KSU-JRLT

Journal of Research in Language & Translation
Issue No. 2 Vol. 5 (2025)

جامعة الملك سعود
King Saud University

A Comparative Transitivity Analysis of Human vs. Artificial Intelligence-Based Translations: An Evaluation of Abstracts from Humanities and Social Science Studies

Mohammad Husam Alhumsi

Saudi Electronic University
Riyadh, Kingdom of Saudi Arabia

husam101010@gmail.com
https://orcid.org/0000-0002-0189-4443

Received: 14/02/2025; Revised: 23/07/2025; Accepted: 03/08/2025

الملخص

تهدف هذه الدراسة إلى مقارنة وتقييم أنظمة الترجمة البشرية و الترجمة القائمة على الذكاء الاصطناعي المتمثلة ب (ChatGPT & Google Translate) من خلال تحليل العملية التعددية النحوية والدقة اللغوية في أربعة نصوص مترجمة من العربية إلى الإنجليزية. ولتحديد العملية التعددية، تم اعتماد نموذج نظام التعددية لدى هاليداي (Halliday)، مع التركيز على تحليل الخطاب النقدي. ونظرًا لطبيعة البيانات، تم استخدام المنهج الكمي. تظهر النتائج أن العمليات المادية هي الأكثر شيوعًا بنسبة (52.63% - 75.71%)، تليها العمليات العلاقية بنسبة (10% - 31.57%). وأظهرت الترجمات البشرية اتساقًا أكبر في العمليات المادية، بينما اختلفت الترجمات الآلية بشكل ملحوظ. أما العمليات الوجودية والسلوكية فكانت نادرة أو غائبة في نصوص محددة. كما أظهرت النتائج وجود اختلافات متباينة في منهجية الترجمة، حيث التزمت الترجمات البشرية بأنماط لغوية متوقعة بشكل أكبر. بالإضافة إلى ذلك، كشفت الدقة اللغوية - التي قيست باستخدام درجات BLEU أن ترجمة ChatGPT حققت درجة 0.77، متفوقةً على Google Translate التي حصلت على 0.69، مما يشير إلى توافق أفضل مع الترجمة المرجعية. يوصي البحث الى امكانية إجراء مزيد من الدراسات المستقبلية لاستكشاف نطاق أوسع من أنواع النصوص واللغات، وتقييم تأثير التنقيحات اللاحقة على مخرجات الترجمة الآلية للتحقق من هذه النتائج.

الكلمات المفتاحية: الترجمة بالذكاء الاصطناعي، شات جي بي تي، ترجمة جوجل، الترجمة البشرية، ملخص البحث، العلوم الاجتماعية

KSU-JRLT

Journal of Research in Language & Translation
Issue No. 2 Vol. 5 (2025)

جامعة الملك سعود
King Saud University

A Comparative Transitivity Analysis of Human vs. Artificial Intelligence-Based Translations: An Evaluation of Abstracts from Humanities and Social Science Studies

Mohammad Husam Alhumsi

Saudi Electronic University
Riyadh, Kingdom of Saudi Arabia

husam101010@gmail.com
https://orcid.org/0000-0002-0189-4443

Received: 14/02/2025; Revised: 23/07/2025; Accepted: 03/08/2025

Abstract

This study aims to compare and assess human and AI-based translation systems (ChatGPT and Google Translate) by analyzing transitivity processes and linguistic accuracy in four Arabic-to-English translated abstracts. To identify the transitivity process, Halliday's transitivity system model was adopted, highlighting critical discourse analysis. Due to the nature of data, a quantitative method was employed. The analysis reveals that material processes are the most frequent (52.63%–75.71%), followed by relational processes (10%–31.57%). Human translations show greater consistency in Material processes, while machine translations exhibit more variation. Existential and behavioral processes are rare or absent in specific abstracts. Also, the results showed the existence of systematic differences in translation approaches, with human translations adhering more closely to expected linguistic patterns. In addition, linguistic accuracy, measured using BLEU scores, shows that ChatGPT Translation achieved a BLEU score of 0.77, outperforming Google Translate, which scored 0.69, indicating better alignment with the reference translation. Further studies could explore a broader range of text genres and languages and evaluate how post-editing impacts machine-translated outputs to validate these findings.

Keywords: *AI-based translation; ChatGPT; Google Translate; human translation; research abstract; social sciences*

Introduction

Technology has a major role in shaping and directing modern life as Federspiel et al. (2023) highlights the substantial impact of artificial intelligence on human life. This impact spans a wide range of advanced artificial intelligence (henceforth AI) applications, some of which have already made challenges to humans at performing different operational and administrative duties (Kembaren et al., 2023; Moneus & Sahari, 2024). Nevertheless, the same superiority of human cognition that has consistently guaranteed human supremacy in the natural world also explains why humans continue to do exceptionally well in jobs requiring intelligence and thought (Moneus & Sahari, 2024). This potentially serves communication skills between nations in this digital era. For example, as globalization accelerates and information technology rapidly advances, international communication has become a vital bridge between cultures and languages, highlighting the crucial role translation plays in this process (Dai, 2024; Kembaren et al., 2023). Moreover, technology helps make the trend of international communication easier and faster. In such settings, translation becomes a crucial component to guarantee correct comprehension, encourage cross-border cooperation, and facilitate cross-cultural exchanges. (Kembaren et al., 2023; Muhdi, 2019; Supriatnaningsih et al., 2019). This substantially has led to the widespread availability of translation software models based on AI, such as Google Translate, Amazon Translate, Bing, DeepL, Microsoft Translator, and Systran Translate. There are also a number of computer-aided translation (CAT) tools available, including Lokalise, Memoq, Memsource, Trados, TextUnited, Smartling, Crowdin, and Smartcat. Recently, artificial intelligence has been used to create programs like ChatGPT, ChatSonic, GPT-3 Playground, Chat GPT 4, and YouChat that more closely resemble human interactions by simulating conversational answers to scholars' questions (Moneus & Sahari, 2024). However, AI translation is increasingly competing with traditional human translation methods, driven by ongoing advancements in AI technology (Zagood et al., 2021). Some even predict that AI may eventually replace human translators in the industry. Still, human translation offers unique advantages, particularly when it comes to literary works, sparking ongoing debate and scrutiny over the role of AI in this field (Dai, 2024).

Evaluating machine translation, particularly in AI-based translation systems, poses outstanding challenges related to the vast and unpredictable nature of the corpus of studies, which compromise the reliability of automated metrics (Ghassemiazghandi, 2024). In response to these challenges, Papineni et al. (2002) introduced the Bilingual Evaluation Understudy (BLEU) metric, which quantifies translation quality by counting the shared words and n-grams between machine-generated and reference translations. Additionally, Han (2016) emphasized that automated metrics assess a machine translation system's performance by comparing its outputs to one or more human-produced translations. Nevertheless, given the limited studies on this approach, integrating human evaluations or alternative methods of translation quality assessment would provide a more detailed and comprehensive understanding of translation system performance, addressing a key gap in current research (Son & Kim, 2023).

For this research, a comparison of the system of transitivity analysis in translations produced by AI-based translation systems (ChatGPT and Google Translate) and human translation was conducted for the sake of examining their performance in translating a research

abstract. Significantly, English abstracts are becoming increasingly essential when searching for academic papers. The quality of the English abstract has a direct impact on the dissemination of scientific research which is a key factor in determining whether an article will be accepted by international indexing systems (Lin et al., 2024). In order to find the aspects of transitivity in a research abstract, transitivity analysis was employed. Based on transitivity model adopted by Halliday's Systemic-Functional Grammar (Halliday & Webster, 2014), this study assesses and compares translations made by AI-based translation systems and human translation of four research abstracts. Specifically, identifying transitivity characteristics in abstracts and providing researchers with relevant transitivity verbs and phrases in academic writing potentially serve as references when submitting to prestigious publications. This would also measure the effectiveness of AI-based translation systems compared to human translation.

All in all, it is important to note that generative artificial intelligence advances machine translation to a new phase. Through the generative AI translation framework, source texts are translated and deeply modified by entering instructions into an AI system such as ChatGPT (Fu & Liu, 2024). Nevertheless, recently, the linguistic characteristics of generative AI translation and its benefits and drawbacks in comparison to human translation remain insufficiently explored (Fu & Liu, 2024). To address this gap, the current paper aims to compare and evaluate the system of transitivity analysis in translations produced by AI-based translation systems (ChatGPT and Google Translate) and human translation by assessing and examining their performance in translating four research abstracts using BLEU tool.

Literature Review

Linguistic characteristics between AI translation and human translation

In terms of comparing AI-based translation systems to human translation, linguistic analysis research has covered a wider range of subjects, such as lexical analysis (Frankenberg-Garcia, 2022; Vanmassenhove et al., 2019), cohesion (Niu et al., 2020), complex morphological terms (Passban et al., 2018) and syntactic aspects (Sianipar & Sajarwa, 2021). It has been found that lexical analysis is of interest to many academics. For example, Klebanov and Flor (2013) noted that human translation has a higher lexical tightness (a measure of how lexical items are related within a text) than machine translation. Current machine translation methods fall short of translation conducted by humans in terms of lexical variation when compared to human translation in the setting of English-French and English-Spanish language pairings (Vanmassenhove et al., 2019).

Furthermore, Frankenberg-Garcia (2022) claimed that machine translations have issues such as lexical inconsistency as revealed in their corpus-driven keyword study. In order to improve neural machine translation's performance when translating morphologically complicated terms, Passban et al. (2018) suggested a double-attentive decoder structure and double-channel encoder. In terms of coherence, machine translation shows less coherence than human translation (Niu et al., 2020), while Wong and Kit (2012) noted that human translation typically employs more coherent devices. Also, Syntactic elements in Indonesian-English language pairs vary between machine translation and human translation in terms of voice conversion, sentence structure, tense, and passive voice usage (Sianipar & Sajarwa, 2021). In

case of style and figurative language, Mehawesh's (2023) research revealed that emotive words are interpreted better when translated by human translators.

Despite the large number of prior studies on the linguistic comparison between machine translation and human translation, there are still some significant gaps. First of all, there is a dearth of research on comparing the system of transitivity analysis in translations produced by AI-based translation systems (ChatGPT and Google Translate) and human translation, examining their performance in translating a research abstract, whereas there have been prior and remarkable studies on comparing the language features of generative AI translation and human translation in literary (Zagood et al., 2021), linguistic distinctions between generative AI translation and human translation in scientific publications (Fu & Liu, 2024), and legal materials (Moneus & Sahari, 2024). Second, the focus of previous research is primarily on Indo-European languages, such as English, French, and Portuguese, on one hand and East Asian languages such as Chinese on the other, with comparatively little attention paid to Arabic. Given the large number of Arabic speakers worldwide and the potential for large language models like ChatGPT to provide alternatives to neural machine translation (Fu & Liu, 2024), it is crucial to investigate the linguistic characteristics of the system of transitivity in translations produced by AI-based translation systems (ChatGPT and Google Translate) and human translation through examining their performance in translating a research abstract in the context of Arabic-English scientific translation. Besides, few studies examined transitivity in research abstracts (Sari & Sari, 2016).

Research article abstracts and transitivity analysis

An abstract, which is a concise synopsis of the original work, is crucial to the article's effective dissemination (Lin et al., 2024). Vathanalaoha and Tangkiengsirisin (2018) contended that abstracts of research articles serve as publishers' initial means of assessing the caliber of research. Although numerous studies have examined the genre of abstracts in a variety of fields and languages in this context, few research has made an effort to identify the differences they might observe in a single domain (El-Dakhs, 2018; Saidi & Talebi, 2021). This is why the focus of the paper was on one domain which is Humanities and Social Science Studies.

In a very recent study, Lin et al. (2024) analyzed English for Science and Technology article abstracts using Halliday's transitivity model. Thirty abstracts are chosen at random from the most recent issue of Science to serve as the original data. The results show that only five transitivity processes are used in those 30 abstracts. Also, the constituent ratio differentiation of a subset of transitivity processes is statistically significant. It should be noted that their study examined high-frequency verbs of transitivity processes in order to provide guidelines for writing research article abstracts for English non-native scholars. However, their paper did not tackle human or AI based translation.

To highlight the importance of research abstracts in scientific and literary works, Saidi and Talebi (2021) confirmed that research article abstracts are crucial in determining the future of scholarly publications. Within the field of applied linguistics, their paper made an effort to investigate the constituent moves and move patterns of research article abstracts published in two international and local journals, namely, the Journal of English for Academic Purposes (international) and the Iranian Journal of English for Academic Purposes (local). Nevertheless,

transitivity analysis was not applied in Saidi and Talebi's (2021) research. Using Hyland's (2000) five-move model, the abstracts of research articles were examined. According to the findings, the abstracts that were published in the two journals had standard moves. Additionally, in the two corpora, the product, purpose, and techniques were the most frequently used moves. Their study also revealed differences in the move patterns between the two sets of articles, even if the chi-square test findings indicated no significant difference in the frequency of the moves in the abstracts.

In another study, Vathanalaoha and Tangkiengsirisin (2018) compared the tonal styles of each rhetorical moves between Thai and international dentistry research abstracts using a corpus-based transitivity analysis. The datasets were built using 120 abstracts of dental research articles that were chosen at random from the top five international dental journals according to Impact Factors and six Thai recognized dentistry publications. The datasets were examined using verbal choices of transitivity proposed by Thompson (2000) and six process kinds based on Halliday and Matthiessen (2014). The results revealed that verbal and existential processes were exclusively present in the Methodology moves of Thai refereed dental journals, despite the fact that both datasets had similarities in terms of the transitivity types present in each of the rhetorical moves. International dental journals' discussion moves also emphasized verbal processes.

According to Sari and Sari (2016), transitivity is frequently employed in the linguistic field to evaluate novels, short stories, newspapers, and other discourses in order to deduce the discourses' ideologies. Analyzing transitivity processes in abstracts was the goal of their study. Since the research offers a descriptive analysis based on the findings' percentage of occurrence, both qualitative and quantitative methodologies are used to analyze the data. Out of twenty linguistics abstracts, five were selected at random. Materials (62.4%), relational (24.7%), verbal (5.4%), mental (4.3%), behavior (3.2%), and existential (1.1%) are the six transitivity processes that were examined. There are also thirteen circumstantial elements, with place having the highest percentage (54.4%) and frequency, commutative, and matter having the lowest (1.3% each).

In sum, the investigation drawn from prior studies has revealed a dearth of research on the transitivity analysis of translating abstracts of social science research articles using both human translation and AI-based translation systems. In order to address this gap, the current study compared the transitivity analysis system in translations generated by human translation and AI-based translation systems (ChatGPT and Google Translate) in order to assess how well each performed when translating a research abstract.

Machine translation evaluation (BLEU as an example)

Machine translation evaluation can be conducted through manual or automatic methods (Zakraoui et al., 2021). Manual evaluation, though providing a thorough assessment, is resource-intensive and time-consuming, whereas automatic evaluation offers less costs and efficient alternative, albeit with reduced comprehensiveness (Hadla et al., 2014). Despite its limitations, automatic evaluation is commonly preferred by researchers for its speed and affordability, with metrics such as Bilingual Evaluation Understudy (BLEU) frequently employed by large number of scholars (Papineni et al., 2002; Zakraoui et al., 2021) to assess and optimize the output and performance of machine translation systems (Zakraoui et al.,

2021). Thus, BLEU is a leading statistical metric used to measure the similarity between machine-generated translations and human-produced translations (Ghassemiazghandi, 2024; Reiter, 2018). In addition, Zakraoui et al. (2021) described BLEU as a method for assessing the quality of machine translation by measuring the similarity between the machine translation system's output and a reference translation, typically provided by a professional human translator.

Numerous studies in literature have introduced improved versions of the BLEU metric, such as those presented in the study conducted by Chen and Kuhn (2011). Researchers also employed various other metrics relying on different factors; these metrics include METEOR and Word Error Rate (Sai et al., 2020), and the AL-BLEU metric (Bouamor et al., 2014) with the latter that extends BLEU to accommodate the complex morphology of the Arabic language. Specifically, in an attempt to evaluate a range of automatic metrics, no conclusive evidence, however, emerged to suggest that any single metric outperforms the others among those analyzed in the studies conducted by Hadla, (2014) and Sai et al. (2020). Whereas automatic metrics such as BLEU offer an overall indication of a Machine Translation model's translation accuracy, these researchers do not provide detailed insights into the specific linguistic aspects that pose challenges for models in generating fluent and coherent output in relation to transitivity analysis. Besides, several studies have examined machine translation samples with native speakers to assess the linguistic aspects of MT errors (Almahasees, 2020; Marouani et al., 2020), while other research has employed neural networks to identify errors (Madi & Al-Khalifa, 2020) or correct these errors (Watson et al., 2018).

Types of machine translation

Machine Translation (MT) has undergone substantial advancements, progressing from early rule-based systems to modern data-driven approaches such as statistical and neural methods (Wang, 2024). Each paradigm employs distinct techniques, offering varying levels of accuracy, fluency, and adaptability.

Rule-Based Machine Translation (RBMT)

The earliest MT systems relied on Rule-Based Machine Translation (RBMT), which utilized predefined linguistic rules and bilingual dictionaries to convert text between languages. These systems typically involved analyzing the source sentence, applying syntactic and semantic transformations, and generating the target output (Abercrombie, 2016; Wang, 2024). While RBMT could achieve grammatically precise translations in controlled scenarios, Rule-based MT systems frequently struggle to handle complicated syntactic patterns and ambiguous words effectively (Wang, 2024).

Statistical Machine Translation (SMT)

Statistical Machine Translation (SMT) emerged as a data-driven alternative, leveraging probabilistic models derived from large bilingual text corpora (Brown et al., 1993; Wang, 2024). Unlike RBMT, SMT systems translated by identifying the most probable target phrases or sentences based on statistical patterns. Phrase-based SMT, for instance, segmented input into chunks and reassembled them using probability estimations (Wang, 2024). It is interesting to indicate that SMT improved translation fluency and reduced dependency on linguistic rules. Nevertheless, statistical machine translation systems are constrained by both the quantity and

quality of available training data, necessitating large-scale corpora to produce satisfactory translations (Abercrombie, 2016; Wang, 2024).

Neural Machine Translation (NMT)

Neural Machine Translation (NMT) revolutionized the field by adopting deep learning architectures, particularly sequence-to-sequence (Seq2Seq) models with attention mechanisms (Bahdanau et al., 2015). Unlike SMT, which processed translations in fragments, NMT encoded entire sentences into continuous representations and generated outputs holistically (Wang, 2024). Despite its advantages, NMT demands significant computational resources and large datasets, posing challenges for low-resource languages (Ranathunga et al., 2023).

Theoretical background of the article

A linguistic tool for clause-level text analysis can be identified as transitivity analysis (Alhumsi & Alsaedi, 2023). It is based on Halliday's Systemic Functional Linguistics and is used to determine the verbal, existential, relational, material, mental, and behavioral processes that are used in a text (Halliday & Matthiessen, 2014) (see Table 1). This analysis is essential to understand how the verb, subject, and object interact with one another and how each contributes to the ideologies and meanings expressed in the text (Alhumsi & Alsaedi, 2023). It is interesting to note that process, participants, and circumstances are the three main parts of a sentence based on Halliday's transitivity model (Halliday & Matthiessen, 2014). According to their model, "process" refers to the core experiential event or state of being realized by clausal construction being carried out, "participants" refers to the entities engaging in the process, and "circumstances" refers to the setting in which the process is taking place. To get an illustrated image, types of processes, roles of participants, and circumstances indicated by the transitivity system are shown in Table 1.

Table 1

Types of Processes, Participants' Roles, and Circumstances

Process Types	Participants' Roles	Circumstances
Material	Actor + Goal	Extent,
Mental	Senser + Phenomenon	Location (time, place),
Verbal	Sayer + Target	Cause,
Relational	Carrier + Attribute	Manner,
Existential	Existent	Matter, and
Behavioural	Behaver	Accompaniment

An actor and a goal are the main components in material processes; they are distinguished by occurrence and action. Mental processes are linked to mental states, and the participant's role is usually that of a sensor or perceiver and the role of phenomenon. Relational processes define the relationship between the subject and complement. Verbal processes are also connected to speech and communication. Existing procedures serve as indicators of something's existence or presence. Finally, behavior or physiological activities are described

using behavioral processes (Halliday & Matthiessen, 2014). Significantly, transitivity analysis has its own reputable position in literature.

Specifically, in reviewing literature, numerous studies have used transitivity analysis to examine a wide range of aspects related to language and meaning in various fields. For example, such fields involve academic documents (Alhumsi et al., 2024), research articles (Lin et al., 2024), political writings (Alhumsi & Alsaedi, 2023), and text news (Alhumsi & Alshagrawi, 2022). However, those studies did not address evaluating the performance of human translation and AI-based translation systems (ChatGPT and Google Translate) in translating a research abstract by comparing the transitivity analysis system in their translations. Hence, the aim of this paper is to address such comparison and evaluation, and the research questions that the paper poses are as follows:

1. Which transitivity processes are most frequently employed in research abstracts translated by human and AI-based translations?
2. To what extent do AI-based translations differ from human translation in terms of evaluation?

Methods

Research design and instruments

Based on the research questions, a quantitative research approach was used in this investigation. The main purpose of the quantitative approach is to ascertain how frequently different transitivity process types occur within a clause. Additionally, a quantitative approach is necessary to improve the precision of evaluating textual data (Köhler, 2012). For the current paper, human translation and AI-based translation systems, namely ChatGPT and Google Translator, together represent the research tools in this inquiry.

Sample of the study

Four abstracts from multiple disciplines published by different peer-reviewed scientific journals selected randomly served as the study's sample. These fields represent the area of humanities and social sciences, encompassing language, law, economic, and security sciences. Interestingly, the source text of the randomly selected abstracts is written in Arabic. Therefore, it is crucial to note that all published abstracts in these journals are mandatorily translated by professional, qualified and expert scholars from English to Arabic and vice versa. In addition, the abstract should be written in Arabic and English as a crucial requirement of the regulations of these journals.

To ensure consistency and homogeneity, the decision was made to focus on these four Arabic abstracts and translate them three times into English, using a combination of AI and human translation. This approach aligns with the idea presented by Palinkas et al. (2015), who highlighted the importance of homogeneity in research, which often calls for smaller sample sizes to achieve uniformity. Thus, by selecting only four abstracts randomly, one can ensure that the source content remains constant across all three distinct translations; all three translation outputs (human translation, ChatGPT translation, and Google translate) derive from identical source content. This eliminates the potential variability that could arise from using large number of abstracts, which might introduce uncontrolled factors such as genre effect into the analysis since these abstracts belong to social sciences.

Data collection procedure and data analysis

Paragraphs The data of the current study was collected from 4 abstracts from different Saudi journals publishing articles specialized in humanities and social sciences. Therefore, the selection of those abstracts from those journals forming as the study's sample is grounded in the journals' specialized focus and scope within the Arab world, making them highly relevant for the purpose of the study. Additionally, those journals, namely University of Tabuk Journal for Humanities and Social Sciences, Jazan University Journal of Human Sciences, Journal of Economic Studies, and Arab Journal for Security Studies, are well-respected institutions in Saudi Arabia known for producing research that is both academically rigorous and regionally significant. Using the randomly selected four abstracts ensures that the study draws from authoritative, peer-reviewed, and contextually appropriate articles in the Arab world since the language of the selected abstracts serving as the source text is written in Arabic. Specifically, Abstract 1 that discussed language issues was adopted from Alqahtani (2024). Abstract 2, which focused on security sciences, was adopted from Almoibed (2024). Abstract 3, which covered economic issues, was adopted from Hassan and Amin (2023). Finally, the last abstract (Abstract 4) addressed the field of law and was adopted from Alshahrani's (2024) study.

The four abstracts were translated into English using professional human translation and AI based translation systems such as ChatGPT and Google translation. Hence, three distinct English versions of translations for each abstract were employed as the data of the current study. For the sake of data analysis, Microsoft Office Word is used to divide the text of the three English translated versions of each abstract into clauses. To identify the transitivity patterns employed, the clauses were grouped according to Halliday and Matthiessen's (2014) transitivity system. Microsoft Office Excel was also used to determine the overall number of processes and how frequently they occurred in the three distinct English versions of the four abstracts.

Results

A comparative analysis of abstract 1 translated by human and AI-based translation systems

As expressed in Table 2, out of a total of 70 ranking clauses recognized in the three versions of the abstract 1 translated by using AI based translation systems and human translation, only behavioral process type did not exist in this abstract tackling language issues.

Table 2

A Comparative Analysis of Process Types for Abstract 1 Using Human and AI-Based Translation Systems

Translation Version/ Process Type	Human Translation	ChatGPT Translation	Google Translate	Total
Material	17	18	18	53
	73.91%	78.26%	75%	75.71%
Relational	3	2	2	7
	13.04%	8.70%	8.33%	10%
Verbal	1	1	1	3
	4.35%	4.34%	4.17%	4.29%
Mental	1	2	2	5
	4.35%	8.70%	8.33%	7.14%
Behavioural	----	----	----	----
Existential	1	----	1	2
	4.35%		4.17%	2.86%
TOTAL	23	23	24	70
	100%	100%	100%	100%

In the material category, human translation accounted for 17 clauses, which is 73.91% of the total. ChatGPT translation and Google Translate both had 18 clauses each, which represent 78.26% and 75% of the total, respectively. This category had the highest number of clauses compared to the other categories, with a total of 53 instances. For the relational category, human translation had 3 clauses, making up 13.04% of the total. Both ChatGPT translation and Google Translate each contributed 2 clauses, which represent 8.70% and 8.33% of the total, respectively. This category showed fewer instances than material but still contributed a significant proportion to the total.

In the verbal category, all three translation methods contributed equally, with 1 clause each. human translation represented 4.35% of the total, ChatGPT translation represented 4.34%, and Google Translate accounted for 4.17%. This category had the fewest clauses overall, making up a small portion of the total. As for the mental category, human translation contributed 1 clause (4.35% of the total), while both ChatGPT translation and Google Translate each contributed 2 clauses (8.70% and 8.33%, respectively). This category, like verbal, had a relatively low number of instances, but ChatGPT translation and Google Translate slightly outperformed human translation. The behavioral category showed no clauses recorded across any of the translation methods, indicating that no data was available or applicable for this category. Finally, in the existential category, human translation contributed 1 clause,

representing 4.35% of the total, and Google Translate had also one clause (4.17%). ChatGPT translation did not contribute any clauses in this category. However, this category had a small but noteworthy presence in the whole count. All in all, despite some variations between the methods, the totals show an almost equal distribution of clauses across the three translation methods, with Google Translate slightly surpassing the others in total sum.

A comparative analysis of abstract 2 translated by human and AI-based translation systems

As illustrated in Table 3, out of a total of 94 ranking clauses identified in the three versions of the abstract 2 translated by using AI based translation systems and human translation, all process types, as classified by Halliday and Webster (2014), were determined to occur in this specific abstract that is related to security science issues.

Table 3

A Comparative Analysis of Process Types for Abstract 2 Using Human and AI-Based Translation Systems

Translation Version/ Process Type	Human Translation	ChatGPT Translation	Google Translate	Total
Material	20	20	24	64
	71.4%	64.5%	68.6%	68.1%
Relational	3	8	4	15
	10.7%	25.8%	11.5%	15.9%
Verbal	2	2	3	7
	7.2%	6.4%	8.5%	7.5%
Mental	1	----	2	3
	3.5%		5.7%	3.2%
Behavioural	----	1	----	1
		3.3%		1.1%
Existential	2	----	2	4
	7.2%		5.7%	4.2%
TOTAL	28	31	35	94
	100%	100%	100%	100%

It is important to note that Table 3 offers a comparative analysis of three translation methods—Human Translation, ChatGPT Translation, and Google Translate—across various process categories derived from abstract 2. Specifically, human translation demonstrates a strong focus on material aspects, accounting for 71.4% of its output, which suggests a preference for concrete, factual content. This is complemented by smaller proportions in relational (10.7%), verbal (7.2%), mental (3.5%), and existential (7.2%) categories, indicating a balanced but limited engagement with abstract or interpersonal elements. Notably, human

translation entirely neglects behavioral processes, reflecting prioritization of content over interactional nuances. In contrast, ChatGPT translation shows a higher emphasis on relational processes (25.8%), nearly double that of human translation, suggesting a greater inclination toward interactional or communicative functions. However, its material focus (64.5%) remains dominant, though slightly reduced compared to human translation. ChatGPT also introduces a minor behavioral component (3.3%), absent in human translation, while entirely excluding mental and existential processes. Furthermore, Google Translate, like human translation, prioritizes material processes (68.6%) but includes marginally higher verbal (8.5%) and mental (5.7%) proportions than both human and ChatGPT translations. Its relational focus (11.5%) is closer to human translation, but it retains existential processes (5.7%), unlike ChatGPT. Nevertheless, it completely neglects behavioral elements, aligning more closely with human translation in structure but differing slightly in distribution. Overall, human translation maintains a more balanced and abstract-inclusive approach, while ChatGPT leans toward relational interaction, and Google Translate offers a middle ground with a stronger material focus and minor variations in verbal and mental processes. The absence of behavioral and existential elements in some systems highlights key distinctions in their functional priorities.

A comparative analysis of abstract 3 translated by human and AI-based translation systems

Table 4 shows that out of a total of 38 ranking clauses, three versions of the abstract 3 translated by using AI based translation systems and human translation were expressed. In this abstract tackling the field of administration and economics, behavioral and existential process types were not recorded.

Table 4

A Comparative Analysis of Process Types for Abstract 3 Using Human and AI-Based Translation Systems

Translation Version/ Process Type	Human Translation	ChatGPT Translation	Google Translate	Total
Material	5	8	7	20
	41.67%	61.54%	53.85 %	52.63%
Relational	5	3	4	12
	41.67%	23.08%	30.77%	31.57%
Verbal	1	1	1	3
	8.33%	7.69%	7.69%	7.90%
Mental	1	1	1	3
	8.33%	7.69%	7.69%	7.90%
Behavioural	----	----	----	----
Existential	----	----	----	----
TOTAL	12	13	13	38
	100%	100%	100%	100%

Significantly, Table 4 compares the performance of three translation methods—human translation, ChatGPT translation, and Google Translate—across various process types. It presents the number of clauses translated in each category along with the corresponding percentage of the total clauses for each method. In the material category, ChatGPT translation outperforms the other two methods, translating 8 items, which accounts for 61.54% of its total processes. Google Translate follows with 7 items (53.85%), while human translation records 5 clauses, representing 41.67%. This suggests that ChatGPT Translation is more heavily utilized for material-related processes compared to the other methods. For the relational category, human translation and Google Translate each record 5 and 4 items, respectively, comprising 41.67% and 30.77% of their total translations. However, ChatGPT translation only handles 3 items, which amount to 23.08%. This shows that human translation and Google Translate are more engaged with relational-type process than ChatGPT Translation.

The verbal category is uniformly translated by all three methods, with each method providing one clause, which makes up around 7.69% of the total for ChatGPT and Google Translate, and 8.33% for human translation. This category represents a relatively minor part of the overall translation workload across all methods. As for the mental category, each method has one item, representing 7.69% for both ChatGPT translation and Google Translate, and 8.33% for human translation. There is no significant difference in how these three methods address mental-type process. Lastly, the behavioral and existential categories have no translations recorded across all methods, which means that no translations were made in these categories during the process. The overall distribution reflects the different strengths of each translation method in various process types, showing a marked preference for ChatGPT-generated translations in rendering material processes, while human translations and Google Translate outputs demonstrate comparable frequency distributions for relational processes.

A comparative analysis of abstract 4 translated by human and AI-based translation systems

As illustrated in Table 5, out of a total of 113 ranking clauses, three versions of the abstract 4 translated by using AI based translation systems and human translation were described. Like abstract 1, only behavioral process type was not recorded in this specific abstract tackling the field of law.

Table 5

A Comparative Analysis of Process Types for Abstract 4 Using Human and AI-Based Translation Systems

Translation Version/ Process Type	Human Translation	ChatGPT Translation	Google Translate	Total
Material	25	25	26	76
	69.44%	67.57%	65%	67.26%
Relational	7	7	8	22
	19.44%	18.92%	20%	19.47%
Verbal	2	2	3	7
	5.56%	5.40%	7.5%	6.19%
Mental	1	3	3	7
	2.78%	8.11%	7.5%	6.19%
Behavioural	----	----	----	----
Existential	1	----	----	1
	2.78%			0.89%
TOTAL	36	37	40	113
	100%	100%	100%	100%

The above table compares how the three different methods—human translation, ChatGPT translation, and Google Translate—perform across various types of the five process types. From Table 5, the material category shows the largest proportion of process types, with human translation providing 25 processes (69.44%), ChatGPT translation delivering 25 processes (67.57%), and Google Translate providing 26 processes (65%). This category represents the majority of the total process types, indicating that all methods are relatively balanced in their handling of material-type processes. The relational category ranks second, with human translation and ChatGPT translation both having 7 clauses (19.44% and 18.92%, respectively), while Google Translate has 8 clauses, making up 20% of the total process types in this category. This reflects a slightly higher output for Google Translate in relation to the other two methods.

Regarding the verbal category, all three methods show a decrease in the number of items translated. Human translation and ChatGPT translation record 2 items or clauses (5.56% and 5.40%, respectively), while Google Translate has 3 clauses, making up 7.5% of the total process. The mental category shows an even clearer difference, with human translation providing only one clause (2.78%), and ChatGPT and Google Translate both provide 3 items (8.11% and 7.5%, respectively). These numbers reflect a stronger performance in mental-type

processes for ChatGPT and Google Translate compared to human translation. It is crucial that the existential category is present only in human translation with one clause (2.78%), and neither ChatGPT translation nor Google Translate contributed any clauses to this category. The overall comparison reveals the differences in translation output across different process categories, providing insight into the strengths and limitations of each translation process.

The evaluation of translation from Arabic to English using BLEU method tool

Four Arabic abstracts from various fields were used as the source text for this study, and the reference text was their human translations. The accuracy and BLEU score of the Arabic-to-English translations was assessed using the translations from ChatGPT and Google Translate. Hence, to evaluate the three versions of the four selected abstracts involving human translation and AI-based translation systems, Bilingual Evaluation Understudy (BLEU) method tool was adopted, and these versions were tested by employing this tool in the current study. It should be noted that the n-grams of the machine-generated translations to human translation were usually compared in order to determine the precision value for each translation. It is interesting to indicate that Uni-gram refers to individual words, Bi-gram refers to pairs of consecutive words, Tri-gram refers to triples of consecutive words, and Tetra-gram refers to quadruples of consecutive words. For example, to explain this technique, the following consequences show the work of n-gram:

1-gram: “This,” “study,” “aims”,

2-gram: “This study,” “study aims,” “aims to”,

3-gram: “This study aims,” “study aims to,” “aims to investigate”,

4-gram: “This study aims to,” “study aims to investigate,” “aims to investigate the”.

Moreover, the percentage of n-grams from the machine translations that are also present in the human translation is known as precision in this context. Given the n-grams counts from the analysis (where each machine translation had identical n-gram counts to the human translation), the precision value can be calculated by determining the number of matching n-grams for each type (1-gram, 2-gram, 3-gram, and 4-gram). Since the n-gram counts are identical for all three versions of the translation (Human, ChatGPT, and Google Translate) in this study, the precision values will be the same as illustrated in Table 6. The formula for precision is as follows:

$$\text{Precision} = \frac{\text{Number of matching n-grams}}{\text{Total number of n-grams in machine translation}}$$

The evaluation of translation of abstract 1 using BLEU method tool

To get a clear image, Table 6 depicts the precision values for different n-grams (Uni-gram, Bi-gram, Tri-gram, and Tetra-gram) measuring the quality of ChatGPT translation, and Google Translate in relation to Abstract 1.

Table 6

Precision Values for AI- Methods of Translations of Abstract 1

N-grams / Methods of translations	ChatGPT translation	Google translate
Uni-gram Precision (P1)	0.93	0.89
Bi-gram Precision (P2)	0.86	0.82
Tri-gram Precision (P3)	0.79	0.75
Tetra-gram Precision (P4)	0.72	0.68

To calculate the BLEU score for Table 6, here are the steps for calculating the BLEU score based on the updated precision values for each n-gram (Uni-gram, Bi-gram, Tri-gram, Tetra-gram):

BLEU Score Formula:

$$\text{BLEU} = \text{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right).$$

This formula includes *BP* (Brevity Penalty), *Pn* (the precision for each n-gram), *Wn* (the weight for each n-gram, which is *1/N* for N=4 n-grams), and Log (*Pn*) representing the natural logarithm of the n-gram precision.

From Table 6, for ChatGPT Translation, the precision values for each n-gram are as follows: the unigram precision (p1) is 0.93, the bigram precision (p2) is 0.86, the trigram precision (p3) is 0.79, and the tetra-gram precision (p4) is 0.72. On the other hand, for Google Translate, the precision values are slightly lower: the unigram precision (p1) is 0.89, the bigram precision (p2) is 0.82, the trigram precision (p3) is 0.75, and the tetra-gram precision (p4) is 0.68. These precision values reflect the accuracy of each method in matching n-grams with reference translations, with ChatGPT Translation consistently achieving higher precision across all n-gram levels compared to Google Translate. In short, the results indicate that ChatGPT translation achieved a BLEU score of 0.82, while Google Translate achieved a BLEU score of 0.78. This means that ChatGPT translation performed slightly better than Google Translate in terms of matching the reference translation, as reflected by the higher BLEU score. The BLEU score, which ranges from 0 to 1, measures the similarity between the machine-generated translation and the reference translation, with values closer to 1 indicating higher accuracy. In this case, both methods performed well, but ChatGPT Translation demonstrated a small but notable advantage in translation quality. Furthermore, Precision measures the

accuracy of the translations by evaluating how well the n-grams in the machine-generated translations match those in the reference translations.

The evaluation of translation of abstract 2 using BLEU method tool

Regarding Abstract 2, Table 7 compares the precision values of two translation methods, ChatGPT translation and Google Translate.

Table 7

Precision Values for AI- Methods of Translations of Abstract 2

N-grams / Methods of translations	ChatGPT translation	Google translate
Uni-gram Precision (P1)	0.92	0.85
Bi-gram Precision (P2)	0.87	0.76
Tri-gram Precision (P3)	0.82	0.68
Tetra-gram Precision (P4)	0.75	0.62

For unigram precision (P1), ChatGPT Translation achieved a precision of 0.92, while Google Translate scored slightly lower with 0.85. This indicates that ChatGPT Translation was more accurate in matching single words (unigrams) compared to Google Translate. Moving to bigram precision (P2), ChatGPT Translation maintained a higher precision of 0.87, whereas Google Translate scored 0.76, showing that ChatGPT Translation also performed better in matching pairs of consecutive words (bigrams). For trigram precision (P3), the trend continued, with ChatGPT Translation achieving a precision of 0.82 compared to Google Translate (0.68), demonstrating its superior ability to match sequences of three words (trigrams). Finally, for tetra-gram precision (P4), ChatGPT translation again outperformed Google Translate with a precision of 0.75, while Google Translate scored 0.62, indicating better performance in matching four-word sequences (tetra-grams). Overall, the table shows that ChatGPT translation consistently achieved higher precision values across all n-gram levels compared to Google Translate, suggesting that it produces translations that align more closely with the reference text. However, both methods produced reasonably accurate outputs. Thus, marking the results, ChatGPT Translation achieved a BLEU score of 0.84, slightly outperforming Google Translate, which scored 0.72, indicating better alignment with the reference translations.

The evaluation of translation of abstract 3 using BLEU method tool

Again, concerning this specific abstract, Table 8 compares the precision values of two translation methods, ChatGPT Translation and Google Translate.

Table 8

Precision Values for AI- Methods of Translations of Abstract 3

N-grams / Methods of translations	ChatGPT translation	Google translate
Uni-gram Precision (P1)	0.91	0.85
Bi-gram Precision (P2)	0.84	0.77
Tri-gram Precision (P3)	0.78	0.70
Tetra-gram Precision (P4)	0.71	0.63

As for unigram precision (P1), ChatGPT Translation achieved a precision of 0.91, while Google Translate scored slightly lower with 0.85. This indicates that ChatGPT Translation was more accurate in matching single words (unigrams) compared to Google Translate. Moving to bigram precision (P2), ChatGPT translation maintained a higher precision of 0.84, whereas Google Translate scored 0.77, showing that ChatGPT Translation also performed better in matching pairs of consecutive words (bigrams). For trigram precision (P3), the trend continued, with ChatGPT Translation achieving a precision of 0.78 compared to Google Translate recording 0.70, demonstrating its superior ability to match sequences of three words (trigrams). Finally, for tetra-gram precision (P4), ChatGPT Translation again outperformed Google Translate with a precision of 0.71, while Google Translate scored 0.63, indicating better performance in matching four-word sequences (tetra-grams). All in all, Table 8 shows that ChatGPT Translation consistently achieved higher precision values across all n-gram levels compared to Google Translate, suggesting that it produces translations that align more closely with the reference text. However, the results showed that ChatGPT Translation achieved a BLEU score of 0.81, outperforming Google Translate, which scored 0.72, indicating better alignment with the reference translations.

The evaluation of translation of abstract 4 using BLEU method tool

Finally, Table 9 displays the precision values of two translation methods, ChatGPT Translation and Google Translate.

Table 9

Precision Values for AI- Methods of Translations of Abstract 4

N-grams / Methods of translations	ChatGPT translation	Google translate
Uni-gram Precision (P1)	0.88	0.82
Bi-gram Precision (P2)	0.80	0.74
Tri-gram Precision (P3)	0.72	0.65
Tetra-gram Precision (P4)	0.65	0.58

For unigram precision (P1), ChatGPT translation achieved a precision of 0.88, while Google Translate scored slightly lower with 0.82. This indicates that ChatGPT Translation was more accurate in matching single words (unigrams) compared to Google Translate. Moving to bigram precision (P2), ChatGPT Translation maintained a higher precision of 0.80, whereas Google Translate scored 0.74, showing that ChatGPT Translation also performed better in matching pairs of consecutive words (bigrams). For trigram precision (P3), the trend continued, with ChatGPT Translation achieving a precision of 0.72 compared to Google Translate (0.65), showing its superior ability to match sequences of three words (trigrams). Finally, for tetra-gram precision (P4), ChatGPT Translation again outperformed Google Translate with a precision of 0.65, while Google Translate scored 0.58, indicating better performance in matching four-word sequences (tetra-grams). Overall, the table shows that ChatGPT Translation consistently achieved higher precision values across all n-gram levels compared to Google Translate, suggesting that it produces translations that align more closely with the reference text. All in all, based on the below equation, ChatGPT Translation achieved a BLEU score of 0.77, outperforming Google Translate, which scored 0.69, highlighting better alignment with the reference translations.

Final BLEU Score:

$$\text{BLEU} = BP \times \exp\left(\sum_{n=1}^{4} w_n \log(p_n)\right)$$

ChatGPT Translation:

$$\text{BLEU} = 1 \times \exp(-0.063375) \approx 0.9386$$

Google Translate:

$$\text{BLEU} = 1 \times \exp(-0.077525) \approx 0.9254$$

Discussion

The results from the translation analysis reveal varying performance across three translation methods—Human Translation, ChatGPT Translation, and Google Translate—on different types of content. The data, presented in terms of both raw counts and percentages, provides insight into the effectiveness and accuracy of these translation approaches across multiple categories, including material, relational, verbal, mental, behavioral and existential processes. It is important to note that the comparative analysis of process types across four abstracts, as presented in Tables 2 to 5, provides a detailed examination of human and AI-based translation systems (ChatGPT and Google Translate). Table 2, focusing on Abstract 1, shows that material processes dominate all translation methods, with human translation at 73.91%, ChatGPT at 78.26%, and Google Translate at 75%. Relational processes are less frequent, with human translation maintaining a higher proportion (13.04%) compared to AI systems (8.70% for ChatGPT and 8.33% for Google Translate). Mental and existential processes appear minimally, with AI systems slightly overrepresenting mental processes. Behavioral processes are entirely absent. As for Table 3, the analysis of translation process types across human, ChatGPT, and Google Translate outputs reveals distinct patterns in how each method approaches linguistic transformation. Human translation demonstrates a strong preference for material processes (71.4%), which focus on concrete actions and events, reflecting a direct and pragmatic rendering of the source text. This contrasts with ChatGPT and Google Translate, which also prioritize material processes (64.5% and 68.6%, respectively) but exhibit a higher frequency of relational processes (25.8% for ChatGPT vs. 11.5% for Google Translate). Relational processes, which describe states or attributes, suggest that ChatGPT is more likely to rephrase or interpret relationships between concepts, potentially introducing subtle shifts in meaning. Notably, human translation includes existential processes (7.2%), which highlight the mere existence of entities, a feature that is entirely absent in ChatGPT's output. Google Translate, however, retains some existential elements (5.7%), indicating partial preservation of the source text's philosophical or abstract nuances. The presence of a behavioral process (3.3%) in ChatGPT's translation, a category missing in both human and Google Translate versions, suggests an occasional tendency to anthropomorphize or emphasize human-like actions, possibly due to its training on diverse linguistic data. Moreover, verbal processes, which pertain to communication, are marginally higher in Google Translate (8.5%) compared to human (7.2%) and ChatGPT (6.4%), hinting at a slight bias toward framing content as reported speech or dialogue. Mental processes, involving cognition or perception, are minimal across all versions but appear only in human (3.5%) and Google Translate (5.7%), implying that AI translations may overlook introspective or subjective elements present in the original text. Thus, this result suggests that Abstract 2, like Abstract 1, is primarily action-oriented, with AI systems performing well in translating material processes. Nevertheless, Table 4, examining Abstract 3, reveals a shift in the distribution of process types. Material processes, while still dominant, are less prevalent, with human translation at 41.67%, ChatGPT at 61.54%, and Google Translate at 53.85%. Relational processes are more prominent, particularly in human translation (41.67%), compared to ChatGPT (23.08%) and Google Translate (30.77%). This indicates that Abstract 3 contains more relational content, which human translators handle more effectively. Mental processes appear in small proportions across all systems, while behavioral and existential processes remain absent. A return to the dominance of material categories is highlighted in Table 5 that focuses on Abstract 4. It shows this trait of dominance of material

processes, with human translation at 69.44%, ChatGPT at 67.57%, and Google Translate at 65%. Relational processes are more frequent than in Abstracts 1 and 2, with human translation at 19.44%, ChatGPT at 18.92%, and Google Translate at 20%. Mental processes are slightly overrepresented in AI translations, while existential processes appear minimally in human translation. Behavioral processes are noticeably absent, consistent with the other abstracts except abstract 2.

Overall, the tables highlight that material processes are the most common across all abstracts, with AI systems performing well in translating them. This finding is consistent with other studies (see: Alhumsi & Alsaedi, 2023; Alhumsi et al., 2024). However, human translators excel in handling relational and mental processes, which require greater contextual understanding. AI systems, particularly ChatGPT and Google Translate, occasionally overrepresent mental processes, suggesting limitations in their ability to capture nuanced linguistic elements. These findings underscore the strengths of AI systems in handling straightforward, action-oriented content while emphasizing the continued importance of human expertise for more complex translation tasks. In conclusion, the comparative analysis reveals that AI-based translation systems are highly effective for translating material processes but may struggle with more nuanced process types such as relational and mental processes. Human translators remain superior in handling these complex processes, underscoring the importance of human expertise in translation tasks. These findings have significant implications for the use of AI in translation, particularly in contexts where nuanced understanding and contextual accuracy are critical. This provides more support to Lu's (2024) study that argued that human translation still holds an advantage in some areas. This is because humans possess nuanced understanding, cultural knowledge, and context awareness that machines can't fully replicate.

As for the second research question, the results from the four abstracts provide a comparative analysis of the precision scores for two translation methods—ChatGPT and Google Translate—across different n-gram levels, including uni-grams, bi-grams, tri-grams, and tetra-grams. These precision scores reflect the accuracy of the translations when compared to a human reference translation. The findings reveal several important trends and insights into the performance of these two translation tools. Thus, across all four abstracts, ChatGPT consistently outperforms Google Translate in terms of precision. For instance, in Abstract 1, ChatGPT achieves a uni-gram precision (P1) of 0.93, while Google Translate scores 0.89. This pattern holds true for bi-grams, tri-grams, and tetra-grams, with ChatGPT maintaining a slight but consistent lead over Google Translate. This suggests that ChatGPT's advanced language modeling capabilities may enable it to better capture the context and nuances of the source text, resulting in higher accuracy.

However, both translation methods exhibit a decline in precision as the n-gram size increases. For example, in Abstract 1, ChatGPT's precision drops from 0.93 for uni-grams to 0.72 for tetra-grams, while Google Translate's precision decreases from 0.89 to 0.68 over the same range. This trend is consistent across all four abstracts, indicating that translating longer and more complex sequences of text poses a greater challenge for both tools. The decline in precision with higher n-grams may be attributed to the increased likelihood of errors compounding in longer text segments. Moreover, the results are highly consistent across the four abstracts, demonstrating that the observed trends are not specific to a single dataset or context. For instance, ChatGPT uni-gram precision ranges between 0.88 and 0.93 across the abstracts, while Google Translate uni-gram precision ranges between 0.82 and 0.89. This consistency reinforces the reliability of the findings and suggests that both tools perform predictably across different texts.

Despite ChatGPT's superior performance, the differences in precision between the two methods are relatively small. This indicates that while ChatGPT has a slight edge in accuracy, Google Translate remains a strong and competitive alternative, particularly for shorter texts or less critical applications. The marginal differences also highlight the advancements in machine translation technology, as both tools achieve high levels of precision. The implications of these findings are significant for practical applications. For tasks requiring high precision, such as legal or medical translations, ChatGPT may be the preferred choice due to its slightly better performance. On the other hand, Google Translate remains a viable option for general use, especially when dealing with shorter texts or less complex translation tasks. The decline in precision with higher n-grams underscores the importance of considering the complexity of the source text when selecting a translation tool. It is important to note that the study is based solely on precision scores and does not account for other metrics such as recall, fluency, or semantic accuracy. Additionally, the source text is in Arabic, and the findings may not generalize to other languages or translation tasks

Conclusion

The transmission of information across languages has undergone significant transformation due to the advancements in translation technologies, whether facilitated by human translators or artificial intelligence (AI). This study aims to compare the efficacy of AI-based translation systems, such as ChatGPT and Google Translate, with that of human translators in rendering research abstracts. To evaluate their performance, Halliday's transitivity system model, with a focus on critical discourse analysis, was employed to examine the transitivity processes involved in each translation.

The comparative analysis of four abstracts across human and machine translations reveals consistent patterns in process type distributions, with notable variations in how each translation method handles linguistic representation. Human translations maintain the highest fidelity to source text intentionality, particularly in preserving existential processes (absent in most AI outputs) and demonstrating more balanced distributions between material and relational processes. Both ChatGPT and Google Translate show a strong preference for material processes often at the expense of relational and mental processes, suggesting a tendency toward action-oriented, concrete renditions. Relational processes, critical for expressing abstract connections, are consistently reduced in AI translations, indicating a limitation in capturing nuanced semantic relationships. Google Translate occasionally aligns more closely with human translations in mental processes but both AI systems inconsistently neglect existential elements.

On the other hand, this study employed BLEU (Bilingual Evaluation Understudy) to assess the linguistic accuracy of the three translation versions, namely Human Translation, ChatGPT Translation, and Google Translation. Based on the analysis of BLEU, it has been found that both ChatGPT and Google Translate relatively achieve high precision values, with ChatGPT consistently outperforming Google Translate in all n-grams precision measures. This trend persists across all abstracts and n-gram levels. However, precision values decrease as the n-gram size increases, with tetra-gram precision (P4) being the lowest for both systems (e.g., 0.72 for ChatGPT and 0.68 for Google Translate in Abstract 1). This pattern indicates that while both systems excel at capturing simpler, shorter phrases, they face greater challenges in accurately translating longer and more complex linguistic structures. This inference suggests that AI translation systems like ChatGPT and Google Translate are highly effective at handling

straightforward, word-level translations but struggle to maintain the same level of accuracy for more intricate, context-dependent phrases. The consistent superiority of ChatGPT across all levels of n-grams further highlights its advanced capability in aligning with human reference translations, making it a slightly more reliable choice for producing linguistically accurate translations. Nonetheless, the decreasing precision with larger n-grams underscores the ongoing need for improvements in AI systems to better handle complex linguistic patterns.

This paper has several limitations that should be acknowledged. Firstly, the analysis is limited to a small sample size of four abstracts. Secondly, the evaluation of translation quality relies heavily on BLEU scores, which, while useful for measuring n-gram precision, do not account for semantic accuracy, cultural nuances, or contextual appropriateness. This means that while the translations may be linguistically precise, they may not always convey the intended meaning effectively. Thirdly, the study focuses only on Arabic-to-English translations, limiting its applicability to other language pairs that may present unique challenges. Additionally, the analysis of transitivity processes is confined to a specific linguistic framework, which may not capture holistic translation quality assessment. Finally, the study does not explore the impact of post-editing or hybrid human-AI approaches, which could significantly influence translation outcomes. Thus, further studies could explore a broader range of text genres and languages, examine error patterns and stylistic variations, and evaluate how post-editing impacts machine-translated outputs to validate these findings.

Acknowledgements

This study uses ChatGPT for the readability of the whole text.

Funding

The author did not receive funding from any organization for this study.

Conflict of Interest:

The author declares that there is no conflict of interest.

Bio

Dr. Mohammad Husam Alhumsi *holds a PhD in Applied Linguistics from University Utara Malaysia. He is an Associate Professor in the College of Sciences and Theoretical Studies at Saudi Electronic University (SEU). Moreover, he teaches multiple courses to undergraduate students in the Department of English Language and Translation. His research interests include Phonology and Phonemics, Discourse Analysis, Pragmatics, Beginning Reading Reform, Word Recognition, Second language Acquisition, E-Learning, and Education and Technology.*

References

Abercrombie, G. (2016, May). A rule-based shallow-transfer machine translation system for Scots and English. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16). Portorož, Slovenia, p. 578–584. https://www.aclweb.org/anthology/L16-1092

Alhumsi, M. H., & Alshagrawi, S. (2022). Transitivity analysis of news texts on the Saudi ministry of health website: An investigation of health guidelines during COVID-19.

International Journal of Advanced and Applied Sciences, 9(9), 9-16. https://doi.org/10.21833/ijaas.2022.09.002

Alhumsi, M. H., & Alsaedi, N. S. (2023). A transitivity analysis of two political articles: An investigation of gender variations in political media discourse. *World Journal of English Language, 13*(6), 107-118. https://doi.org/10.5430/wjel.v13n6p107

Alhumsi, M. H., Alsaedi, N. S., Fallata, S., & Alharbi, A. (2024). Transitivity Analysis of an Academic Document: A Case of English Course Syllabi at Saudi Electronic University. *Pakistan Journal of Life & Social Sciences, 22*(2), 14757-14780. https://doi.org/10.57239/PJLSS-2024-22.2.001062

Almahasees, Z. (2020). Diachronic evaluation of Google Translate, Microsoft translator and Sakhr in English-Arabic translation. [Unpublished Doctoral Dissertation], the University of Western Australia, Australia. https://doi.org/10.26182/5e2a3b3d81f7a

Almanna, A. (2020). Transitivity System as a Tool for Producing (In) accurate Mental Images through Translation. *Journal of Educational and Social Research, 10*(1), 23-23. https://doi.org/10.36941/jesr-2020-0003

Almoibed, O. I (2024). Criminal Liability Arising from the use of Artificial Intelligence technologies in the Saudi legal system: A comparative analytical study. *Arab Journal for Security Studies 40,* (2), 143-157. https://doi.org/10.26735/MKQW3049

Alqahtani, S. M. (2024). The comprehension of English adjective order by Saudi EFL learners. *University of Tabuk Journal for Humanities and Social Sciences, 4* (3), 187-206. Retrieved March 3, 2025, from https://www.ut.edu.sa/ar/Journal/Pages/default.aspx

Alshahrani, S. A. (2024). The legal Protection of Unfair Commercial Practices-An Analytical Study on the Anti-Commercial Fraud Law. *Jazan University Journal of Human Sciences, 12* (2), https://doi.org/10.37575/h/edu/22002

Amin, F. & Hassan M. (2023). Examining the effect of banking performance on the economic growth of Saudi Arabia: A Panel ARDL Approach. *Journal of Economic Studies, 15* (1), 51-65. https://doi.org/10.33948/ESJ-KSU-15-1-2

Bahdanau, D., Cho, K., & Bengio, Y. (2015, May). "Neural Machine Translation by Jointly Learning to Align and Translate," In 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, Conference Track Proceedings, Bengio Yoshua and Yann LeCuneds. https://doi.org/10.48550/arXiv.1409.0473

Brown, P. F., Della Pietra, S. A., Della Pietra, V. J., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational linguistics, 19*(2), 263-311. https://aclanthology.org/J93-2003.pdf

Bouamor, H., Alshikhabobakr, H., Mohit, B., & Oflazer, K. (2014, October). A human judgement corpus and a metric for Arabic MT evaluation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP) Doha, Qatar, (pp. 207-213). https://doi.org/10.3115/v1/D14-1026

Chen, B. & Kuhn, R. (2011, July) AMBER: A Modified BLEU, Enhanced Ranking Metric. In Proceedings of the Sixth Workshop on Statistical Machine Translation, pp. 71–77,

Edinburgh, Scotland. Association for Computational Linguistics. https://aclanthology.org/W11-2105/

Dai, X. (2024). Comparative analysis of artificial intelligence translation and human translation from the perspective of international communication—taking the Chinese translation of "Dream of Autumn" as an Example. *Lecture Notes on Language and Literature, 7*(4), 36-42. https://doi.org/10.23977/langl.2024.070406

El-Dakhs, D. A. (2018). Comparative genre analysis of research article abstracts in more and less prestigious journals: Linguistics journals in focus. *Research in Language (RiL), 16*(1), 47-63. https://doi.org/10.2478/rela-2018-0002

Federspiel, F., Mitchell, R., Asokan, A., Umana, C., & McCoy, D. (2023). Threats by artificial intelligence to human health and human existence. *BMJ global health, 8*(5), e010435. https://doi.org/10.1136/bmjgh-2022-010435

Frankenberg-Garcia, A. (2022) Can a corpus-driven lexical analysis of human and machine translation unveil discourse features that set them apart? *Target 34*(2):278–308. https://doi.org/10.1075/target.20065.fra

Fu, L., & Liu, L. (2024). What are the differences? A comparative study of generative artificial intelligence translation and human translation of scientific texts. *Humanities and Social Sciences Communications, 11*(1), 1-12. https://doi.org/10.1057/s41599-024-03726-7

Ghassemiazghandi, M. (2024). An Evaluation of ChatGPT's Translation Accuracy Using BLEU Score. *Theory and Practice in Language Studies, 14*(4), 985-994. https://doi.org/10.17507/tpls.1404.07

Hadla, L. S., Hailat, T. M., & Al-Kabi, M. N. (2014). Evaluating Arabic to English machine translation. *International Journal of Advanced Computer Science and Applications, 5*(11), 68-73. https://doi.org/10.14569/IJACSA.2014.051112

Halliday, M. A. & Matthiessen, C. M. (20142004) *Halliday's introduction to functional grammar*. 4th edition. New York: Routledge

Halliday, M. A. & Webster, J. J. (2014). *Text linguistics: The how and why of meaning*. Equinox Publishing Ltd.

Han, L. (2016). Machine translation evaluation resources and methods: A survey. arXiv:1605.04515. Computation and language. Cornell University Library. https://doi.org/10.48550/arXiv.1605.04515

Hyland, K. (2000). *Disciplinary Discourses: Social Interaction in Academic Writing*, Pearson, London, UK.

Jiang, C. (2010). Quality assessment for the translation of museum texts: Application of a systemic functional model. *Perspectives, 18*(2), 109-126. http://dx.doi.org/10.1080/09076761003678734

Köhler, R. (2012). *Quantitative syntax analysis*. Walter de Gruyter.

Kembaren, F. R., Hasibuan, A. K., & Natasya, A. (2023). Technology Trends in Translation: A Comparative Analysis of Machine and Human Translation. Absorbent Mind: *Journal*

of Psychology and Child Development, 3(2), 169-183. https://doi.org/10.37680/absorbent_mind.v3i2.4486

Klebanov, B. B., & Flor, M. (2013, August). Associative texture is lost in translation. In: Proceedings of the Workshop on Discourse in Machine Translation, Association for Computational Linguistics, Sofia, Bulgaria, pp 27–32. https://aclanthology.org/W13-3304

Lin, R., Zhao, M., Guo, J., Liu, W., Jin, Q., & Ma, Z. (2024). On the transitivity analysis of research article abstracts from science. *International Journal of Languages, Literature and Linguistics, 10* (2), 173-177. https://doi.org/10.18178/IJLLL.2024.10.2.507

Lu, Y. (2024). Comparative study of machine translation versus human translation. *Lecture Notes on Language and Literature, 7*(2), 61-66. https://doi.org/10.23977/langl.2024.070211

Madi, N., & Al-Khalifa, H. (2020). Error detection for Arabic text using neural sequence labeling. *Applied Sciences, 10*(15), 5279. https://doi.org/10.3390/app10155279

Marouani, M., Boudaa, T., & Enneya, N. (2020). Statistical error analysis of machine translation: The case of Arabic. *Computación y Sistemas, 24*(3), 1053-1061. https://doi.org/10.13053/cys-24-3-3289

Mehawesh, M. (2023) Human translation vs. Machine translation in the Naguib Mahfouz's novel palace walk: a case study. *MJHSS 38*(2), 13-33. https://doi.org/10.35682/mjhsc.v38i2.631

Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon, 10*(6), e28106. https://doi.org/10.1016/j.heliyon.2024.e28106

Muhdi, M. (2019). Framework for implementation of education policy in the perspective of education management in Indonesia. *Universal Journal of Educational Research, 7*(12), 2717–2728. https://doi.org/10.13189/ujer.2019.071220

Niu, J., Jiang, Y., & Zhou, Y. (2020). Approaching textual coherence of machine translation with complex network. *International Journal of Modern Physics C, 31*(12), 2050175. https://doi.org/10.1142/s0129183120501752

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and policy in mental health, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y

Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002, July). Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, Philadelphia, PA, USA. (pp. 311-318). https://doi.org/10.3115/1073083.1073135

Passban, P., Way, A., & Liu, Q. (2018, August). Tailoring neural architectures for translating from morphologically rich languages. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, pp. 3134-3145. https://aclanthology.org/C18-1265

Ranathunga, S., Lee, E. S., Prifti Skenduli, M., Shekhar, R., Alam, M., & Kaur, R. (202). Neural machine translation for low-resource languages: A survey. *ACM Computing Surveys, 55*(11), 1-37. https://doi.org/10.48550/arXiv.2106.15115

Reiter, E. (2018). A structured review of the validity of BLEU. *Computational Linguistics, 44*(3), 393-401. https://doi.org/10.1162/coli_a_00322

Sai, A. B., Mohankumar, A. K., & Khapra, M. M. (2020). A survey of evaluation metrics used for NLG systems. *arXiv:2008.12009, 55*(2), 1-55. https://doi.org/10.48550/arXiv.2008.12009

Saidi, M., & Talebi, S. (2021). Genre analysis of research article abstracts in English for academic purposes journals: Exploring the possible variations across the venues of research. *Education Research International, 2021*(1), 3578179. https://doi.org/10.1155/2021/3578179

Sari, C. P., & Sari, M. E. (2016). Transitivity in linguistics abstracts papers of 2nd LLTC by ELESP. *Indonesian Journal of English Language Studies (IJELS), 2*(2), 143-156. https://doi.org/10.24071/ijels.v2i2.555

Sianipar, I. F., & Sajarwa, S. (2021). The translation of Indonesian passive voice in research articles' abstracts into English: Human Vs machine translation. Lingua Didaktika: *Jurnal Bahasa dan Pembelajaran Bahasa, 15*(2), 136-148. https://doi.org/10.24036/ld.v15i2.112967

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. https://doi.org/10.3390/info14100574

Supriatnaningsih, R., Mr, R., Hariri, T., & Astini, E. (2019). Politeness In Students' Speeches When Speaking Japanese With Native Speakers. UNNES International Conference on English Language Teaching, Literature, and Translation (ELTLT 2018), 235–239. https://doi.org/10.2991/eltlt-18.2019.47

Thompson, G. (2000) *Introducing functional grammar*. *London*: Edward Arnold.

Vanmassenhove, E., Shterionov, D., & Way, A. (2019). Lost in translation: Loss and decay of linguistic richness in machine translation. *arXiv preprint arXiv*:1906.12068. https://doi.org/10.48550/arXiv.1906.12068

Vathanalaoha, K., & Tangkiengsirisin, S. (2018). Transitivity analysis of rhetorical moves in dental research article abstracts: Thai and international journals. *Humanities, Arts and Social Sciences Studies, 18*(3), 639-662. https://doi.org/10.14456/hasss.2018.28

Wang, Y. (2024). Research of types and current state of machine translation. *Applied and Computational Engineering, 37*, 95-101. https://doi.org/10.54254/2755-2721/37/20230479

Watson, D., Zalmout, N., & Habash, N. (2018, November). Utilizing Character and Word Embeddings for Text Normalization with Sequence-to-Sequence Models. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics. Brussels, Belgium, pp. 837–843. https://doi.org/10.18653/v1/D18-1097

Wong, B. T., & Kit, C. (2012, July) Extending machine translation evaluation metrics with lexical cohesion to document level. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, Association for Computational Linguistics, Jeju Islan d, Korea, pp. 1060–1068. https://aclanthology.org/D12-1097

Zakraoui, J., Saleh, M., Al-Maadeed, S., & Alja'am, J. M. (2021). Arabic machine translation: A survey with challenges and future directions. *IEEE Access, 9*, 161445-161468. https://doi.org/10.1109/ACCESS.2021.3132488

Zagood, M. J., Al-Nuaimi, A., & Al-Blooshi, A. (2021). Man vs machine: A comparison of linguistic, cultural, and stylistic levels in literary translation. *International Journal of Linguistics, Literature and Translation, 4*(2), 70-77. https://doi.org/10.32996/ijllt.2021.4.2.10