System 130 (2025) 103618

Contents lists available at ScienceDirect

System

journal homepage: www.elsevier.com/locate/system

GenAI as a translation assistant? A corpus-based study on lexical and syntactic complexity of GPT-post-edited learner translation

Ho Ling Kwok, Yining Shi, Han Xu, Dechao Li, Kanglong Liu *

Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China

ARTICLE INFO

Keywords:
Generative artificial intelligence
Learner translation
Lexical complexity
Syntactic complexity

ABSTRACT

The advent of generative artificial intelligence (GenAI) models, most notably ChatGPT in late 2022, marked a significant milestone in AI development, attracting widespread attention from various research fields. Among its emerging applications, GenAI demonstrates potential in translation education. This study examines the role of GenAI as a post-editing assistant in learner translation by comparing the lexical and syntactic complexity of second language (L2) translations produced by Hong Kong students, with and without post-editing by GPT. The analysis revealed that GPT post-editing improved lexical complexity in learner translations, though its effect on syntactic complexity was inconsistent. While GPT post-editing resulted in longer clauses, more complex nominals, and an increased use of coordinate phrases, non-edited translations featured greater subordination and more verbal structures. These findings suggest that GenAI holds promise in enhancing translation practice but also highlight the need for critical AI literacy to ensure effective use in translation education, particularly in advancing students' linguistic and instrumental competence.

1. Introduction

The launch of ChatGPT, a generative artificial intelligence (GenAI) chatbot, in late 2022 marked a significant milestone in the evolution of AI technologies. Built on a large language model from the GPT-3.5 series, ChatGPT became a widely used platform for human-machine interaction, assisting with various tasks. Its capabilities have garnered interest across multiple sectors. Since its release, many technology companies have invested in advancing GenAI models, leading to the availability of models such as Gemini, Claude, and ERNIE Bot. GenAI has significantly impacted industries and professions, particularly those focused on language-related tasks.

The impact of GenAI on language teaching and learning has drawn increasing attention in the field of language education. Scholars have recognized its potential to enhance learners' language skills and have advocated for integrating these technologies into pedagogical practices to improve both teaching methods and learner outcomes (e.g., Guo et al., 2022; Guo & Wang, 2023; Wiethof et al., 2021). As a result, a growing number of higher education institutions are becoming more open to adopting GenAI tools in their teaching environments.

In addition to language learning, there is increasing debate surrounding the use of GenAI tools in translation practice. Central to the discussion is the potential of GenAI tools to serve as alternatives to traditional machine translation (MT) systems, with particular

* Corresponding author. Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China.
E-mail addresses: ho-ling.kwok@connect.polyu.hk (H.L. Kwok), yn-zoe.shi@polyu.edu.hk (Y. Shi), han12.xu@polyu.edu.hk (H. Xu), dechao.li@polyu.edu.hk (D. Li), kl.liu@polyu.edu.hk (K. Liu).

https://doi.org/10.1016/j.system.2025.103618
Received 3 October 2024; Received in revised form 9 February 2025; Accepted 9 February 2025
Available online 10 February 2025
0346-251X/

attention given to the quality and accuracy of their output (e.g., Bistafa, 2024; Peng et al., 2023). As the GenAI revolution progresses, these advancements have the potential to transform the workflows of professional translators as well as influence translation learners' approaches. However, despite the growing interest in GenAI, few studies have explored its application specifically within the context of learner translation or translation education. This gap highlights the need for further investigation into how GenAI tools might support or hinder the development of translation skills in educational settings.

This study aims to investigate second language (L2) learner translations, both with and without GPT post-editing, through a corpus-based methodology. Specifically, it builds on the simplification hypothesis, which posits that translations are often less complex than original texts in the same language. The research examines the potential impact of GPT post-editing on the lexical and syntactic complexity of learner translations. The findings are expected to contribute to both theoretical and practical domains, offering a deeper understanding of how GenAI post-editing may influence translation simplification. The study aims to explore how GenAI tools, through an analysis of GPT-edited translations as opposed to learner-generated translations, could support translation education. By addressing a gap in the literature, this research seeks to examine potential applications of GenAI in educational contexts.

The remainder of this paper is structured as follows: the next section sets the scene for this study by reviewing the fundamental concepts of GenAI and the features of translational languages. The Methods section outlines the procedures for corpus compilation, complexity feature extraction, and data analysis. The main findings are presented in the Results section, followed by a discussion in the subsequent section, which interprets the results through the lens of AI footprint. The final section concludes the study and offers suggestions for future research.

2. Literature review

2.1. Generative artificial intelligence (GenAI)

GenAI refers to a category of artificial intelligence designed to generate original content, such as text, images, and other data types, in response to user prompts. A key example of this technology is the Generative Pre-trained Transformer (GPT) series, a collection of large language models (LLMs) developed by OpenAI. These models are trained on extensive datasets sourced from the internet, enabling them to recognize patterns, contextual cues, and relationships within the data (Brown et al., 2020; Johri, 2023). By leveraging advanced deep learning and natural language processing (NLP) techniques, GPT models can understand language and generate human-like text (Haleem et al., 2022). As a result, GenAI tools represent a significant breakthrough in AI, with wide-ranging applications, particularly in language-related fields.

GenAI services have sparked extensive discussions about their language capabilities and human-machine interaction. Although these models are not designed for specific tasks, they generate predictions based on vast amounts of training data containing diverse text types from various language activities (Johri, 2023). As a result, they can support a wide array of language-related tasks, including text generation and completion (Wu et al., 2023), editing and proofreading (Al Sawi & Alaa, 2024), and translation (Al Sawi & Allam, 2024; Jiao et al., 2023). GenAI tools demonstrate significant potential in transforming linguistic and communication processes.

In the realm of translation studies, research has increasingly focused on the potential of GenAI models as translation tools and evaluated their performance. Some findings indicate that models like GPT can achieve translation quality comparable to conventional MT tools, particularly in terms of accuracy and fluency (Bistafa, 2024; Gao et al., 2024; Hendy et al., 2023; Jiao et al., 2023). Moreover, unlike traditional MT systems, GenAI offers users the ability to input prompts to refine the translation output, introducing a more interactive and adaptive process. This feature highlights the evolving nature of translation technologies and their potential to reshape translation practices.

Despite its advancements, GenAI-generated translation faces several criticisms, particularly regarding the variability of its output quality, which is influenced by factors such as the GenAI model, language pairs, text types, prompt design, and parameter settings. For instance, while the translation quality of GPT-3.5 diminishes when handling distant or low-resource languages, the more recent GPT-4 model has shown significant improvements, making it more competitive with conventional MT tools (Jiao et al., 2023). However, challenges remain, especially when dealing with certain text types. GPT models are generally less effective for highly technical or specialized texts, such as medical reports, legal documents, and culturally nuanced content like literary works (Khoshafah, 2023). Additionally, prompt engineering plays a critical role in shaping translation quality. Providing clear information about the translation's purpose and target audience can enhance output, whereas overly complex or irrelevant prompts may result in word-for-word translations that degrade quality (Peng et al., 2023; Yamada, 2023). Parameter settings, such as temperature values, also affect performance, with lower settings typically yielding better results (Peng et al., 2023). Given these limitations, some argue that GPT models are more suited for post-editing translations rather than fully replacing traditional translation tools (Sahari et al., 2023). Overall, these ongoing discussions underscore both the potential and limitations of GenAI in translation, highlighting the need for further research to refine its role in this evolving field.

2.2. Translation

Translational language is often regarded as a "third code" (Frawley, 1984), distinct from both the source and target languages. This uniqueness stems from its specific production process, which differs from that of native language use. Unlike native language production, translation and L2 communication are frequently described as "constrained communication," meaning that these processes operate under specific limitations (Lanstyák & Heltai, 2012). These constraints arise primarily from psycholinguistic and social factors, such as the need to activate two distinct languages simultaneously (Grosjean, 2013), the suppression of interference from the source text (Toury, 2012), and the requirement to follow the linguistic norms of the target culture (Kruger & van Rooy, 2016). The constant bilingual and bi-directional switching between two communicative contexts places additional cognitive demands on translators. As a result, the interplay between linguistic constraints, social expectations, and cognitive load shapes the distinctive characteristics of translational language.

In the field of translation studies, analyzing the linguistic features of translated texts is a central area of inquiry. Baker (1993) introduced the translation universals (TU) hypothesis, which seeks to identify "universal features of translation" that result from the act of translation itself, independent of any specific language system. Building on this idea, Chesterman (2004) proposed the concepts of S-universals and T-universals. S-universals refer to "universal differences between translations and their source texts," while T-universals describe differences between translations and non-translations in the target language. However, the notion of "universals" has sparked significant debate. Some scholars argue that the features of translated texts are influenced by factors such as genre, language pairs, and the direction of translation (e.g., House, 2016; Pym, 2008; Tymoczko, 1998). Despite these debates, research continues to explore the universals hypothesis, now understood in a probabilistic sense, implying recurring patterns under certain conditions rather than absolute rules (Tsai, 2021). Baker's hypothesis has shifted the focus of scholarly attention from parallel text comparisons to studying the characteristics of translational language, particularly through the use of comparable corpora to analyze translated texts alongside non-translated ones.

One widely discussed TU candidate is simplification, the inclination to subconsciously simplify translational language (Baker, 1996). The simplification hypothesis, classified as a T-universal, suggests that translated texts are generally simpler than non-translated native texts in the target language (Chesterman, 2004). This phenomenon has been extensively studied at both the lexical (Ferraresi et al., 2018; Laviosa, 1998; Wen, 2009) and syntactic (Chen et al., 2024; Liu & Afzaal, 2021; McWhorter, 2011) levels. While Ferraresi et al. (2018) rejected the simplification hypothesis in their study of English European Parliament texts translated from French and Italian, many other studies have supported it. For example, translated English narrative prose and Chinese fiction have been shown to display lower lexical richness compared to their non-translated equivalents (Laviosa, 1998; Wen, 2009; Xiao & Yue, 2009). Similarly, translated English tends to be syntactically simpler than non-translated texts (Chen et al., 2024; Liu & Afzaal, 2021). These findings suggest that simplification is a common feature across various translation contexts, though its expression may differ depending on language pairs and text types.

While there has been research on learner translations, it remains less extensive compared to the body of work on professional translations. One notable study in this area is by Kunilovskaya et al. (2018), which found that both learner and professional translations exhibit distinct linguistic features that differentiate them from non-translated native texts. However, learner translations showed more pronounced deviations and simplifications, particularly in lexical density, lexical variety, word form frequency distribution, and sentence length. These findings highlight the unique characteristics of learner translations and their divergence from non-translated texts.

In addition to discussions about translation expertise, the issue of translation directionality, particularly translating into a non-native language (L2 translation), has attracted increasing attention. The growing dominance of English globally has led translators to more frequently work in their L2, translating from their native language (L1) into L2 English. Research generally suggests that L1 translations tend to be of higher quality than L2 translations, possibly due to limited proficiency in the non-native target language (Pavlović, 2013; Samuelsson-Brown, 2010). Recent empirical studies, such as Penha-Marion et al. (2024), support these findings, showing lower lexical complexity in L2 student translations compared to L1 translations. Interestingly, they also observed longer mean sentence lengths in L2 translations. These findings highlight the unique challenges and complexities of translating into a non-native language.

In summary, L2 translation, particularly those performed by learner translators, may represent a distinct form of translation that combines features of both translation and L2 production, as well as non-expert task performance. The observed simplification in L2 translations raises important questions about potential interventions. With the rise of advanced GenAI tools like GPT, it becomes compelling to explore whether these technologies can help improve L2 learner translation expression by addressing the issue of simplification.

2.3. Research questions

The literature reveals ongoing debates about AI-assisted and AI-generated translation. Previous research has primarily focused on GenAI as a potential alternative to traditional MT tools, with an emphasis on examining the quality of its translation output. Sahari et al. (2023) suggest that GPT may be more effective in editing translated texts rather than performing the full translation. However, the role of GenAI as a post-editing assistant has not been widely explored. In the context of L2 learner translation, where challenges related to both L2 production and task expertise often result in language simplification, it remains to be seen whether GenAI tools like GPT can help improve translation expression by addressing this issue.

To address the research gap, this study aims to explore how GPT editing may influence the linguistic complexity of L2 translations produced by Hong Kong students. Specifically, it compares translations with and without GPT editing to assess potential changes in learner translation expression. The study addresses the following research questions.

RQ1: Are there differences in lexical complexity between L2 learner translations with GPT editing and those without GPT editing?
RQ2: Are there differences in syntactic complexity between L2 learner translations with GPT editing and those without GPT editing?

3. Methods

3.1. Design

This study employed a corpus-based methodology to examine and compare the linguistic complexity of two sets of L2 learner translations: those with GPT editing and those without. A mixed-methods design was used to address the research questions comprehensively. The quantitative analysis focused on nine lexical complexity indices and 14 syntactic complexity indices, while the qualitative analysis revisited specific text examples to provide context for the quantitative results. As Larsson et al. (2022) noted, "[…] quantifying linguistic phenomena forces us away from our primary object of study, namely language itself" (p. 152). This mixed-methods approach is commonly applied in corpus-based research to provide a more complete interpretation of the findings. The study was conducted in three stages: corpus compilation, feature extraction and selection, and data analysis (Fig. 1).

3.2. Corpus compilation

The two self-compiled corpora, the Parallel Learner Translation Corpus – Original (PLTC–O) and the Parallel Learner Translation Corpus – GPT Edited (PLTC–GPT), represent original L2 learner translations and GPT post-edited L2 learner translations, respectively. Both corpora are multi-genre, including texts from academic writing, popular writing, reportage, instructional writing, persuasive writing, and creative writing.

PLTC–O documents English translations produced by learner translators in Hong Kong, across various written genres. Specifically, it includes translations from L1 Chinese to L2 English, created by undergraduate students majoring in translation during their second to fourth years of study. These students, who received most of their education in Hong Kong, predominantly used Cantonese as their L1 or primary communicative language. The participant recruitment process ensured that all learner translators had similar language and educational backgrounds. Before translating, they were provided with a brief that outlined the target audience (native English speakers), allowed translation tool usage, and detailed the required register of the translated text. For this study, only translations completed without the use of GenAI tools were selected. This resulted in 85 texts for analysis (see Kwok et al. (2023) for corpus compilation details).

PLTC–GPT contains GenAI post-edited learner translations based on GPT outputs. The corpus was compiled using a zero-shot prompt approach for GPT post-editing, intentionally avoiding the use of editing examples to assess GPT's natural editing capabilities without additional training. Instructions for the editing task, along with the source text for reference, were provided to GPT, which then post-edited the learner translations from PLTC–O. These instructions mirrored the translation brief given to students for PLTC–O, ensuring comparability between the two corpora. Specifically, GPT was prompted with details regarding the target audience and the appropriate register for the text. For this study, GPT-3.5 was used, as it represents a foundational model in the GenAI landscape and is significant for examining how early-stage GenAI can perform in translation post-editing. While future advancements in GenAI are expected, analyzing the performance of GPT-3.5 offers a useful benchmark for understanding the progression of these technologies. The parameter settings for the GPT model are outlined in Table 1. Since PLTC–GPT is directly derived from PLTC–O, the two corpora are comparable in terms of modality (written mode), number of texts (85 texts in each), genre (six categories), and target language (English). Further details on both corpora are summarized in Table 2.

3.3. Linguistic complexity features

This study aims to compare GPT-edited translations with L2 learner translations, focusing on two key linguistic dimensions: lexical and syntactic complexity. To accomplish this, we utilized established NLP tools to calculate nine lexical complexity indices and 14 syntactic complexity indices for each text.

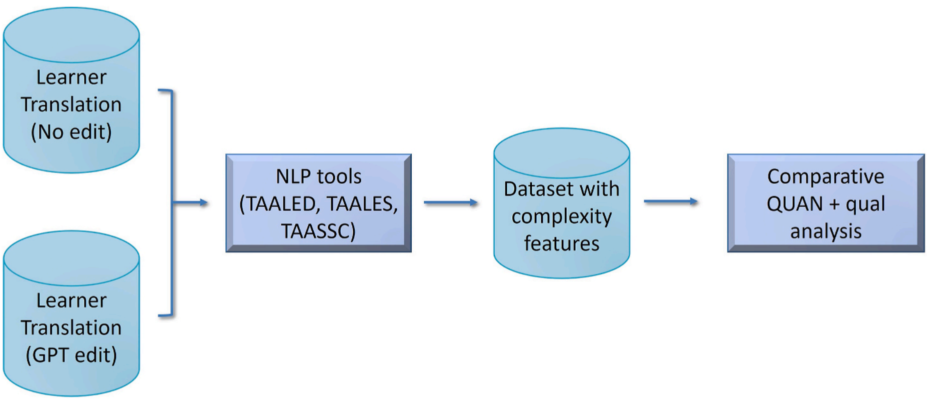

Fig. 1. Research procedures.

Table 1
GPT setup.

Parameter	Settings
Model name	gpt-35-turbo-16k
Model version	0613
Temperature	0.3
Top-p	1
Frequency penalty	0
Presence penalty	0

Note. The temperature and top-p settings were based on Moslem et al. (2023). Default values were used for frequency penalty and presence penalty settings.

Table 2
Details of the corpora.

Corpora	Nature	Genres	Texts	Tokens	Types	STTR
PLTC–O	L2 learner translations without the aid of GenAI	Academic writing, popular writing, reportage, instructional writing, persuasive writing, creative writing	85	202,503	15,255	41.44
PLTC–GPT	GPT post-edited L2 learner translations		85	195,512	14,750	42.46

Note. STTR (standardized type-token ratio) was calculated on the basis of 1000 words.

3.3.1. Lexical complexity indices

Lexical complexity indices. Following Bulté and Housen (2012), Bui (2021), and Zhang and Lu (2024), lexical complexity is conceptualized as a multi-dimensional construct comprising three dimensions: lexical density, diversity, and sophistication. These dimensions have been widely acknowledged in L2 writing research (e.g., Abdi Tabari et al., 2023; Zhang & Lu, 2024) and, to a lesser extent, in translation studies (e.g., Dinh, 2022; Liu & Dou, 2023) to provide a comprehensive view of lexical richness in texts.

Lexical density quantifies the information load carried by content words within texts. It is calculated as the proportion of content words (i.e., nouns, verbs, adjectives, and adverbs) relative to the total number of words (Ure, 1971). A higher lexical density indicates a greater informational load, signifying increased lexical complexity. The lexical density index was extracted using the Tool for the Automatic Analysis of Lexical Diversity (TAALED 1.4.1; Kyle et al., 2021), an NLP tool designed to automatically calculate this metric.

Lexical diversity measures the variation of words within texts. Common indices include type-token ratio (TTR), standardized type-token ratio (STTR), and moving average type-token ratio (MATTR). To ensure the accuracy of comparison, this study employed MATTR, noted for its stability across all text lengths (Zenker & Kyle, 2021). A higher MATTR signifies greater lexical complexity in texts. This index was also extracted by TAALED 1.4.1.

Lexical sophistication refers to the use of advanced vocabulary and encompasses several sub-dimensions, including word frequency, word range, n-gram frequency, lexical familiarity, and semantic networks (Kyle et al., 2018). The word frequency index measures the prevalence of common words in texts, while the word range index assesses the breadth of word usage across different contexts (Kyle & Crossley, 2015). N-gram frequency evaluates the occurrence of high-frequency bigrams and trigrams. Lexical familiarity pertains to psycholinguistic word information, drawing on familiarity scores from the MRC database. For these four sub-dimensions, higher scores indicate the use of more frequent or familiar lexical items, which reflect lower lexical complexity. Semantic networks, the final sub-dimension, account for polysemy and hypernymy. Polysemy scores measure the number of senses (i.e., meanings) associated with word forms, with higher scores indicating the use of more general and simpler words, thereby suggesting lower lexical complexity. In contrast, hypernymy scores reflect the number of superordinate terms, with higher scores implying the use

Table 3
Lexical complexity indices.

Dimensions	Indices	Descriptions
Density	Lexical density tokens	The proportion of content words
Diversity	MATTR50	The variety of words
Sophistication		
Word frequency	BNC written log frequency	Mean word frequency score
Word range	BNC written range	Mean word range (number of documents that a word occurs in) score
N-gram frequency	BNC written bigram proportion	The proportion of bigrams that fall within the top 50,000 most frequent bigrams
	BNC written trigram proportion	The proportion of trigrams that fall within the top 50,000 most frequent trigrams
Lexical familiarity	MRC Familiarity	Mean familiarity score
Semantic network	Polysemy content words	Average number of senses for content words
	Hypernymy nouns and verbs	Mean hypernymy score for nouns and verbs (averaging all senses and paths)

Note. Lexical density and diversity indices were extracted by TAALED 1.4.1 (Kyle et al., 2021). Lexical sophistication indices were extracted by TAALES 2.2 (Kyle & Crossley, 2015; Kyle et al., 2018).

of more specific and complex vocabulary, denoting higher lexical complexity. All lexical sophistication indices were extracted using the Tool for the Automatic Analysis of Lexical Sophistication (TAALES 2.2; Kyle et al., 2018; Kyle & Crossley, 2015), offering a comprehensive evaluation of lexical complexity across multiple dimensions. The operational definitions of all lexical complexity indices used in this study are provided in Table 3 (see Kyle et al. (2018, 2021), and Kyle and Crossley (2015) for further details).

3.3.2. Syntactic complexity indices

In examining syntactic complexity, this study focuses solely on global complexity indices without covering fine-grained complexity measures. While fine-grained measures provide a thorough investigation of complexity features in specific syntactic structures (Biber et al., 2020; Bulté & Housen, 2012), these indices tend to be highly interrelated and overlapping. This interconnectedness can potentially lead to overspecification of research outcomes (Chen et al., 2024), which compromises the clarity and generalizability of findings. As a result, global syntactic complexity was examined in this study to capture the general picture of language structures in the two translation versions.

Global syntactic complexity is understood as a multi-dimensional construct consisting of five components: production unit, phrasal complexity, subordination, coordination, and sentence complexity. It is specifically operationalized through 14 indices derived from the L2 Syntactic Complexity Analyzer (L2SCA) (Table 4; for detailed definitions and operationalizations, see Lu (2010)). These 14 indices have been widely applied in both L2 writing research (e.g., Casal & Lee, 2019; Lu & Ai, 2015) and translation studies (e.g., Chen et al., 2024; Liu & Afzaal, 2021; Wang et al., 2024) due to their comprehensive coverage of syntactic dimensions and strong reliability (Lu, 2010; Lu & Ai, 2015). Higher scores on these indices denote greater syntactic complexity. In this study, the indices were extracted using the Tool for the Automatic Analysis of Syntactic Sophistication and Complexity (TAASSC 1.3.8; Kyle, 2016), an NLP tool that automatically calculates scores for each index.

3.4. Data analysis

A mixed-methods approach, incorporating both quantitative and qualitative methodologies, was employed in the data analysis. For the quantitative analysis, Shapiro-Wilk tests were conducted to assess normality prior to the inferential data analysis. Non-parametric Wilcoxon signed-rank tests were then performed to examine statistically significant differences between learner translations with and without GPT editing. The alpha level was set at 0.05 for all statistical tests. Effect sizes were also obtained to evaluate the effect of GPT editing on each complexity index, providing insight into the practical significance of the results. For the qualitative analysis, prominent linguistic complexity features were extracted to illustrate and supplement the quantitative findings. This mixed-methods approach allows for a more comprehensive interpretation of the data, offering contextual depth to complement the statistical results.

4. Results

4.1. Lexical complexity

Regarding RQ1, which examined the differences in lexical complexity between L2 learner translations with GPT editing and those without, the GPT-edited versions consistently exhibited higher lexical complexity (Fig. 2, Table 5). All comparisons yielded effect sizes greater than 0.65, suggesting that GPT post-editing has a moderately strong impact on lexical complexity outcomes. The following sections present the findings across the three dimensions of lexical complexity: density, diversity, and sophistication.

GPT-edited learner translations ($Mdn = 0.528$) were lexically denser than those without editing ($Mdn = 0.510$; $V = 54, p < .001, r = 0.843$). This result indicates that the GPT-edited versions contain a higher proportion of content words. Such an increase suggests a denser semantic load and higher lexical complexity in the texts after being edited by GPT.

GPT-edited learner translations ($Mdn = 0.780$) demonstrated greater lexical diversity than the original learner translations ($Mdn =$

Table 4
Syntactic complexity indices.

Dimensions	Indices	Descriptions
Production unit	MLC	Mean length of clause
	MLT	Mean length of T-unit
	MLS	Mean length of sentence
Phrasal complexity	VP/T	Verb phrases per T-unit
	CN/C	Complex nominals per clause
	CN/T	Complex nominals per T-unit
Subordination	DC/C	Dependent clause ratio, i.e., dependent clauses per clause
	DC/T	Dependent clauses per T-unit
	C/T	T-unit complexity ratio, i.e., clauses per T-unit
	CT/T	Complex T-unit ratio, i.e., complex T-units per T-unit
Coordination	CP/C	Coordinate phrases per clause
	CP/T	Coordinate phrases per T-unit
	T/S	Sentence coordination ratio, i.e., T-unit per sentence
Sentence complexity	C/S	Sentence complexity ratio, i.e., clauses per sentence

Note. Syntactic complexity indices were extracted by TAASSC 1.3.8 (Kyle, 2016; Lu, 2010).

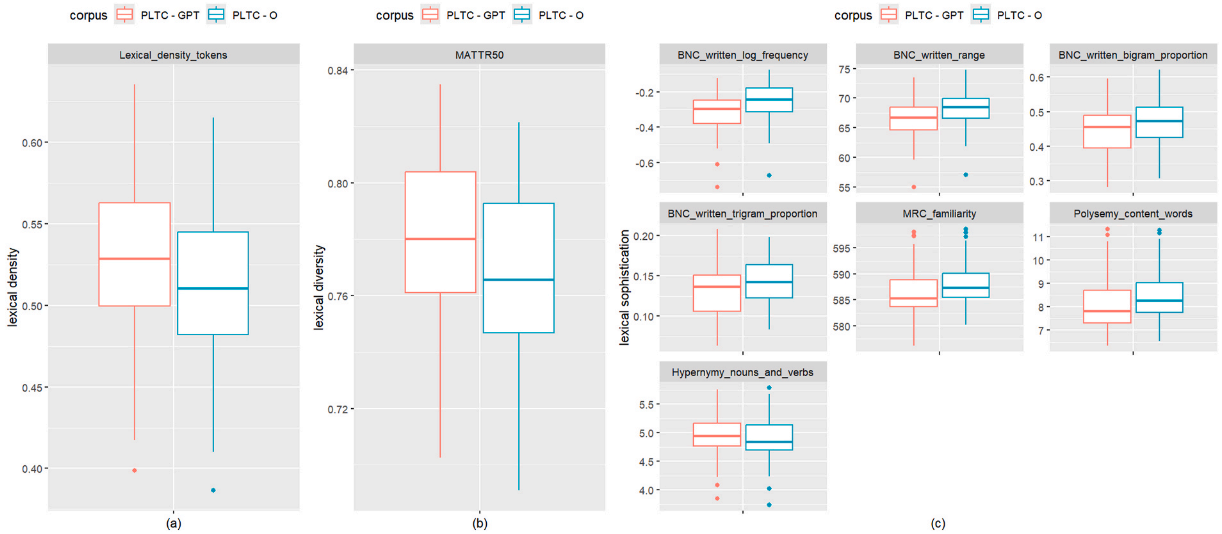

Fig. 2. Lexical complexity feature scores between the two corpora: (a) lexical density; (b) diversity; (c) sophistication.

Table 5
Wilcoxon signed-rank tests for comparing lexical complexity between the two corpora.

Indices	PLTC–O ($n = 85$)	PLTC–GPT ($n = 85$)	V	p	r
	Mdn	*Mdn*			
Density					
Lexical density tokens	0.510	0.528	54	<0.001	0.843
Diversity					
MATTR50	0.766	0.780	59	<0.001	0.841
Sophistication					
BNC written log frequency	−0.245	−0.298	3640	<0.001	0.861
BNC written range	68.511	66.725	3646	<0.001	0.864
BNC written bigram proportion	0.472	0.456	3431	<0.001	0.762
BNC written trigram proportion	0.142	0.136	3313	<0.001	0.706
MRC familiarity	587.30	585.32	3531	<0.001	0.810
Polysemy content words	8.247	7.806	3518	<0.001	0.803
Hypernymy nouns and verbs	4.843	4.939	457	<0.001	0.651

0.766; $V = 59$, $p < .001$, $r = 0.841$). This finding indicates that GPT editing results in increased word variation within texts, thereby enhancing lexical complexity.

Lexical sophistication was analyzed across five sub-dimensions. Regarding word frequency and word range, GPT-edited learner translations were characterized by using more low-frequency words (*Mdn_PLTC–GPT* = −0.298 vs. *Mdn_PLTC–O* = −0.245; $V = 3640$, $p < .001$, $r = 0.861$) and words appearing in a more restricted range of contexts (*Mdn_PLTC–GPT* = 66.725 vs. *Mdn_PLTC–O* = 68.511; $V = 3646$, $p < .001$, $r = 0.864$). These results indicate GPT's preference for uncommon words compared to unedited learner translations.

Considering n-gram frequency, the GPT-edited versions also featured more low-frequency bigrams (*Mdn_PLTC–GPT* = 0.456 vs. *Mdn_PLTC–O* = 0.472; $V = 3431$, $p < .001$, $r = 0.762$) and trigrams (*Mdn_PLTC–GPT* = 0.136 vs. *Mdn_PLTC–O* = 0.142; $V = 3313$, $p < .001$, $r = 0.706$). These findings suggest that GPT editing is not limited to common phrase patterns, instead incorporating more diverse and potentially sophisticated word combinations.

For psycholinguistic properties of words, the GPT-edited versions showed lower familiarity scores (*Mdn_PLTC–GPT* = 585.32 vs. *Mdn_PLTC–O* = 587.30; $V = 3531$, $p < .001$, $r = 0.810$), indicating a higher prevalence of less familiar words following GPT editing.

Finally, in terms of semantic network, GPT-edited learner translations exhibited lower polysemy scores (*Mdn_PLTC–GPT* = 7.806 vs. *Mdn_PLTC–O* = 8.247; $V = 3518$, $p < .001$, $r = 0.803$) but higher hypernymy scores (*Mdn_PLTC–GPT* = 4.939 vs. *Mdn_PLTC–O* = 4.843; $V = 457$, $p < .001$, $r = 0.651$). This suggests that GPT-edited translations contain words with more specific meanings, which are often considered more difficult. Collectively, GPT editing leads to more sophisticated word use, thereby enhancing lexical complexity.

Overall, GPT-edited learner translations demonstrated significantly higher lexical complexity across all examined aspects compared to those without GPT editing. This enhancement included increased lexical density, diversity, and sophistication. These findings suggest that GPT editing substantially augments the lexical richness of L2 learner translations, potentially mitigating the issue of lexical simplification often observed in translated texts.

4.2. Syntactic complexity

Regarding RQ2, which examined the differences in syntactic complexity between the two corpora, L2 learner translations with GPT post-editing showed significant differences from those without GPT editing in 10 of the 14 indices (Fig. 3, Table 6). For the indices with significant differences, the effect sizes ranged from 0.225 to 0.595, indicating that GPT editing has small to moderately strong effects on global syntactic complexity outcomes. The following sections present the results across the five syntactic complexity dimensions: production unit, phrasal complexity, subordination, coordination, and sentence complexity.

Among the three indices of the production unit dimensions, only the differences in mean length of clause (MLC) reached significant difference. Learner translations with GPT editing (*Mdn* = 12.49) exhibited greater clause length compared to those without GPT editing (*Mdn* = 12.16; $V = 576$, $p < .001$, $r = 0.595$). These findings indicate that GPT-edited versions contain more words at the clausal level but not at higher levels, such as T-unit and sentence levels.

The phrasal complexity dimension yielded mixed results. While non-edited versions contained more verb phrases per T-unit (VP/T) (*Mdn_PLTC–GPT* = 2.029 vs. *Mdn_PLTC–O* = 2.035; $V = 2342$, $p < .024$, $r = 0.245$), GPT-edited versions exhibited a greater number of complex nominals per clause (CN/C) (*Mdn_PLTC–GPT* = 1.634 vs. *Mdn_PLTC–O* = 1.628; $V = 770$, $p < .001$, $r = 0.503$). These results suggest that GPT editing shifts the emphasis from verb structures to nominal structures.

Learner translations without GPT editing consistently showed a greater use of subordination than GPT-edited versions across all four indices. These findings highlight distinct syntactic patterns between the two corpora, with GPT reducing the use of subordinate structures in L2 learner translations.

The GPT-edited versions exhibited a higher frequency of coordination compared to the non-edited versions, as evidenced by the number of coordinate phrases per clause (CP/C) (*Mdn_PLTC–GPT* = 0.310 vs. *Mdn_PLTC–O* = 0.292; $V = 781$, $p < .001$, $r = 0.497$) and per T-unit (CP/T) (*Mdn_PLTC–GPT* = 0.464 vs. *Mdn_PLTC–O* = 0.454; $V = 1354$, $p < .038$, $r = 0.225$). However, the sentence coordination ratio (T/S) remained unaffected after GPT editing. These results underscore GPT's preference for coordinate phrases.

Finally, learner translations with GPT editing (*Mdn* = 1.563) exhibited a higher sentence complexity ratio than those without GPT editing (*Mdn* = 1.560; $V = 2734$, $p < .001$, $r = 0.431$). This finding indicates that GPT editing results in more complex sentences containing a greater number of clauses.

Overall, the two groups demonstrated differing patterns of syntactic complexity. Learner translations with GPT editing demonstrated greater syntactic complexity in terms of more complex nominals, increased use of coordinate phrases, and longer clauses. Conversely, non-edited learner translations demonstrated greater syntactic complexity through more frequent verbal structures and a higher degree of subordination. These findings suggest that while GPT editing augments certain aspects of syntactic complexity in L2 learner translations, its effectiveness in mitigating the issue of syntactic simplification, often observed in translated texts, remains questionable.

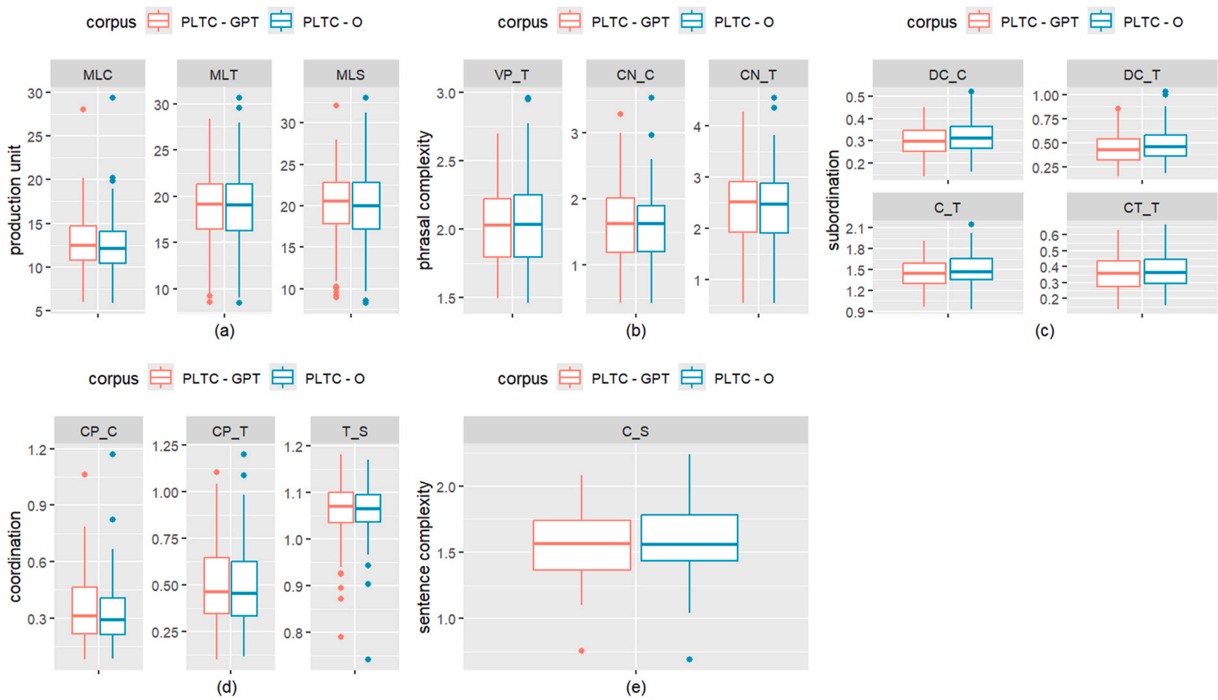

Fig. 3. Syntactic complexity feature scores between the two corpora: (a) production unit, (b) phrasal complexity, (c) subordination, (d) coordination, (e) sentence complexity.

Table 6
Wilcoxon signed-rank tests for comparing syntactic complexity between the two corpora.

Indices	PLTC–O ($n = 85$)	PLTC–GPT ($n = 85$)	V	p	r
	Mdn	*Mdn*			
Production unit					
Mean length of clause (MLC)	12.16	12.49	576	<0.001	0.595
Mean length of T-unit (MLT)	19.07	19.14	1785	>0.05	0.020
Mean length of sentence (MLS)	20.00	20.57	1753	>0.05	0.035
Phrasal complexity					
Verb phrases per T-unit (VP/T)	2.035	2.029	2342	0.024	0.245
Complex nominals per clause (CN/C)	1.628	1.634	770	<0.001	0.503
Complex nominals per T-unit (CN/T)	2.468	2.512	1658	>0.05	0.081
Subordination					
Dependent clause ratio (DC/C)	0.310	0.296	2982	<0.001	0.549
Dependent clauses per T-unit (DC/T)	0.459	0.430	2979	<0.001	0.578
T-unit complexity ratio (C/T)	1.464	1.444	2967	<0.001	0.542
Complex T-unit ratio (CT/T)	0.364	0.359	2555	0.001	0.346
Coordination					
Coordinate phrases per clause (CP/C)	0.292	0.310	781	<0.001	0.497
Coordinate phrases per T-unit (CP/T)	0.454	0.464	1354	0.038	0.225
Sentence coordination ratio (T/S)	1.066	1.070	1572	>0.05	0.104
Sentence complexity					
Sentence complexity ratio (C/S)	1.560	1.563	2734	<0.001	0.431

4.3. Qualitative findings

The differences between the two corpora are illustrated in Excerpts 1 to 4 below, with key changes emphasized. These excerpts reveal patterns and tendencies in language use within learner translations with and without GPT editing, offering qualitative insights that complement the quantitative findings.

Excerpt 1 [w2e-002] (Persuasive writing – press editorials)

(1a) PLTC–O: If we believe that **treasury payments** should only **go** to those in need, then there is **good reason to introduce** stricter asset checks to **weed out most** of **the** recipients, **so that the savings can be spent** elsewhere.

(1b) PLTC–GPT: If we believe that **government funds** should only **be allocated** to those in need, then there is **a strong argument for implementing** stricter asset checks to **exclude the majority** of recipients, **allowing the saved funds to be used** elsewhere.

Excerpt 2 [w2f-004] (Creative writing – novels and stories)

(2a) PLTC–O: I should avoid **doing** any business or investment during **the** year **in order** to prevent **incurring** losses. **Engaging in speculation** would not **bring me** any profit.

(2b) PLTC–GPT: I should avoid **engaging in** any business or investments during **this** year to prevent losses. **Speculative trading** would not **yield** any profits.

Excerpt 3 [w2b-016] (Popular writing – social sciences)

(3a) PLTC–O: Leung announced that Chief Executive Tung Chee-hwa's policy address in October would include **several** "small **initiatives for** stimulating the economy", **which caused lots of** speculation about **what** form of **largesse the** government **would take.**

(3b) PLTC–GPT: Leung announced that Chief Executive Tung Chee-hwa's policy address in October would include **some** "small **measures to** stimulate the economy," **leading to widespread** speculation about the form of government **assistance.**

As illustrated in Excerpts 1 and 2, learner translators often used de-lexical verbs, which GPT subsequently replaced with more specific lexical terms. For instance, GPT substituted *"go"* (Excerpt 1a) with *"be allocated"* (Excerpt 1b), and *"doing"* (Excerpt 2a) with *"engaging in"* (Excerpt 2b). In these examples, *"go"* and *"do"* are regarded as de-lexical verbs, which are common, general verbs whose meanings are primarily conveyed by the noun phrases that follow them. By replacing these de-lexical verbs with more precise alternatives, GPT enhances both lexical sophistication and the overall quality of the translated texts.

As exemplified in excerpts 1 and 3, the GPT-edited translations demonstrated a tendency to use participial phrases as replacements for dependent clauses found in the original versions. In Except 1, GPT changed the dependent clause *"so that the savings can be spent elsewhere"* (Excerpt 1a) to the participial phrase *"allowing the saved* funds *to be used elsewhere"* (Except 1b). A similar modification was observed in Excerpt 3, where GPT substituted the relative clause *"which caused lots of speculation …"* (Excerpt 3a) with the participial phrase *"leading to widespread speculation …"* (Excerpt 3b). These structural adjustments reduce the number of subordinations in the GPT-edited versions, potentially resulting in more concise sentences and smoother textual flow. However, implementing such changes requires careful consideration to ensure the clarity and readability of the texts.

GPT exhibited a preference for nominal forms over verbal forms, as shown in Excerpts 2 and 3. In Excerpt 2, the learner translator used the gerund phrase *"engaging in speculation"* (Excerpt 2a), a verb form ending in *-ing* to function as a noun. GPT modified this phrase into a more straightforward noun phrase, *"speculative trading"* (Excerpt 2b). Similarly, in Excerpt 3, the embedded question *"what form of largesse the government would take"* (Except 3a) was transformed into the noun phrase *"the form of government assistance"* (Except 3b). These examples illustrate GPT's tendency towards more complex nominal structures, contrasting with the non-edited learner translations, which feature denser verbal structures. The preference for nominal forms can potentially increase the conciseness and formality of the texts. However, it is important to note that while formal style may be appropriate for popular writing, i.e., scientific discoveries presented to the general public (Excerpt 3), it does not necessarily contribute to better writing quality in literary genres, particularly in creative writing (Excerpt 2).

Excerpt 4 [w2f-014] (Creative writing – novels and stories)
(4a) PLTC–O: She knew how fragile her life was, and **she** understood the rules of the forest.
(4b) PLTC–GPT: She knew how fragile her life was and understood the rules of the forest.

GPT tended to coordinate similar clausal structures to form coordinate phrases. As shown in Excerpt 4, two independent coordinate clauses in the original learner translation (Excerpt 4a) were integrated into coordinate phrases by eliminating the repeated subject *"she"* (Excerpt 4b). This change results in a higher number of coordinate phrases in the GPT-edited version. While such modification enhances conciseness and reduces redundancy, its appropriateness in literary translation is subject to debate. In fictional contexts, the original version (Excerpt 4a) may sometimes be more effective, as its slightly longer and more rhythmic expression can serve specific stylistic purposes or evoke emotional impact within the narrative. The choice between the two versions depends on the overall narrative style and the intended effect within the story.

These excerpts highlight the differences between learner translations with and without GPT editing, as well as the patterns of language use introduced by GPT. Notably, these patterns appear across different registers, despite GPT being prompted with specific register information for each text. Such systematic revisions demonstrate the potential of GPT to enhance both lexical and syntactic complexity in learner translations, while also raising concerns about the preservation of genre-specific stylistic features and the authorial voice of the source text during the editing process.

5. Discussion

This study compared the lexical and syntactic complexity features in L2 translations with and without GPT editing. The findings address the first research question that translations post-edited by GPT exhibited greater lexical complexity compared to those without GPT editing. Subsequent qualitative analysis revealed that GPT consistently replaced simple, de-lexical verbs with more specific lexical choices. However, results were mixed in terms of syntactic complexity. While GPT post-editing produced lengthier clauses, more complex nominals, and an increase in coordinate phrases, the original learner translations featured a greater use of subordination and denser verbal structures. Qualitative analysis further revealed that GPT consistently substituted subordinating structures with participial phrases, favored nominal forms over verbal ones, and coordinated similar clauses into compound phrases. The following section discusses these findings in relation to learner translators' linguistic proficiency and GPT's editing mechanisms, as well as ongoing debates regarding AI translationese and its implications for GenAI development and the potential roles of GenAI in translation practices.

5.1. Learner translators' linguistic proficiency and simplification

The linguistic competence of learner translators is unlikely to match that of GPT, contributing to the observed differences in linguistic complexity before and after GPT editing. In this study, translations produced by learner translators are conceptualized as a form of constrained communication (Lanstyák & Heltai, 2012), similar to the challenges encountered in L2 writing. Both processes are influenced by linguistic limitations, cognitive load, and contextual factors, all of which collectively affect the quality and complexity of the output. From a cognitive perspective, linguistic items are conceptualized as form-meaning pairs structured within networks of interconnected meanings (Langacker, 1987). During the translation process, the robustness of these networks can impact lexical retrieval and selection (Halverson, 2003, 2017), thereby shaping the final translation product. This mechanism highlights the pivotal role of linguistic competence in determining the lexical complexity of translations produced by learners. While L2 translation benefits from accurate comprehension of the source text (Campbell, 1998), it is often hindered by less proficient expression in the non-native target language, particularly for learner translators (Samuelsson-Brown, 2010). The combined challenges of non-expert translation performance and non-native language production may result in a simplification of their translation outputs. In contrast, GPT has been trained on vast amounts of language data, comprising 175 billion parameters in GPT-3 (Brown et al., 2020; Johri, 2023), making vocabulary size an area in which learner translators are unlikely to compete. This vast training enables GPT to counteract lexical simplification by replacing basic, de-lexicalized verbs with more precise and contextually appropriate alternatives. The qualitative findings demonstrate how GPT's extensive linguistic knowledge allows it to enhance lexical variety and complexity, producing outputs that are typically beyond the reach of learner translators.

5.2. GPT's mechanisms and its influence on linguistic complexity

The mechanisms underlying GPT's operation may also account for its robust lexical performance. As a large language model, GPT is initially trained to predict the next word in a sequence based on contextual information. Through a series of subsequent training and fine-tuning, the model develops the ability to respond to various instructions and generate coherent and relevant outputs (Johri, 2023). This mechanism indicates that GPT operates primarily at the lexical level while maintaining a strong awareness of textual context, enabling it to effectively mitigate lexical simplification in learner translations.

While GPT demonstrates a strong capacity to address lexical simplification in learner translations, its ability to handle syntactic complexity remains less certain. As evidenced by our findings, GPT-edited texts exhibit a reduction in subordination and verbal structures, coupled with an increase in the use of lengthier clauses, more intricate nominal constructions, and an increased reliance on coordinate phrases. Similar patterns have been observed in GPT-revised student argumentative writing (Wang, 2023). These findings suggest that syntactic simplicity in certain aspects may be offset by increased complexity in others following GPT editing. This phenomenon is likely a byproduct of GPT's next-word prediction training mechanism. While this approach often yields lexically refined and sophisticated texts, it may also favor certain word sequences, resulting in predictable and formulaic sentence patterns. Overall, these observations underscore both the strengths and limitations of GPT's editing capabilities in addressing linguistic complexity in learner translations.

5.3. *AI translationese and its implications for GenAI development*

The linguistic patterns generated by GPT observed in this study have contributed to discussions surrounding the concept of an AI footprint. Gellerstam (1986) introduced the term "translationese" to describe the linguistic markers left by the translation process. With the rise of MT, the notion of "machine translationese" has emerged, referring to the linguistic effects introduced by MT algorithms (Vanmassenhove et al., 2021). It is plausible that GenAI algorithms similarly influence their language output, exhibiting favorable language patterns.

This study examines the use of GPT-3.5, a foundational GenAI model in the field, for post-editing learner translations. Recent advancements in GenAI have introduced newer models that incorporate larger training datasets, enhanced architectures, and refined fine-tuning techniques. These advancements have improved their language processing capabilities (Rahaman et al., 2023). However, even with advanced models, certain challenges persist. For example, research indicates that GPT-4 can occasionally produce hallucinated edits, underscoring the caution when using them as translation post-editors (Raunak et al., 2023). As future models continue to emerge, their linguistic patterns are likely to shift, potentially transforming (not necessarily eliminating) the identifiable "AI footprint" observed in earlier models. Therefore, it remains uncertain whether newer models can effectively address translation simplification issues. Continuous investigations and regular reassessment of these technologies are essential to track their capabilities and limitations, alongside ongoing efforts to mitigate the AI footprint.

5.4. *Practical implications*

The advent of GenAI may also reshape the standards of linguistic and instrumental competence in translation. Previous research has affirmed the potential of GenAI to provide valuable support in general language learning (Guo et al., 2022; Guo & Wang, 2023). The findings of this study also highlight a collaborative role for GenAI as a translation-assistance tool, offering several benefits for translation learners. These benefits include providing alternative word choices, aiding in text elaboration, and facilitating the connection of similar ideas through parallel structures. To some extent, GenAI post-edited texts may be viewed as alternative solutions or feedback for learners' translations.

Despite these benefits, caution is needed when using GenAI tools for translation. This study reveals that GenAI post-editing does not adequately address the issue of syntactic simplification. Moreover, the analyzed excerpts suggest the potential emergence of an AI translationese, as GenAI tends to favor certain sentence patterns and may be less likely to adjust language style according to specific text genres. It remains unclear whether more targeted prompts could enhance the production of genre-appropriate translations.

As GenAI systems evolve, the need for critical AI literacy becomes increasingly important. Boden et al. (2017) contended that AI tools cannot be regarded as responsible agents, and, as with other translation technologies, the decision to adopt AI-generated translations ultimately depends on the user's critical judgment. This approach aligns with the concept of critical AI literacy proposed by Giustini and Dastyar (2024), which emphasizes not only the digital skills required to operate AI technologies but also a deeper understanding of their capabilities and limitations. Key challenges include interpreting AI-generated outputs, navigating the complex autonomy of AI systems, and addressing privacy concerns, all of which are critical issues associated with the adoption of GenAI (Abdelaal & Al Sawy, 2024).

In summary, while GenAI enhances certain aspects of learner translations, its limitations underscore the continued importance of human expertise and the development of translation competence. Further research is needed to explore how these tools can be effectively integrated into translation pedagogy and practice.

6. Conclusion

This study examined how GPT alters learner translational language by comparing L2 learner translations with and without GPT editing, using lexical and syntactic complexity as key measures. The analysis yielded two key findings that address the research questions. First, translations post-edited by GPT consistently demonstrated greater lexical complexity compared to those without GPT editing. Second, while GPT post-editing resulted in longer clauses, more complex nominals, and a higher frequency of coordinate phrases, non-GPT-edited translations made greater use of subordination and featured denser verbal structures. These differences can be attributed to the linguistic competence of learner translators and the mechanisms underlying GPT's operations. Finally, the implications of these findings, particularly concerning the potential AI footprint and the role of GenAI in translation, have been explored.

There are a few factors that may limit the generalizability of the results. First, the absence of target native texts as a reference corpus makes it challenging to evaluate how closely the translations align with the complexity of native language expressions. Second, for the sake of research control, GPT editing in this study followed a one-off input-process-output model. While this approach ensured consistency, it does not fully represent the more dynamic, iterative use of GenAI in practical settings, where users frequently ask follow-up questions or make additional edits. Finally, to ensure the validity of the results, we deliberately excluded learner translations produced using GenAI tools. While this exclusion strengthens the internal validity of the study, it may limit the extent to which the findings can be generalized to broader GenAI-assisted translation contexts, particularly in post-editing tasks. Taken together, these factors suggest that the findings should be interpreted with caution when applied to real-world GenAI-assisted translation practices.

Despite these considerations, this study makes valuable contributions to both translation pedagogy and our initial understanding of the features of GenAI post-edited translations. It lays a foundation for further research into GenAI-assisted translation practices. Looking ahead, future research should explore several key areas: incorporating a reference corpus to better assess the naturalness of GenAI-assisted translations, investigating interactive GenAI-human collaboration in translation tasks, and examining how different prompt designs influence translation outcomes. These avenues of inquiry will help refine our understanding of GenAI's role in translation and its potential to reshape translation pedagogy.

CRediT authorship contribution statement

Ho Ling Kwok: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yining Shi:** Methodology, Investigation, Formal analysis. **Han Xu:** Writing – review & editing, Validation, Methodology. **Dechao Li:** Writing – review & editing, Validation, Methodology. **Kanglong Liu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization.

Author notes

Relevant data concerning the study are publicly available on Open Science Framework (https://osf.io/7xkdu/).

Declaration of competing interest

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.

Acknowledgment

The research presented in this paper was supported by multiple grants from The Hong Kong Polytechnic University (Project Nos: TDG22-25/VTL-6, P0051009, and CBS/24-25/LD).

References

Abdelaal, N., & Al Sawy, I. (2024). Perceptions, challenges, and prospects: University professors' use of artificial intelligence in education. *Australian Journal of Adult Learning, 7*(1), 1309. https://doi.org/10.29140/ajal.v7n1.1309

Abdi Tabari, M., Lu, X., & Wang, Y. (2023). The effects of task complexity on lexical complexity in L2 writing: An exploratory study. *System, 114*, Article 103021. https://doi.org/10.1016/j.system.2023.103021

Al Sawi, I., & Alaa, A. (2024). Navigating the impact: A study of editors' and proofreaders' perceptions of AI tools in editing and proofreading. *Discover Artificial Intelligence, 4*(23). https://doi.org/10.1007/s44163-024-00116-5

Al Sawi, I., & Allam, R. (2024). Exploring challenges in audiovisual translation: A comparative analysis of human- and AI-generated Arabic subtitles in birdman. *PLoS One, 19*(10), Article e0311020. https://doi.org/10.1371/journal.pone.0311020

Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, F. Gill, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of john sinclair* (pp. 233–250). John Benjamins Publishing Company.

Baker, M. (1996). Corpus-based translation studies: The challenges that lie ahead. In H. Somers (Ed.), *Terminology, LSP and translation: Studies in language engineering in honour of Juan C. Sager* (pp. 175–186). John Benjamins Publishing Company.

Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes, 46*, Article 100869. https://doi.org/10.1016/j.jeap.2020.100869

Bistafa, S. R. (2024). Translating scientific Latin texts with artificial intelligence: The works of Euler and contemporaries. *arXiv*. https://doi.org/10.48550/arXiv.2307.07520

Boden, M., Bryson, J., Caldwell, D., Dautenhahn, K., Edwards, L., Kember, S., Newman, P., Parry, V., Pegman, G., Rodden, T., Sorrell, T., Wallis, M., Whitby, B., & Winfield, A. (2017). Principles of robotics: Regulating robots in the real world. *Connection Science, 29*(2), 124–129. https://doi.org/10.1080/09540091.2016.1271400

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *arXiv*. https://doi.org/10.48550/arXiv.2005.14165

Bui, G. (2021). Influence of learners' prior knowledge, L2 proficiency and pre-task planning on L2 lexical complexity. *International Review of Applied Linguistics in Language Teaching, 59*(4), 543–567. https://doi.org/10.1515/iral-2018-0244

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 21–46). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.32.02bul.

Campbell, S. (1998). *Translation into the second language*. Longman.

Casal, J. E., & Lee, J. J. (2019). Syntactic complexity and writing quality in assessed first-year L2 writing. *Journal of Second Language Writing, 44*, 51–62. https://doi.org/10.1016/j.jslw.2019.03.005

Chen, J., Li, D., & Liu, K. (2024). Unraveling cognitive constraints in constrained languages: A comparative study of syntactic complexity in translated, EFL, and native varieties. *Language Sciences, 102*, Article 101612. https://doi.org/10.1016/j.langsci.2024.101612

Chesterman, A. (2004). Hypotheses about translation universals. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, changes and challenges in translation studies: Selected contributions from the EST Congress, Copenhagen 2001* (pp. 1–13). John Benjamins.

Dinh, H. (2022). Synergic concepts, lexical idiosyncrasies, and lexical complexities in bilingual students' translated texts as efforts to resolve conceptual inequivalences. *Language, 7*(2), 94. https://doi.org/10.3390/languages7020094

Ferraresi, A., Bernardini, S., Petrović, M., & Lefer, M.-A. (2018). Simplified or not simplified? The different guises of mediated English at the European parliament. *Meta, 63*(3), 717–738. https://doi.org/10.7202/1060170ar

Frawley, W. (1984). *Translation: Literary, linguistic, and philosophical perspectives*. University of Delaware Press.

Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: A comparison among ChatGPT, google translate, and DeepL translator. *Humanities and Social Sciences Communications, 11*(1), 835. https://doi.org/10.1057/s41599-024-03363-0

Gellerstam, M. (1986). Translationese in Swedish novels translated from English. In L. Wollin, & H. Lindquist (Eds.), *Translation studies in scandinavia* (pp. 88–95). CWK Gleerup.

Giustini, D., & Dastyar, V. (2024). Critical AI literacy for interpreting in the age of AI. *Interpreting and society*. https://doi.org/10.1177/27523810241247259

Grosjean, F. (2013). Bilingualism: A short introduction. In F. Grosjean, & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 5–25). Wiley-Blackwell.

Guo, K., & Wang, D. (2023). To resist it or to embrace it? Examining ChatGPT's potential to support teacher feedback in EFL writing. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-12146-0

Guo, K., Wang, J., & Chu, S. K. W. (2022). Using chatbots to scaffold EFL students' argumentative writing. *Assessing Writing, 54*, Article 100666. https://doi.org/10.1016/j.asw.2022.100666

Haleem, A., Javaid, M., & Singh, R. P. (2022). An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 2*(4), Article 100089. https://doi.org/10.1016/j.tbench.2023.100089

Halverson, S. L. (2003). The cognitive basis of translation universals. *Target, 15*(2), 197–241. https://doi.org/10.1075/target.15.2.02hal

Halverson, S. L. (2017). Gravitational pull in translation. Testing a revised model. In G. De Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical translation studies: New methodological and theoretical traditions* (pp. 9–46). De Gruyter Mouton. https://doi.org/10.1515/9783110459586-002.

Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., Kim, Y. J., Afify, M., & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv*. https://doi.org/10.48550/arXiv.2302.09210

House, J. (2016). Universals of translation? In J. House (Ed.), *Translation as communication across languages and cultures* (pp. 58–64). Routledge.

Jiao, W., Wang, W., Huang, J-t., Wang, X., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv*. https://doi.org/10.48550/arXiv.2301.08745

Johri, S. (2023). The making of ChatGPT: From data to dialogue. Retrieved from https://sitn.hms.harvard.edu/flash/2023/the-making-of-chatgpt-from-data-to-dialogue/. (Accessed 6 August 2024).

Khoshafah, F. (2023). ChatGPT for Arabic-English translation: Evaluating the accuracy. *Research Square*. https://doi.org/10.21203/rs.3.rs-2814154/v2

Kruger, H., & van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide, 37*(1), 26–57. https://doi.org/10.1075/eww.37.1.02kru

Kunilovskaya, M., Morgoun, N., & Pariy, A. (2018). Learner vs. professional translations into Russian: Lexical profiles. *Translation & Interpreting, 10*(1), 33–52. https://doi.org/10.12807/ti.110201.2018.a03

Kwok, H. L., Laviosa, S., & Liu, K. (2023). Lexical simplification in learner translation: A corpus-based approach. *Research in Corpus Linguistics, 11*(2), 103–124. https://doi.org/10.32714/ricl.11.02.06

Kyle, K. (2016). *Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication*. Georgia State University. Doctorial dissertation.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *Tesol Quarterly, 49*(4), 757–786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*(3), 1030–1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity indices using direct judgements. *Language Assessment Quarterly, 18*(2), 154–170. https://doi.org/10.1080/15434303.2020.1844205

Langacker, R. W. (1987). *Foundations of cognitive grammar* (vol. 1). Stanford, Calif: Stanford University Press. *Theoretical prerequisites*.

Lanstyák, I., & Heltai, P. (2012). Universals in language contact and translation. *Across Languages and Cultures, 13*(1), 99–121. https://doi.org/10.1556/Acr.13.2012.1.6

Larsson, T., Egbert, J., & Biber, D. (2022). On the status of statistical reporting versus linguistic description in corpus linguistics: A ten-year perspective. *Corpora, 17*(1), 137–157. https://doi.org/10.3366/cor.2022.0238

Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. *Meta, 43*(4), 557–570. https://doi.org/10.7202/003425ar

Liu, K., & Afzaal, M. (2021). Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification. *PLoS One, 16*(6), Article e0253454. https://doi.org/10.1371/journal.pone.0253454

Liu, Z., & Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective. *Frontiers in Psychology, 14*, Article 1276705. https://doi.org/10.3389/fpsyg.2023.1276705

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International Journal of Corpus Linguistics, 15*(4), 474–496. https://doi.org/10.1075/ijcl.15.4.02lu

Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing, 29*, 16–27. https://doi.org/10.1016/j.jslw.2015.06.003

McWhorter, J. H. (2011). *Linguistic simplicity and complexity: Why do languages undress?* De Gruyter Mouton. https://doi.org/10.1515/9781934078402

Moslem, Y., Haque, R., Kelleher, J. D., & Way, A. (2023). Adaptive machine translation with large language models. *arXiv*. https://doi.org/10.48550/arXiv.2301.13294

Pavlović, T. (2013). Exploring directionality in translation studies. Explorations in English *Language and linguistics. 1*(2), 149–165.

Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). Towards making the most of ChatGPT for machine translation. *arXiv*. https://doi.org/10.48550/arXiv.2303.13780

Penha-Marion, L., Gilquin, G., & Lefer, M.-A. (2024). The effect of directionality on lexico-syntactic simplification in French><English student translation. In B. van Rooy, & H. Kotze (Eds.), *Constraints on language variation and change in complex multilingual contact settings* (pp. 153–190). John Benjamins Publishing Company.

Pym, A. (2008). On Toury's laws of how translators translate. In A. Pym, M. Shlesinger, & D. Simeoni (Eds.), *Beyond descriptive translation studies: Investigations in homage to Gideon Toury* (pp. 311–328). John Benjamins Publishing Company.

Rahaman, M. S., Ahsan, M. M. T., Anjum, N., Terano, H. J. R., & Rahman, M. M. (2023). From ChatGPT-3 to GPT-4: A significant advancement in AI-driven NLP tools. *Journal of Engineering and Emerging Technologies, 2*(1), 1–11. https://doi.org/10.52631/jeet.v2i1.188

Raunak, V., Sharaf, A., Awadallah, H. H., & Menezes, A. (2023). *Leveraging GPT-4 for automatic translation post-editing* (2305.14878v2). *arXiv*. https://doi.org/10.48550/arXiv.2305.14878

Sahari, Y., Al-Kadi, A. M. T., & Ali, J. K. M. (2023). A cross sectional study of ChatGPT in translation: Magnitude of use, attitudes, and uncertainties. *Journal of Psycholinguistic Research*. https://doi.org/10.1007/s10936-023-10031-y

Samuelsson-Brown, G. (2010). A practical guide for translators. *Multilingual Matters*. https://doi.org/10.21832/9781847692610

Toury, G. (2012). *Descriptive translation studies – and beyond* (Revised edition). John Benjamins Publishing Company.

Tsai, Y. (2021). Diachronic observation of lexical and syntactic patterns in English translations of Taiwan patent texts. *Perspectives, 29*(6), 865–885. https://doi.org/10.1080/0907676X.2020.1849325

Tymoczko, M. (1998). Computerized corpora and the future of translation studies. *Meta, 43*(4), 652–660. https://doi.org/10.7202/004515ar

Ure, J. (1971). Lexical density and register differentiation. In G. E. Perren, & J. L. M. Trim (Eds.), *Applications of linguistics* (pp. 443–452). Cambridge University Press.

Vanmassenhove, E., Shterionov, D., & Gwilliam, M. (2021). Machine translationese: Effects of algorithmic bias on linguistic complexity in machine translation. In P. Merlo, J. Tiedemann, & R. Tsarfaty (Eds.), *Proceedings of the 16th conference of the European chapter of the association for computational linguistics: Main volume* (pp. 2203–2213). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.188.

Wang, C. (2023). A syntactic complexity analysis of revised composition through artificial intelligence-based question-answering systems. In *Proceedings of the 2023 2nd international conference on artificial intelligence and computer information technology (AICIT)* (pp. 1–3). IEEE. https://doi.org/10.1109/AICIT59054.2023.10277827.

Wang, Z., Liu, M., & Liu, K. (2024). Utilizing machine learning techniques for classifying translated and non-translated corporate annual reports. *Applied Artificial Intelligence, 38*(1), Article 2340393. https://doi.org/10.1080/08839514.2024.2340393

Wen, T. (2009). *Simplification as a recurrent translation feature: A corpus-based study of modern Chinese translated mystery fiction in taiwan*. University of Manchester. PhD thesis.

Wiethof, C., Tavanapour, N., & Bittner, E. (2021). Implementing an intelligent collaborative agent as teammate in collaborative writing: Toward a synergy of humans and AI. In *Proceedings of the 54th Hawaii international conference on system sciences* (pp. 400–409).

Wu, J., Gan, W., Chen, Z., Wan, S., & Lin, H. (2023). AI-generated content (aigc): A survey. *arXiv*. https://doi.org/10.48550/arXiv.2304.06632

Xiao, R., & Yue, M. (2009). Using corpora in translation studies: The state of the art. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 237–262). Continuum International Publishing Group.

Yamada, M. (2023). Optimizing machine translation through prompt engineering: An investigation into ChatGPT's customizability. In M. Yamada, & F. do Carmo (Eds.), *Proceedings of machine translation summit XIX, vol. 2: Users track* (pp. 195–204). Asia-Pacific Association for Machine Translation.

Zenker, F., & Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing, 47*, Article 100505. https://doi.org/10.1016/j.asw.2020.100505

Zhang, X., & Lu, X. (2024). Testing the relationship of linguistic complexity to second language learners' comparative judgment on text difficulty. *Language Learning, 74*(3), 672–706. https://doi.org/10.1111/lang.12633