide an audience with an accurate interpretation of a film with the help of language and cultural background. New developments in AI have created translation systems with the help of AI, such as ChatGPT, which can be seen as potential competitors to human translators. Many researchers (Borger et al., 2023; Ray, 2023; Hadi, 2023; Roumeliotis & Tselikas, 2023) have explored the application of AI-based translation tools, specifically ChatGPT, in different domains. However, there is limited research on the efficiency and disadvantages of AI-driven translation systems that employ multidimensional analysis in film translations.

This paper involves a comparative analysis of the film translations done by human translators and ChatGPT to determine the efficiency and drawbacks of AI translation systems in preserving the quality of Arabic film translations. In this paper, we aim to investigate the features of translation made by humans and Chatbots to identify the similarities and differences in their translation techniques and styles based on the framework developed by Biber. This framework considers aspects such as involved versus informational discourse, narrative versus non-narrative concerns, explicit versus situation-dependent expressions, overt expression of argumentation/persuasion, and abstract versus non-abstract discourse.

Five films addressing social issues, history, gender, survival, and conflict resolution from Middle Eastern countries such as Lebanon, Saudi Arabia, and Jordan were chosen. The translations were gathered for data analysis with reference to Biber's five dimensions, including humans and ChatGPT. Therefore, by analyzing these translations across the mentioned dimensions, the present work intends to contribute to understanding AI translation systems' limitations and potential using lexico-grammatical aspects.

The results of this analysis will help to enrich the discussion of the role of AI in the field of translation. This paper shows that accurate and culturally appropriate translations require human creativity and cultural knowledge, especially in film translations, which are essential cultural products with complex subtleties. Thus, acknowledging and appreciating these human factors in translation makes it possible to guarantee the preservation of the meaning across translations and improve communication between representatives of different cultures. It stresses that although AI-driven systems such as ChatGPT can produce accurate and linguistically correct translations, difficulties appear because of the functional translation differences between AI-generated and human-generated texts.

This research study emphasizes the significance of AI systems comprehending the context to enhance the quality of translation. The specific research purpose of this study is to enhance the AI-based translation systems to promote the exchange of culture via films. This study aims to advance the knowledge within translation studies since the incorporation of AI in translation enhances the efficiency of the translation process. Because of this, it has the potential to influence significantly the future of film translation. The subsequent sections will discuss the literature review of similar studies and the methods used in this study. This shall be followed by a comparative analysis of the results and the proposal about enhancing AI translation systems in film translation. Our discussion will cover implications and recommendations for future advancements in this area.

2. Literature Review

Artificial Intelligence (AI) has significantly impacted various fields, including translation. AI-powered translation systems, like ChatGPT, are becoming popular because they can translate things quickly (Ozdemir, 2023). However, using AI to translate movies raises some critical questions about how well these systems can maintain the quality of the translation (Wilks, 2008; Abdallah, 2012; Basmatkar, 2019). Much research (Jiang & Zhang, 2023; Iqbal, 2023; Yang et al., 2023; Santiago, 2023) has been done to compare translations done by people and by machines.

While there has been some research on the quality of AI-driven translation systems (Sardinha, 2024; Barag, 2023), film translation in Arabic has not been studied extensively. It would be beneficial to explore this topic further to better understand the challenges and opportunities in this field.

Various studies have been done on the challenges and solutions in subtitling and dubbing English-language films into Arabic. Alkandi's (2010) work is a valuable addition to the field since the study focuses on the lack of quality in the translations delivered by translators. Furgani's (2016) paper offered a detailed account of the main challenges translators encounter in translating films from English to Arabic, alongside practical solutions to these issues.

The paper (Debbas & Haider, 2020) on translation limitations of the American animated sitcom Family Guy into Arabic is a significant segment to consider. Alharthi's (2016) research study is important and valuable in contributing to the existing knowledge as it focuses on the subtitling strategies employed by Arab translators to overcome the technical, linguistic, and cultural drawbacks in translating humor in Seinfeld. The study also reveals the potential precursors that may have preceded the subtitlers' choices, which are crucial in subtitling and dubbing.

Linguistic differences discussed in most studies were categorized based on the MD approach, focusing on different genres (Ali & Ali, 2023; Ali & Thompson, 2022). However, no prior study has compared the film translation that was carried out by ChatGPT to human translation using the identified parameters from Biber's (1991) multidimensional analysis. Further research is required to investigate the efficiency of these translation techniques and the extent of their influence on the quality of offered translations.

The present study seeks to fill this gap by comparing film translations that human translators and ChatGPT have done. It also uses Biber's multidimensional analysis framework that systematically compares the human and AI translations and systematically compares involved vs. informational, narrative vs. non-narrative, explicit vs. situation-dependent, overt argumentation/persuasion, and abstract vs. non-abstract dimensions.

3. Delimitations

The study primarily centers around a linguistic analysis of lexico-grammatical variations between human-written and machine-generated text on Biber's five textual dimensions. Human-written and machine-generated translations were considered two linguistic varieties for the analysis. The research delimits the movies based on their Middle Eastern Arabic origin and the availability of human subtitling. The movies were taken from the last decade. Further, the focus was on the quantitative analysis and providing a functional interpretation of each dimension. It offers a continuous range of linguistic features associated with each of the dimensions (involved versus informational discourse, narrative versus non-narrative concerns, explicit versus situation-dependent expressions, overt expression of argumentation/persuasion, and abstract versus non-abstract discourse), which helps in determining the position of understudy linguistic productions on the continuum of the five factors.

4. Methodology

As part of our research, we compared how humans and machines translate Middle Eastern Arabic movies. To cover a diverse range of films, we chose various ones appropriately. In this work, we have gathered translations from real human translators and an artificial intelligence translator named ChatGPT. The Middle Eastern selected films from 2011 to 2018 embrace various societal topics. All these films are from different countries in the Middle East, such as Lebanon, Saudi Arabia, and Jordan, thus providing a regional view. Issues of social relevance, history, relationships, roles of men and women, existence, and ways of handling conflicts are depicted in the films. This diversity guarantees the essential coverage of Middle Eastern societies; therefore, these films are highly beneficial for research purposes.

Brown et al. (2020) described the ability of language models such as ChatGPT to translate language. They showed how the model can produce reasonable and correct translations without fine-tuning for translation tasks. Moreover, Banat and Adla (2023) revealed that ChatGPT can translate Arabic text into English. Unlike other translation tools that may need additional installation or technical skills to run, ChatGPT has an easy-to-use interface that anyone can easily access. The human translations were collected from an OpenSubtitles database, while the other translations were generated by ChatGPT 3.5. The data from the last decade was used, thus enabling a

comparison of the film translation practices within a given period. The selected movies were split into 25 text files for further analysis.

To explain the translations' features in detail, we employed Biber's multidimensional analysis. Thus, the present study provides a richer picture of translation quality beyond a single metric. This framework consists of five dimensions that assist in assessing the translations and comparing the translations done by human translators and ChatGPT. The five dimensions discussed are involved versus informational, narrative versus non-narrative, explicit versus situation, expression of argumentation/persuasion, and abstract versus non-abstract. These dimensions help in distinguishing between the general and specific types of texts, narratives, and non-narratives, texts that can be understood without situational knowledge and those that require it, texts with opinions or attempts at persuasion, and texts with no views and no attempt at persuasion, and texts that deal with concepts and ideas as opposed to the texts that deal with concrete issues.

The data was labeled with the help of distinct codes. Next, the specific linguistic features were found with the help of the Biber tagger developed in 1991. Applying factor analysis, a factor solution was achieved based on identifying more than 150 linguistic features. In addition, the frequencies of the linguistic characteristics were computed using Biber's Tag Count program. These frequencies were then normalized to meet equal weightage of all the features. The research that was conducted involved using statistics and, in particular, the SPSS software to determine whether there are statistically significant differences in any identified linguistic features between the texts written by humans and the texts written by the AI models.

After defining the dimensions, the research compared the translations by human translators and ChatGPT based on Biber's framework. Hence, this analysis aimed to determine how aspects of lexico-grammatical features were similar or different. This study also presents a functional interpretation of the dimensions to support the analysis of translation quality and enhance machine translation. The conclusion was drawn based on the data obtained from the comparative study of the identified factors, and recommendations were made for the further improvement of AI-based translation systems in film translation.

The study focused on improving machine learning and AI methods to help people understand movies in different languages. The study procedure was quite rigorous, which provided us with specific insights into the effectiveness of AI systems in the movie translation process. The findings of this research can help design improved AI technologies for translating films in the future.

Analysis and Discussion

Table 1 presents the results of comparing the translation of the Middle Eastern Movies done by Human and Chat GPT. The analysis and discussion are based on descriptive statistics. On Dimension one (D1), "Involved versus Information discourse", both the translations show involved discoursed to a varying degree. Table 1. Presents the results of the two categories being analyzed: Translation of Middle Eastern Movies by AI (Artificial Intelligence) and Humans.

Table 1. Linguistic variations between the translation by humans and AI on D1

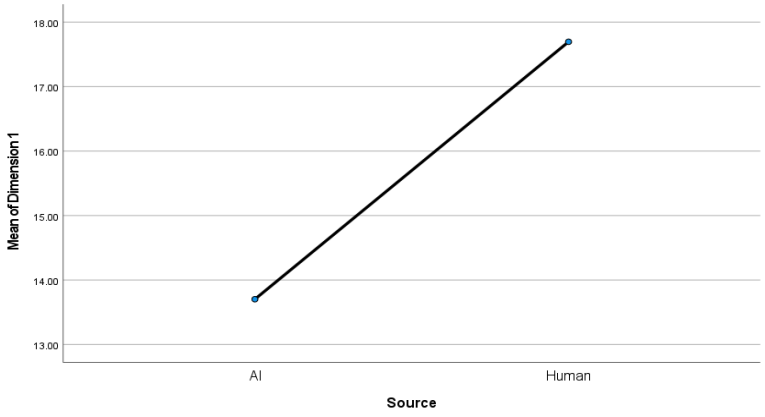
		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
AI		25	13.7032	6.51003	1.30201	11.0160	16.3904
Human		25	17.6956	6.99004	1.39801	14.8103	20.5809
Total		50	15.6994	6.98253	.98748	13.7150	17.6838
Model	**Fixed Effects**			6.75430	.95520	17.6200	17.6200
	Random Effects				1.99620	41.0635	41.0635

The mean score of AI translations is 13.70. It indicates that it produces involved discourse. In a text where involved discourse is produced, it means that the writing is using such types of linguistic features that involve the audience/ readers. Among various linguistic features, wh-clauses, wh-questions, hedges, amplifiers, and Qualifier-Emphatic are some of the linguistic features that mark the presence of involved discourse. A feature like an amplifier shows 'solidarity with the listener' (Biber, 1991, p. 240) thus involving the audience. Regarding translation by AI, it shows a 1.30 standard error. The standard error is a statistical measure that provides insight into the accuracy of a sample's population representation. It is calculated by dividing the sample's standard deviation by the sample size's square root. The resulting value indicates how much the sample means to deviate from the actual population. A minor standard error indicates that a sample is more representative of the population, while a significant standard error indicates that the sample is less accurate. It is an essential concept in statistical analysis as it assists in determining the reliability of the sample data and provides a basis for making inferences about the larger population.

The results also suggest that the standard deviation is 6.51. In statistics, the standard deviation is one of the most important measures to determine the spread or variation of a given data set. This tells us how dispersed the data is from the medium or the average value of the data. The standard deviation is small when the spread of data points is small and close to the mean, while the large standard deviation shows a large spread of data points. It can be useful in providing specific insights into the kind of data available or the information we will arrive at once we have analyzed it. As a result of our investigation, the estimated mean value of the parameter is between 11.01 and 16.39, with a 95% confidence level. This means there is a high likelihood that the true population mean will lie within this range.

Concerning the descriptive outcomes of human translations, the mean value is 17.69, meaning that it creates the involved discourse. The standard deviation is 6.99, and the standard error is 1.39. Thus, the confidence interval of 95% for the mean ranges from 14.81 and 20.58.

20.58.



A clear pattern can be observed when comparing the translations of Middle Eastern movies by humans and AI. Human translations, with a mean score of 17.69, exhibit a more involved discourse than AI translations (13.70). The statistical results indicate a dense presence of linguistic features in human translation that produce involved discourse, a level that AI translations are yet to achieve. The combined results of both categories yield a mean score of 15.69, a standard deviation of 6.98, and a standard error of 0.98. These combined results establish a 95% confidence interval for the mean between 13.71 and 17.68, further reinforcing the statistical significance of the findings.

Table two presents the results of descriptive statistics of the translations of Middle Eastern Movies done by AI and Humans on dimension 2 (D2). In this dimension, "Narrative versus Non-narrative Concerns," both AI and humans produce non-narrative discourse to varying degrees. Present tense verbs and pronouns are among those linguistic features that mark the presence of non-narrative concerns. It is clear from the name that non-narrative means related to the present time.

Table 2. Linguistic variations between the translation by humans and AI on D2

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
AI		25	-1.8132	2.51190	.50238	-2.8501	-.7763
Human		25	-1.1728	2.62111	.52422	-2.2547	-.0909
Total		50	-1.4900	2.56126	.36222	-2.2209	-.7651
Model	**Fixed Effects**			2.56709	.36304	-.7631	-.7631
	Random Effects				.36304	3.1199	3.1199

Translations by the AI with a mean score of -1.81 produce non-narrative discourse. The results indicate that the standard deviation is 2.51, with a standard error of 0.50. The 95% confidence interval for the mean falls between -2.85 and -0.77. Meanwhile, translation by humans, with a mean score of -0.17, produces relatively less non-narrative discourse than produced by AI. Its standard deviation and standard error are 2.62 and 0.52, respectively. The 95% confidence interval for the mean falls between -2.25 and -0.09.

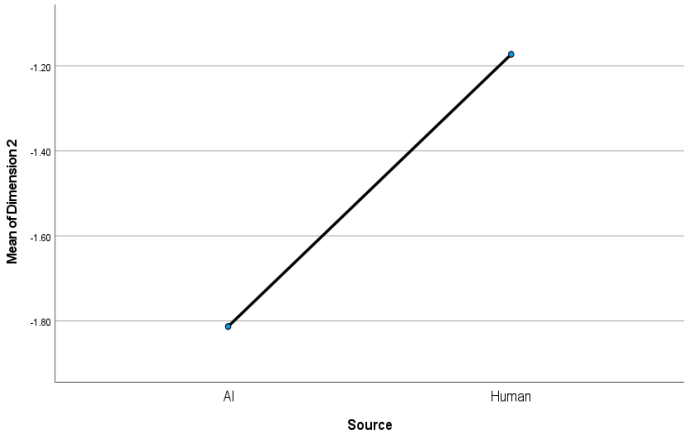


Figure 2. Comparison between the translation by Human and AI on D2

The combined results show a mean score of -1.49, with a standard deviation of 2.56. The standard error is 0.36. The 95% confidence interval for the mean falls between -2.22 and -0.76. The fixed effects estimate provides information about the average difference between the categories. In this case, the estimate is -0.76.

In Table 2, it has been estimated that there is a random effect of 0.36. The random effects estimate indicates that there may be variations beyond the fixed effects, which may suggest differences among the samples or other factors not accounted for in the analysis.

Table 3. Linguistic variations between the translation by humans and AI on D3

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
AI		25	-.6356	3.26265	.65253	-1.9824	.7112
Human		25	-2.1944	2.85271	.57054	-3.3719	-1.0169
Total		50	-1.4150	3.13363	.44316	-2.3056	-.5244
Model	**Fixed Effects**			3.06455	.43339	-.5436	-.5436
	Random Effects				.77940	8.4882	8.4882

Based on the information provided in Table Three, it is observed that both human and AI translations generate a context-specific conversation that heavily relies on the situation on dimension 3 (D3). Specific linguistic features like adverbs of time and place characterize this type of conversation. When these features occur frequently, it indicates the presence of dense situation-dependent discourse. Therefore, using these markers plays a significant role in producing contextually relevant and meaningful translations.

According to the data presented in Table Three, the AI translation has a mean score of -0.63, which suggests that the produced discourse serves mixed purposes. According to Biber's (1991) observations, a score closer to zero means that the discourse is mixed purpose, which suggests that the discourse contains both positive and negative linguistic features. The standard deviation of the scores is 3.26, with a standard error of 0.65. The mean score falls between -1.98 and 0.71 with a 95% confidence interval, indicating a range of possible values for the actual mean.

The analysis reveals that human translations exhibit a discourse heavily influenced by the context of the situation. This is supported by a mean score of -2.19, which indicates the presence of negative linguistic features. Human translations show a tilt toward situation-dependent discourse with the presence of the linguistic features which fall on the negative polarity of the continuum.

The standard deviation of 2.85 indicates a wide range of scores, while the standard error of 0.57 indicates a relatively small margin of error. Furthermore, the 95% confidence interval for the mean score is between -3.37 and -1.01, which suggests that the actual mean score could fall within this range with 95% certainty. According to the information provided in Table 3, it can be observed that both categories possess an identical mean score of -1.41. Moreover, the standard deviation for these categories is 3.13, while the standard error is 0.44. Further, the 95% confidence interval for the mean score falls between -2.30 and -0.52.

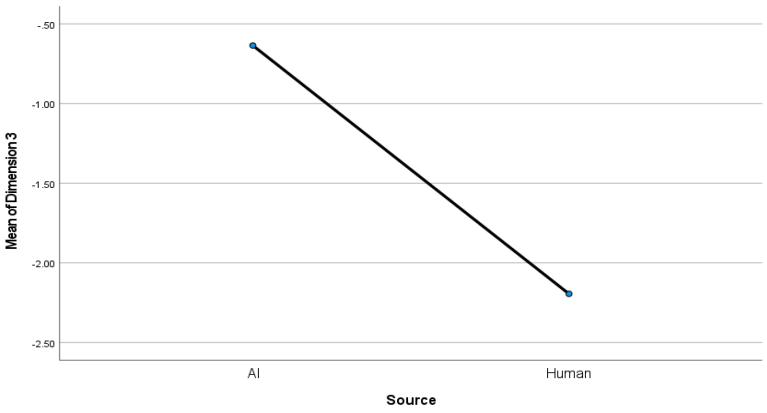


Figure 3. Comparison between AI and Human Translations on D3

Figure three suggests that both AI and human translations received overall negative scores. This implies that both translations were context-based. Nonetheless, a notable distinction exists between the two, as there was a statistically significant contrast in the mean scores of AI versus human translations.

Upon further examination, it became apparent that AI translations are heavily influenced by the specific circumstances in which they are generated. This highlights the critical impact of context on the overall quality of the translation. In contrast, human translations demonstrate less reliance on the situation, resulting in a more consistent level of quality regardless of the context. This dimension specifically showed the most difference in the mean scores. The following excerpts will clarify this difference.

The selected dialogue from the SubTitles exemplifies situation-dependent discourse, as highlighted in Biber's third dimension of multidimensional analysis (explicit vs. situation-dependent discourse). The passage reads:

> Daoud: "What's the matter with you? What's the matter with all of you? What's going on? Don't you realize who just got killed? Have you all forgotten who saved you?! Where are you now? Where are you?! You're not going anywhere! I'm gonna kill him." (SCAFH-5)

This dialogue demonstrates situation-dependent discourse through the use of wh-pronoun-relative clauses ("What's the matter with you?" and "What's going on?"), which anchor the discourse in the immediate context. The absence of adverbs of time and place ("now," "here") underscores the implicit nature of the dialogue, relying on the shared situational understanding of the characters.

In contrast, the AI-translated subtitles show a balanced discourse on Biber's third dimension, with a dimension score of -0.63.

> Shuaib: "Moses must stay with us."
>
> Jesus: "Moses only? Issa, we barely know them. We do not know the good of Moses from his evil."
>
> Jesus: "But he saved your life. What more do you want to know?"
>
> Shuaib: "We have to be careful and cautious, the world has changed."
>
> Isa: "Being careful did not protect my mother from getting sick from the water." (SCAFAI-5)

This conversation combines explicit and situation-dependent features. Using wh-pronoun-relative clauses ("What more do you want to know?") provides clarity. At the same time, the absence of explicit temporal and spatial references ties the conversation closely to the specific situation.

Overall, the human-translated passage exemplifies situation-dependent discourse. In contrast, the AI-translated passage balances explicitness and situational dependence, supporting the claim that AI translations yield a more balanced discourse according to Biber's multidimensional analysis.

In terms of dimension 4 (D4), it was found that both automated translations by AI and human translation produce argumentative discourse. This was deduced by observing the positive mean scores that fall on the positive polarity of the continuum on "overt expression of argumentation/persuasion." Recent findings have revealed the influential impact of specific language elements in fostering persuasive discussions, namely persuasive verbs, infinitive verbs, and predictive and essential models. Conversely, the data revealed that texts with lower scores also contained negative language components. Notably, third-person pronouns and both private and public verbs were found to be the primary discourse markers in the negative category. Both translations presented discourse that served dual purposes to some extent.

Table 4. Linguistic variations between the translation by humans and AI on D4

		N	M	St. D	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
AI		25	.2624	3.49070	.69814	-1.1785	1.7033
Human		25	.6696	3.17472	.63494	-.6409	1.9801
Total		50	.4660	3.30863	.46791	-.4743	1.4063
Model	**Fixed Effects**			3.33645	.47185	1.4147	1.4147
	Random Effects				.47185	6.4614	6.4614

Table 4 displays the data, revealing AI with a mean score of 0.26, placing it on the positive end of the continuum. However, the fact that its score is relatively close to the zero-dimension score suggests that AI's discourse is multifaceted in purpose. This is further supported by a significant standard deviation of 3.49, indicating a wide range of performance. Additionally, the 95% confidence interval for the mean score falls between -1.17 and 1.70, highlighting the considerable uncertainty surrounding the actual value of the mean score.

Regarding human translation, the average score of 0.66 indicates that the most desirable linguistic features are present. However, the proximity to zero suggests that the language used in the discourse may have multiple intentions, indicating a blend of purposes. The data is widespread, as shown by the standard deviation of 3.17, but the small standard error of 0.63 indicates a narrow margin of error. With a 95% confidence level, the average lies between -0.64 and 1.98.

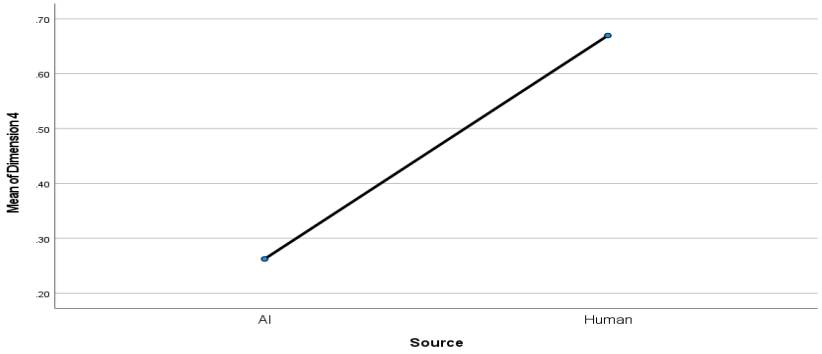


Figure 4. Comparison between AI and Human Translations on D4

Based on the analysis of descriptive statistics, it was observed that both the AI and Human categories displayed positive mean scores, with the total mean score also falling in the positive range. According to Figure 4, the Human category has a slightly higher mean score (0.66) on Dimension 4 than the AI category (0.26), indicating that, on average, the Human category tends to be slightly more argumentative in their discourse. The scores being close to zero indicate that the translations of Middle Eastern movies exhibit mixed-purpose discourse, with an overall mean score of 0.46. The standard deviation of 3.30 suggests that the scores are widely spread out, while the standard error is 0.46. The 95% confidence interval for the mean falls between -0.47 and 1.40, which means that the mean score most likely lies between these two values with high certainty.

The data presented in Table 5 reveals some interesting insights about the style of language used in both AI-generated and human translations. The average scores suggest that negative linguistic features are used in both translations. Moreover, the language used in both types of translations tends to be non-abstract/ non-impersonal, which falls under dimension 5 (D5). This dimension is known as "Impersonal/ Abstract versus Non-impersonal (Non-Abstract Style)."

Table 5. Linguistic variations between the translation by humans and AI on D5

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound
AI		25	-1.4800	1.68177	.33635	-2.1742	-.7858
Human		25	-1.9804	1.52314	.30463	-2.6091	-1.3517
Total		50	-1.7302	1.60795	.22740	-2.1872	-1.2732
Model	**Fixed Effects**			1.60442	.22690	-1.2740	-1.2740
	Random Effects				.25020	1.4489	1.4489

Dimension 5 refers to the two categories being analyzed: AI (Artificial Intelligence) and Human. The mean score of AI (-1.48) suggests that the discourse related to AI is predominantly negative and lacks abstraction. The standard deviation of 1.68 indicates that the data points are dispersed around the mean, while the standard error of 0.33 shows how accurately the mean represents the population. The 95% confidence

interval for the mean score is between -2.17 and -0.78, which means that there is a high probability that the actual mean score falls within this range.

According to the analysis of human translations, certain linguistic features indicate the existence of non-abstract discourse. The mean score of these translations is -1.98, which suggests the presence of these features. The standard deviation of the scores is 1.52, which means that the scores varied quite a bit. The 95% confidence interval for the mean falls between -2.60 and -1.35, which provides a range of values within which we can be 95% confident that the actual population mean lies.

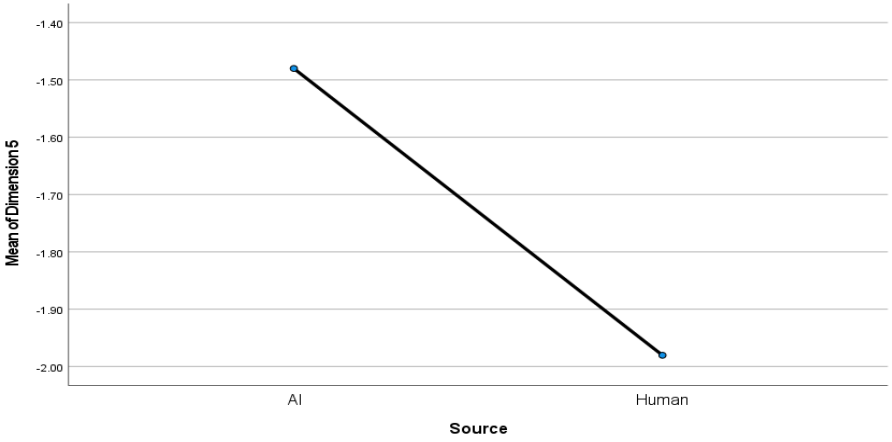


Figure 5 Comparison between AI and Human Translations on D5

The data in question compares the function of translations produced by AI and humans on Dimension 5. This dimension measures the level of abstractness in the translations.

According to the data, AI has a mean score of -1.48 on Dimension 5, which is lower than the score of -1.98 achieved by humans. When we combine the data from both categories, we obtain a mean score of -1.73, indicating that both AI and human categories have negative mean scores. The standard deviation for the combined data is 1.60, while the standard error is 0.22. The mean value can be estimated with 95% confidence to lie within the range of -2.18 and -1.27.

These descriptive statistics suggest that both AI and human categories exhibit negative mean scores in translation. However, the AI category has a slightly higher mean score than the human category, which indicates that, on average, AI produces less non-abstract discourse than humans on Dimension 5.

5. Conclusion

The comparison of human and AI translations of Middle Eastern movies has yielded some findings. In terms of the dimension of "Involved versus Information Discourse," it was observed that human translations exhibit a higher level of involved discourse compared to AI translations. However, AI translations tend to produce more non-narrative discourse. It is worth noting that both human and AI translations exhibit varying degrees of non-narrative concerns.

Furthermore, the analysis of situation-dependent discourse suggests that the context of the situation influences both human and AI translations. However, AI translations exhibit a mixed-purpose discourse, indicating positive and negative linguistic features. On the other hand, human translations exhibit a discourse heavily influenced by the context, with a higher presence of negative linguistic features.

In terms of argumentative discourse, both human and AI translations generate argumentative discourse, with human translations displaying a slightly higher mean score. The analysis indicates the presence of non-abstract discourse in both types of translations, with human translations showing a higher mean score.

Upon thoroughly evaluating the descriptive statistics, it becomes apparent that the mean scores for human translations surpass those of AI translations. The overall average score falls between the individual averages for the two categories. The fixed effects estimate exposes a notable contrast between the AI and human classifications. However, further examination is necessary to assess the extent and importance of this distinction. These findings offer initial insights into the dimension scores and pave the way for further statistical exploration into the significance and consequences of the observed variances between AI and human categories. More studies can be conducted to explore the other important aspects of translation quality, such as cultural nuances, stylistic choices, or audience reception.

Acknowledgments

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1446).

Authors' contributions

Dr. Sadia Ali was responsible for data collection and drafting the manuscript. Dr. Naeem Afzal was responsible for the study design and revising. Both authors read and approved the final manuscript.

Funding

This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2024/R/1446).

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Sciedu Press.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Abdallah, K. (2012). *Translators in production networks: Reflections on agency, quality and ethics.* Ita-Suomen yliopisto.

Alharthi, A. (2016). *Challenges and strategies of subtitling humour: a case study of the American sitcom Seinfeld, with particular reference to English and Arabic (28467715)* [Doctoral dissertation]. ProQuest Dissertations and Theses Global.

Ali, S., & Ali, M. (2023). Exploring Lexico-grammatical patterns in Saudi press reportage: A multidimensional analysis. *Arab World English Journal, 14*(1), 502-512. https://doi.org/10.24093/awej/vol14no1.32

Ali, S., & Thompson, P. (2022). Reporting local cultures in the globalised world: How indigenised can English be in the free world? *Corpora, 17*(2), 197-224. https://doi.org/10.3366/cor.2022.0254

Alkadi, T. (2010). *Issues in the subtitling and dubbing of English-language films into Arabic: Problems and* solutions [Doctoral dissertation]. Retrieved from https://www.semanticscholar.org/paper/ISSUES-IN-THE-SUBTITLING-AND-DUBBING-OF-FILMS-INTO-Alkadi/f9e8e1708ceb11cb9e6840aab403e19f6317e23f

Banat, M., & Adla, Y. A. (2023). Exploring the Effectiveness of GPT-3 in Translating Specialized Religious Text from Arabic to English: A Comparative Study with Human Translation. *Journal of translation and language studies*, *4*(2), 1-23. https://doi.org/10.48185/jtls.v4i2.762

Barag, S., Raja, A., & Shieh, J. (2023). Human-generated vs. AI-generated articles on XLH: A comparative analysis. *Medical Journal of Southern California Clinicians,* 48-56. https://doi.org/10.38206/160107

Basmatkar, P., Holani, H., & Kaushal, S. (2019). Survey on neural machine translation for multilingual translation system. *2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), 53*(5), 1-38. https://doi.org/10.1109/iccmc.2019.8819788

Biber, D. (1991). *Variation across speech and writing*. Cambridge University Press.

Borger, J. G., Ng, A. P., Anderton, H., Ashdown, G. W., Auld, M., Blewitt, M. E., … Naik, S. H. (2023). Artificial intelligence takes center stage: Exploring the capabilities and implications of ChatGPT and other AI-assisted technologies in scientific research and education. *Immunology & Cell Biology, 101*(10), 923-935. https://doi.org/10.1111/imcb.12689

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., … Amodei, D. (2020). Language models are few-shot learners. *Advances in neural information processing systems, 33*, 1877-1901. Retrieved from https://arxiv.org/abs/2005.14165

Debbas, M., & Haider, A. S. (2020). Overcoming cultural constraints in translating English series: A case study of subtitling family guy into Arabic. *3L The Southeast Asian Journal of English Language Studies, 26*(1), 1-17. https://doi.org/10.17576/3l-2020-2601-01

Furgani, K. T. (2016). *A study into the challenges of subtitling English into Arabic* [Doctoral dissertation]. Retrieved from https://researchonline.ljmu.ac.uk/id/eprint/4809/1/A%20study%20into%20the%20challenges%20of%20subtitling%20English%20into%20Arabic.pdf

Hadi, M. U., Tashi, Q. A., Qureshi, R., Shah, A., Muneer, A., Irfan, M., … Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *TechRxiv*, 1-43. https://doi.org/10.36227/techrxiv.23589741.v1

Iqbal, N. (2023). *A comparison of the linguistic encoding of artificial-intelligence-generated academic essays and academic essays written by MSc ALSLA 2022-23 students* [Unpublished doctoral dissertation]. University of Oxford.

Jiang, Z., Lv, Q., & Zhang, Z. (2023). *Distinguishing Translations by Human, NMT, and ChatGPT: A Linguistic and Statistical Approach.* arXiv. Retrieved from https://arxiv.org/abs/2312.10750

Ozdemir, S. (2023). *Quick start guide to large language models: Strategies and best practices for using ChatGPT and other LLMs*. Addison-Wesley Professional.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems, 3,* 121-154. https://doi.org/10.1016/j.iotcps.2023.04.003

Roumeliotis, K. I., & Tselikas, N. D. (2023). ChatGPT and Open-AI models: A preliminary review. *Future Internet, 15*(6), 192. https://doi.org/10.3390/fi15060192

Santiago, G. (2023). *A Quantitative Comparative Study of Automatic Speech Recognition (ASR) Systems Using English and Lojban Languages in the Health Care Industry* [Doctoral dissertation]. Retrieved from https://www.researchgate.net/publication/260647887_Comparative_study_of_automatic_speech_recognition_techniques

Sardinha, T. B. (2024). AI-generated vs human-authored texts: A multidimensional comparison. *Applied Corpus Linguistics, 4*(1), 100083. https://doi.org/10.1016/j.acorp.2023.100083

Wilks, Y. (2008). *Machine translation: Its scope and limits.* Springer Science & Business Media.

Yang, Z., Li, L., Lin, K., Wang, J., Lin, C. C., Liu, Z., & Wang, L. (2023). *The dawn of lmms: Preliminary explorations with gpt-4v (ision).* arXiv. https://doi.org/10.48550/arXiv.2309.17421