Linguistic Variation in L2 Writing and Learner Translation: A Corpus-Based Multidimensional Analysis

Zhongliang Wang[1] *Ho Ling Kwok*[2] *Kanglong Liu*[3]

Address: [1]School of Foreign Languages and Commerce, Guangzhou Railway Polytechnic, Guangzhou, China;
[2]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China；
[3]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
E-mail: [1]wangzhongliang@gtxy.edu.cn; [2]ho-ling.kwok@connect.polyu.hk;
[3]kl.liu@polyu.edu.hk
Correspondence: Kanglong Liu
Citation: Wang, Zhongliang, Ho Ling Kwok and Kanglong Liu. "Linguistic Variation in L2 Writing and Learner Translation: A Corpus-Based Multidimensional Analysis." *Translation Quarterly* 110: 90-113.

Abstract

This study examines the linguistic features of L2 (second language) writing and L2 learner translation among Chinese learners, employing Biber's (1988) multi-dimensional analysis framework. The analysis is based on two corpora: 53 texts from the International Corpus of English - Hong Kong (ICE-HK) representing L2 writing, and 53 translated texts from the Parallel Learner Translation Corpus (PLTC), which were self-compiled for this study. Both corpora encompass six written genres—academic writing, popular writing, reportage, instructional writing, persuasive writing, and creative writing. Dimension scores were calculated for each text and compared between the two corpora using Wilcoxon signed-rank tests. Although both corpora primarily align with the "general narrative exposition" genre, significant differences were observed in Dimensions 1 (Involved vs. Informational Discourse), 4 (Overt Expression of Persuasion), and 6 (Online Informational Elaboration). Specifically, ICE-HK exhibited significantly lower scores on these dimensions compared to PLTC, indicating that L2 writing is more informational, less persuasive, and shows reduced online informational elaboration than L2 translations. These findings highlight the distinct constraints and pro-

duction processes inherent to L2 writing versus L2 translation, offering valuable insights into the linguistic variation between these two modes of L2 production and informing practices in second language writing instruction and translation training.

1. Introduction

Direct English composition and translation involve different linguistic skills and mental processes (Göpferich and Jääskeläinen 2009). However, research in applied linguistics has yet to fully explore the distinctions between these two modes of language production, particularly in the Hong Kong context. In direct composition, L2 (second language) speakers generate ideas and produce language simultaneously in English, while translation involves expressing pre-formed ideas originally conceived in Chinese, requiring a cross-linguistic transfer into English. These processes appear to engage distinct cognitive resources—immediate language processing for direct writing and cross-linguistic transfer for translation. Both skills are essential in academic and professional contexts where Chinese speakers may need to both compose directly in English and translate Chinese materials, yet current pedagogical practices often do not fully address the unique demands of each. Given that Chinese and English are typologically distant languages, a deeper understanding of their linguistic differences could greatly enhance teaching methods for both English writing and translation.

This study addresses this key research gap by applying the multi-dimensional analysis proposed by Biber (1988) to compare L2 learner translation with L2 writing, both in English. This approach facilitates a comprehensive examination of linguistic features across multiple dimensions, shedding light on how these two tasks differ. Previous research has explored differences between native and non-native writing (Vo 2019) and between professional and learner translations (Kunilovskaya et al. 2018). However, specific differences between L2 writing and L2 learner translation remain underexplored. A valuable theoretical framework for examining these distinctions is the concept of constrained communication. Both L2 writing and translation involve constrained language production, though each presents unique constraints (Chen et al. 2024; Lanstyák and Heltai 2012). In L2 writing, processing is inherently constraint-based (Wu et al. 2023) and influenced by factors such as limited linguistic resources in the target language, while in translation, constraints arise from the need to accurately convey the source text's meaning and style (e.g., Snape et al. 2023; Wetzel et al. 2024). This framework suggests that observed differences between writing and translation may reflect distinct ways L2 speakers manage these constraints. Understanding these linguistic differences is essential for advancing language acquisition and translation studies (Kwok et al. 2023). By comparing these two modes of L2 production through Biber's multi-dimensional analysis, this study aims to provide nuanced insights into how L2 speakers navigate the unique demands of writing and translation tasks. The findings have the potential to significantly impact second

language acquisition theory, translation studies, and language pedagogy, contributing to more effective L2 instruction and assessment practices.

2. Literature Review

2.1 Constrained Communication

The concept of "constrained communication" was introduced to describe communicative situations where "one or several of the potential limiting factors play a greater than average role" (Lanstyák and Heltai 2012, 100). It suggests that various types and degrees of constraints affect all communication instances. Certain contexts, such as language contact scenarios—including translation and bilingual interactions—are particularly impacted by these constraints (ibid.). Expanding on this concept, some scholars have introduced the term "constrained language" to describe language output produced under heightened restrictions (Kruger and van Rooy 2016, 27). Both L2 production and translated texts are considered to share similar constraints, mainly due to the fact that they both involve the transfer of meaning from one language to another, which can result in similar linguistic and cognitive challenges, such as language interference, cultural differences, and limitations in expressing nuances and idiomatic expressions. This perspective provides a valuable framework for understanding the unique challenges and characteristics of language produced in bilingual contexts, whether through translation or direct second-language use.

L2 writing has long been recognized as a constrained form of writing, with L2 writers facing numerous limitations, such as restricted vocabulary, incomplete mastery of grammar, and interference from their first language (Hinkel 2011). This perspective suggests that L2 writers operate under unique constraints that shape their linguistic choices and overall writing performance, often resulting in distinctive features in L2 writing. Kormos (2011) argued that L2 writing is inherently more cognitively demanding than L1 (first language) writing due to the additional constraints imposed by limited linguistic resources. These constraints can affect various aspects of the writing process, including lexical retrieval, syntactic formulation, and discourse organization, leading to an increased cognitive load that may impact the fluency and quality of L2 written output. Wang and Wen (2002), through their think-aloud study on L1 use during L2 writing, found that linguistic constraints prompted strategic L1 use, especially for idea generation and organizational planning, though linguistic encoding in L2 remained challenging.

Translation is a key area where the concept of constrained communication plays a crucial role. In the translation process, there is continuous bilingual activation (Kruger and van Rooy 2016), as translators must frequently alternate between two linguistic systems, placing significant demands on working memory. Additionally, the lack of a shared communica-

tive context—due to linguistic differences—can lead to breakdowns or misunderstandings. To address these challenges, translators often employ strategies to reduce cognitive load and mitigate risk. These strategies may include literal translation, explication, and simplification (Carl and Dragsted 2012; Pym 2015). Furthermore, the translation process is inherently constrained by the source text, which can influence the structure and content of the target language. This interference may manifest in various linguistic features of the translated output (Toury 2012). Collectively, these insights highlight the complex interplay between language activation, translation processes, and the constraints imposed by source-text fidelity in bilingual communication.

Despite the growing body of research in both L2 writing and translation studies, significant gaps remain in our understanding of the linguistic differences between L2 writing and L2 learner translation. While both forms of language production share the common characteristic of being constrained communication (Lanstyák and Heltai 2012), they present distinct cognitive and linguistic challenges that warrant further investigation. L2 writing has been extensively examined from various perspectives, including cognitive processes (Manchón 2012), linguistic features (Crossley and McNamara 2012), and developmental trajectories (Polio 2017). Similarly, translation studies have explored the linguistic characteristics of translated texts, often focusing on professional translations or comparing them to native language production (Baker 1993; Liu and Afzaal 2021; Liu et al. 2022; Wang et al. 2023; Wang, Liu et al. 2024; Wang, Cheung et al. 2024). However, the specific linguistic features that differentiate L2 writing from L2 learner translation remain underexplored. This gap is particularly noteworthy given that both tasks are commonly undertaken by L2 speakers in academic and professional settings. Furthermore, most previous studies on the linguistic features of translations tend to focus on isolated features, making it crucial to examine the linguistic characteristics of both text types systematically from multiple dimensions.

2.2 Multi-Dimensional Analysis (MDA)

MDA (Biber 1988) is a methodology for analyzing genre variation that utilizes 67 linguistic features and employs factor analysis to identify dimensions of language variation. The MDA process consists of three main stages: feature selection and frequency calculation, factor analysis for feature clustering, and dimension score computation. In the first stage, researchers select relevant linguistic features and calculate their frequency within a corpus. In the second stage, factor analysis is used to group these features into factors, which are then interpreted as textual dimensions based on the communicative functions of their constituent elements. Finally, dimension scores are calculated by considering the positive and negative factor loadings of the linguistic features within each factor. As McEnery et al. (2006, 288) explain, the dimension score is computed by adding together the factor scores of features with positive loadings and subtracting the factor scores of features with negative loadings.

Two primary approaches to MDA have been employed in linguistic and translation studies. The first is the standard MDA model (Biber 1988), which incorporates six dimensions and 67 linguistic features (e.g., Kruger and van Rooy 2016; Wang and Liu 2024). The second approach is customized, where researchers select their own linguistic features, conduct factor analysis, and interpret the resulting dimensions (e.g., Hu et al. 2016; Xiao 2009).

Over the past three decades, MDA has gained widespread adoption in genre studies, with applications across a variety of text types, including student writing (Gardner et al. 2019), annual reports (Ren and Lu 2021), fiction (Su and Liu 2022), diplomatic discourse (Yao et al. 2024), speeches (Xu and Li 2024) and academic articles (Liu and Xiao 2022). Liu and Xiao (2022) used MDA to examine the linguistic strategies employed in the conclusion sections of research articles across different disciplines. They found significant variations in writing styles between natural and social science papers, as well as evidence that the communicative intent of the conclusion sections influences linguistic features across multiple dimensions. Similarly, Gardner et al. (2019) applied MDA to investigate the British Academic Written English corpus, uncovering and interpreting four distinct dimensions of linguistic features. Their findings provided a framework for mapping academic disciplines, disciplinary groups, study levels, and genre families.

In recent years, MDA has found increasing applications in translation studies. Researchers have used this method to investigate translated language characteristics and L2 English varieties (Hu et al. 2016; Kruger and van Rooy 2016), as well as areas like financial translation (Wang and Liu 2024) and literary translation (Su and Liu 2022). Kruger and van Rooy (2016) applied MDA to analyze a translation corpus and a parallel set of texts from the International Corpus of English for East Africa, using the International Corpus of English for Great Britain as a control corpus. Their goal was to determine whether translated and non-native indigenized English variants share common features due to similar barriers in bilingual language production. The study found three common characteristics: enhanced formality, explicitation through elaboration and specification, and features resulting from processing strain. The key structural difference revealed that translated English tends to favor compressed and complex syntactic patterns, while the non-native indigenized variety predominantly uses clausal structures. Similarly, Hu et al. (2016) analyzed the Corpus of Translational English and the Freiburg-LOB Corpus of British English to identify linguistic features that distinguish translated texts from non-translated English writings. By examining 96 lexical, syntactic, and textual features, they found distinctive traits such as a higher frequency of the passive voice, greater use of abstract nouns, and more frequent explicit connectives. These features were shown to co-occur systematically, forming a "translational" dimension. Wang and Liu (2024) employed MDA to examine chairman's statements in corporate annual reports, revealing significant differences between translated and non-translated English statements in dimensions such as informational density and context independence. These studies collectively contribute to a deeper under-

standing of linguistic variations across genres in both original and translated texts.

3. The Study

Both translation and L2 writing are forms of constrained communication, yet a key distinction exists between the two. Translation involves the production of text based on an existing source, whereas L2 writing requires creating content independently (Lanstyák and Heltai 2012). Student translators, who generally possess adequate L2 proficiency but are still developing their translation expertise, may face varying levels of constraint. These constraints can influence their selection and application of translation strategies during the text production process (Kwok et al. 2023). Despite the growing body of literature on both L2 writing and translation, there is a notable gap in systematic comparisons of the linguistic features of L2 learner translation and L2 writing. To address this gap, the present study seeks to investigate the linguistic differences between L2 learner translation and L2 writing, employing MDA (Biber 1988) framework. The study aims to answer the following research questions:

RQ1: In which language dimensions do L2 writing and L2 learner translation differ?

RQ2: What individual linguistic features characterize the differences between L2 writing and L2 learner translation?

RQ3: What are the potential factors that account for these differences if they are identified?

4. Methods

4.1 Corpus

The International Corpus of English - Hong Kong (ICE-HK) is part of the broader *International Corpus of English (ICE)* project, which was initiated by Greenbaum (1988). Specifically, *ICE-HK* was developed in the early 1990s under the direction of Bolt and Bolton (1996). This corpus was designed to collect comparable English language data from various regional varieties. In line with the overall structure of the *ICE* project, *ICE-HK* focuses on gathering English language samples from Hong Kong's Chinese-speaking population, who are native Cantonese speakers and have completed their primary and secondary education in Hong Kong. The corpus includes a wide range of text types, encompassing both spoken and written communication across various genres. Given its composition, *ICE-HK* is a representative corpus of English as a Second Language (ESL) as it is used in Hong Kong (Nelson 2006). As such, it serves as a valuable resource for studying the distinctive features of Hong Kong English within the global context of World Englishes.

The *Parallel Learner Translation Corpus (PLTC)* is a project at The Hong Kong Poly-

technic University, developed to align with the written component (specifically the printed subcategory) of *ICE-HK* in terms of text categories and proportional size. The *PLTC* aims to document English translations produced by learners across various written registers in Hong Kong, including academic writing and fictional works. The corpus was developed through a systematic two-phase process. In the first phase, Chinese source materials were created through professional translations of *ICE-HK* texts from English into Chinese. Quality assurance was ensured through rigorous copyediting and collaborative review processes. The second phase involved learner translations from Chinese to English, contributed by undergraduate translation students from The Hong Kong Polytechnic University. These students, native Cantonese speakers who had completed their secondary education in Hong Kong, translated the texts into English. This methodologically sound compilation process resulted in a corpus that effectively captures the characteristics of learner translations in the Hong Kong context, making it a valuable resource for research in translation studies and second language acquisition. *PLTC* is a project at The Hong Kong Polytechnic University, designed to match the written component (specifically the printed subcategory) of *ICE-HK* in terms of text categories and proportional size. The *PLTC* aims to document English translations produced by learners across various written registers in Hong Kong, including academic writing and fictional works. *PLTC* has been developed through a systematic two-phase process. The first phase established the Chinese source materials through professional translations of *ICE-HK* texts from English into Chinese, with quality assurance maintained through rigorous copyediting and collaborative review processes. The second phase comprised learner translations from Chinese to English, contributed by undergraduate translation students from The Hong Kong Polytechnic University. Those translation students, were native Cantonese speakers who had completed their secondary education in Hong Kong. This methodologically sound compilation process resulted in a corpus that effectively captures the characteristics of learner translations in the Hong Kong context, making it a valuable resource for research in translation studies and second language acquisition. The development of *PLTC* involves a two-phase process. The initial phase focuses on preparing Chinese source materials. This is accomplished by having professional translators translate *ICE-HK* texts from English into Chinese. These translations undergo a rigorous quality control process, involving copyediting and collaborative review by both translators and copyeditors. The second phase centers on collecting learner translations from Chinese to English. Undergraduate students in their second to fourth year of translation studies at The Hong Kong Polytechnic University are invited to take part in via email. Interested students provide consent, and the researchers then select participants based on specific criteria. The eligible participants must be fluent Cantonese speakers who completed their secondary education in Hong Kong. This approach allows for the creation of a corpus that reflects the characteristics of learner translations in Hong Kong, providing valuable data for research in translation studies and second language acquisition.

The *PLTC* was initially designed to match the size and structure of *ICE-HK*. However, the recent emergence of Generative AI tools, such as ChatGPT, introduced an unexpected complication. Some participants began using these tools for translation tasks, leading to the inclusion of AI-generated content in the corpus. This AI-generated content, or "noise", posed a potential threat to the integrity of the *PLTC*, as it could obscure the true characteristics of L2 learner translations. To address this challenge and preserve the corpus' utility for studying L2 learner translation, we decided to focus on a subset of 53 translation samples from the *PLTC* (hereafter referred to as *PLTC*). These samples were specifically selected because they were completed without the assistance of Generative AI tools, ensuring they accurately reflect the L2 learners' translation abilities. To ensure a balanced comparison, we selected a corresponding set of 53 writing samples from the *ICE-HK* corpus (hereafter referred to as *ICE-HK*). This paired approach allows for a more controlled and reliable comparison between L2 learner translations and L2 writing texts, while mitigating the potential bias introduced by AI-assisted translations. Both *ICE-HK* and *PLTC* encompass six genres: academic writing, popular writing, reportage, instructional writing, persuasive writing, and creative writing. Each text is approximately 2,000 words in length. *ICE-HK* contains 127,835 tokens and 12,704 types, while *PLTC* has 125,178 tokens and 11,880 types (see Table 1). In this study, *ICE-HK* represents L2 writing, while *PLTC* represents L2 learner translation.

Table 1: Composition of the corpora

Corpus	***ICE-HK***	***PLTC***
Genres	Academic writing, popular writing, reportage,	
	instructional writing, persuasive writing, creative writing	
Number of texts	53	53
	(A total of 53 printed written texts are extracted to match the 53 samples of *PLTC*)	
Tokens	127,835	125,178
Types	12,704	11,880
TTR	9.94%	9.49%
STTR	41.47%	41.05%

4.2 Data analysis

This study employed MDA (Biber 1988) framework, which examines 67 linguistic features. While more recent MDA models exist, such as Xiao's (2009) 140-feature model and Hu et al.'s (2016) 96-feature model, Biber's (1988) approach was selected due to its robust theoretical and methodological foundations. It is particularly well-suited for exploring the quantitative distribution of textual features across different registers using multivariate statis-

tical techniques (Biber 2006). MDA identifies significant co-occurrence patterns of linguistic features within a language (Biber and Conrad 2001, 5). Biber's (1988) MDA framework encompasses 23 genres, including 17 written and 6 spoken genres (see Appendix). The model features six key dimensions: 1) Involved versus Informational Production, 2) Narrative versus Non-Narrative Concerns, 3) Explicit versus Situation-Dependent Reference, 4) Overt Expression of Persuasion, 5) Abstract versus Non-Abstract Information, and 6) On-line Informational Elaboration (p.115). Its versatility has been demonstrated in applications across a variety of registers, including annual reports (e.g., Wang and Liu 2024), novels (Su and Liu 2022), and scholarly articles (Liu and Xiao 2022).

For our analysis, we retrieved dimension and linguistic feature scores for individual texts from both the *ICE-HK* and *PLTC* corpora. To address our research questions, we adopted a systematic approach to data analysis. This began with assessing the distributional properties of the dependent variables. We conducted normality tests, including the Shapiro-Wilk test and visual inspection of Q-Q plots, to determine whether the dependent variables followed a normal distribution. The results of these normality tests are presented in Table 2. Four of the six dimensions deviated from a normal distribution. Consequently, the non-parametric Wilcoxon signed-rank test was employed to compare the ranks of *ICE-HK* and *PLTC*. We set the significance level at $\alpha = .05$ for all statistical tests, balancing the minimization of Type I errors (false positives) with maintaining adequate statistical power. This threshold indicates that results with p-values less than .05 were considered statistically significant, meaning we can be 95% confident that the observed differences were not due to chance. This comprehensive statistical approach allowed us to effectively quantify and compare the linguistic characteristics of learner translations and L2 writing, providing a solid foundation for addressing our research questions and drawing meaningful conclusions about the relationship between the two.

5. Results

Table 3 presents the descriptive statistics for the six dimensions of *ICE-HK* and *PLTC*, alongside the genre most closely associated with each dimension. Figure 1 provides boxplots of the six dimensions for both corpora. Although both *ICE-HK* and *PLTC* share the same closest genre, *general narrative exposition*, when considering the entire corpus, the two corpora are most closely aligned with this genre. However, a closer examination of individual dimensions reveals significant differences between the corpora in some areas. Specifically, they exhibit the same closest genre in four dimensions: Dimension 1 (Academic Prose), Dimension 2 (Conversation), Dimension 3 (Academic Prose), and Dimension 5 (Press Reportage). In contrast, Dimensions 4 and 6 show notable differences between the corpora. The results of the Wilcoxon signed-rank test indicate significant differences between *ICE-HK* and *PLTC* in Dimensions 1, 4, and 6, as shown in Table 4. In these three dimensions, *ICE-HK* demonstrates

statistically lower values compared to *PLTC*.

Table 2: Results of Shapiro-Wilk (normality test)

	Statistic	df	Sig.
Dimension 1	0.864	106	< .001
Dimension 2	0.964	106	.006
Dimension 3	0.979	106	.090
Dimension 4	0.960	106	.003
Dimension 5	0.983	106	.194
Dimension 6	0.958	106	.002

Table 3: Descriptive statistics for *ICE-HK* and *PLTC*

Dimension	***ICK-HK (n=53)***					***PLTC (n=53)***				
	Mean	**SD**	**Min**	**Max**	**Closest Genre**	**Mean**	**SD**	**Min**	**Max**	**Closest Genre**
Dimension 1	-11.35	8.66	-27.08	24.30	Academic prose	-10.63	8.69	-21.26	27.13	Academic prose
Dimension 2	-0.49	3.24	-5.52	6.44	Conversation	-0.50	3.10	-5.29	5.83	Conversation
Dimension 3	4.77	3.38	-2.71	12.12	Academic prose	4.89	3.55	-3.19	12.62	Academic prose
Dimension 4	0.55	3.17	-5.00	10.43	Prepared speeches	1.07	3.32	-3.83	10.16	General fiction
Dimension 5	1.52	2.23	-3.29	7.37	Press reportage	1.57	2.12	-2.52	6.23	Press reportage
Dimension 6	-1.37	0.85	-2.57	1.57	Personal letters	-0.56	1.29	-3.02	3.15	Official documents
Closest genre			General narrative exposition					General narrative exposition		

Dimension 1: *Involved Versus Informational Discourse*

In the context of multidimensional analysis, Dimension 1 examines the contrast between interactive and information-focused communication. This dimension assesses the extent to which a given text prioritizes the conveyance of information or engagement in interaction. Texts that score low on this scale tend to be rich in informational content, whereas those with higher scores typically exhibit a greater degree of interactivity between the author and the audience (Biber 1988). The mean scores of *ICE-HK* and *PLTC* on this dimension are relatively low, with *ICE-HK* at -11.35, *PLTC* at -10.63, indicating that both of them are largely informational in nature. However, the Wilcoxon test results reveal significant differences between them, with the *ICE-HK* being lower than *PLTC*, which indicates that L2 writing is more informationally dense than L2 learner translation. *ICE-HK* and *PLTC* differ significantly in 9 linguistic features on this dimension, with seven involved features and two informational features (see Table 5), the boxplots of which are shown in Figure 2 and Figure 3 respectively.

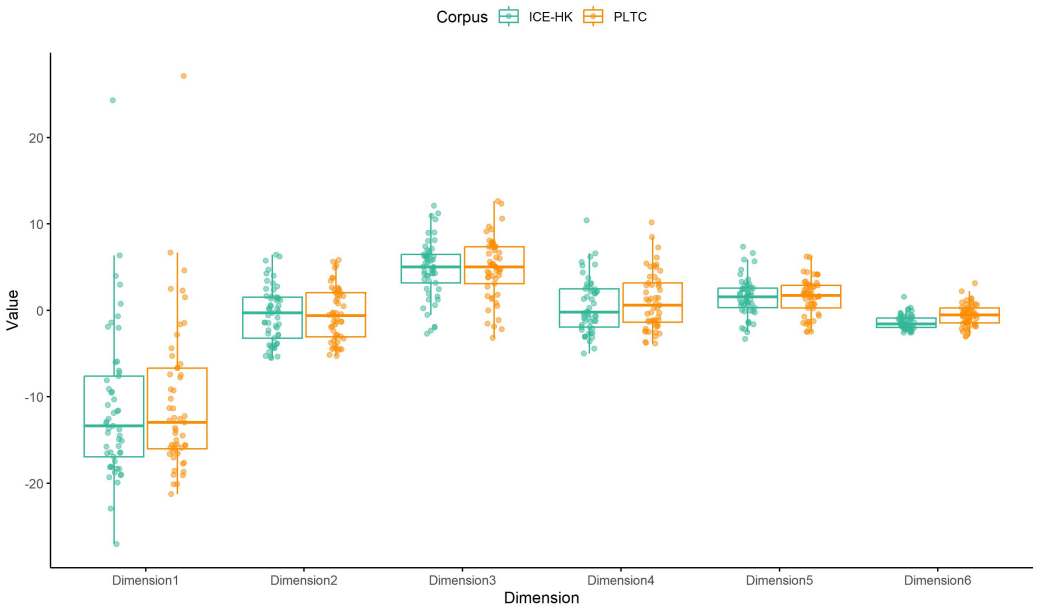


Figure 1: Boxplots of the six dimensions of *ICE-HK* and *PLTC*

Table 4: Wilcoxon test results for the comparison between the two corpora

Dimension	***PLTC (n=53) VS. ICE-HK (n=53)***						
	Negative ranks		**Positive ranks**		**Ties**	**Z**	***p***
	N	**Mean rank**	**N**	**Mean rank**	**N**		
Dimension 1	18	24.33	35	28.37	0	-2.457	.014*
Dimension 2	29	25.57	24	28.73	0	-.230	.818
Dimension 3	24	25.48	29	28.26	0	-.91	.357
Dimension 4	16	22.78	36	28.15	1	-2.955	.003*
Dimension 5	23	27.70	30	26.47	0	-.695	.487
Dimension 6	12	16.08	41	30.20	0	-4.626	<.001*

Note: Negative ranks indicate that scores for *PLTC* are lower than those for *ICE-HK*.
Positive ranks indicate that scores for *PLTC* are higher than those for *ICE-HK*.
Ties indicate no difference between *PLTC* and *ICE-HK* scores.
Significance threshold: $p < .05$.

Specifically, *ICE-HK* is significantly lower than *PLTC* in six involved features: second person pronouns, analytic negation, demonstrative pronouns, "be" as main verb, sentence relatives, possibility modals and in one informational feature (total other nouns). However, *ICE-HK* is significantly higher than *PLTC* in subordinator "that" deletion (involved feature) and in total prepositional phrases (informational feature). In sum, the comparison suggests that L2 writing tends to be more informationally dense than L2 learner translation, pointing to the possibility that these two types of texts are shaped by different communicative purposes and constraints. L2 writing, with its greater informational density, likely reflects a more formal, content-driven

approach, while L2 learner translation may prioritize interactional strategies, possibly due to the influence of the source text or the learner's focus on conveying meaning within a bilingual context.

Example 1 illustrates differences in verb choice and pronoun usage between *ICE-HK* (L2 writers) and *PLTC* (learner translators). In describing a modern, tasteful setting, the L2 writer in the *ICE-HK* text writes, "such as one would find amongst the younger affluent professionals in a place like Hong Kong," using the impersonal pronoun "one" and a more formal, descriptive structure. In contrast, the learner translator in the *PLTC* text describes the same setting with, "It is like what you can see in young, rich, successful professionals in a place similar to Hong Kong", opting for the second-person pronoun "you" and a direct, conversational tone. The choice of "such as one would find" in *ICE-HK* conveys a detached, formal style, typical of informational discourse, whereas "it is like what you can see" in *PLTC* creates a more immediate, interactive effect. These shifts reflect the learner translator's tendency toward a more direct and engaging style, using both "be" as a main verb and personal pronouns, while the L2 writer maintains a formal, impersonal approach.

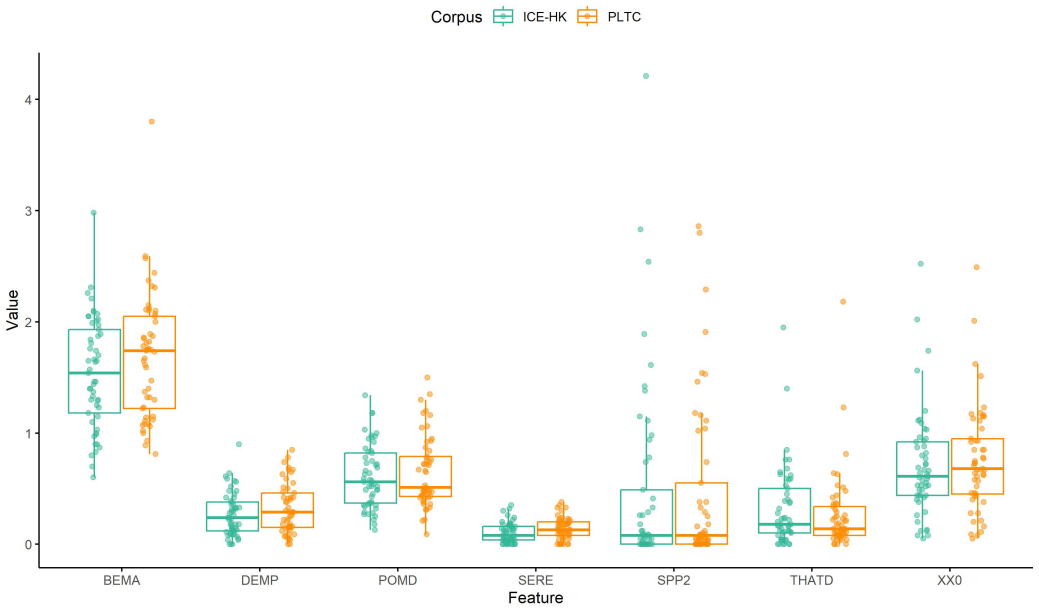
Example 1:

The backdrop and sidedrops (made up of vertical panels of cloth in a gradation of light colors from cream to yellow to gold to green) and the costumes (silvery gray silk with faint, simple designs for both sexes) suggest a modern setting which has good taste, such as one would find amongst the younger affluent professionals in a place like Hong Kong. (*ICE-HK*)

The background, side curtains (composed of vertical cloth panels, color changes from cream to gold, then to green) and costumes (made from silver-grey silk with simple pattern on both male and female costumes) showed an environment of a tasteful modern society. It is like what **you** can see in young, rich, successful professionals in a place similar to Hong Kong. (*PLTC*)

Dimension 4: *Overt Expression of Persuasion*

The fourth dimension of the analysis focuses on the explicit use of persuasive language. This dimension is distinctive in that it only includes features with positive loadings, which reflect the extent to which a text openly presents the author's viewpoints and arguments. Texts scoring high on this dimension tend to clearly express the author's stance, whether through direct argumentation, subjective assessment, or attempts to influence the reader's perception of events' likelihood or desirability (Biber, 1988). Analyzing the descriptive data reveals that both *ICE-HK* and *PLTC* show positive mean scores for overt persuasion, with *ICE-HK* at 0.55 and PLTC at 1.07. Statistical analysis using the Wilcoxon test highlights a significant difference between the two corpora, with *ICE-HK* scoring lower on this dimension (see Table 6). This suggests that L2 learner translation (*PLTC*) employs more overtly persuasive language than L2 writing (*ICE-HK*). Further analysis reveals significant differences in two out of the six linguistic features on this dimension. Boxplots of these features are presented in Figure 4.

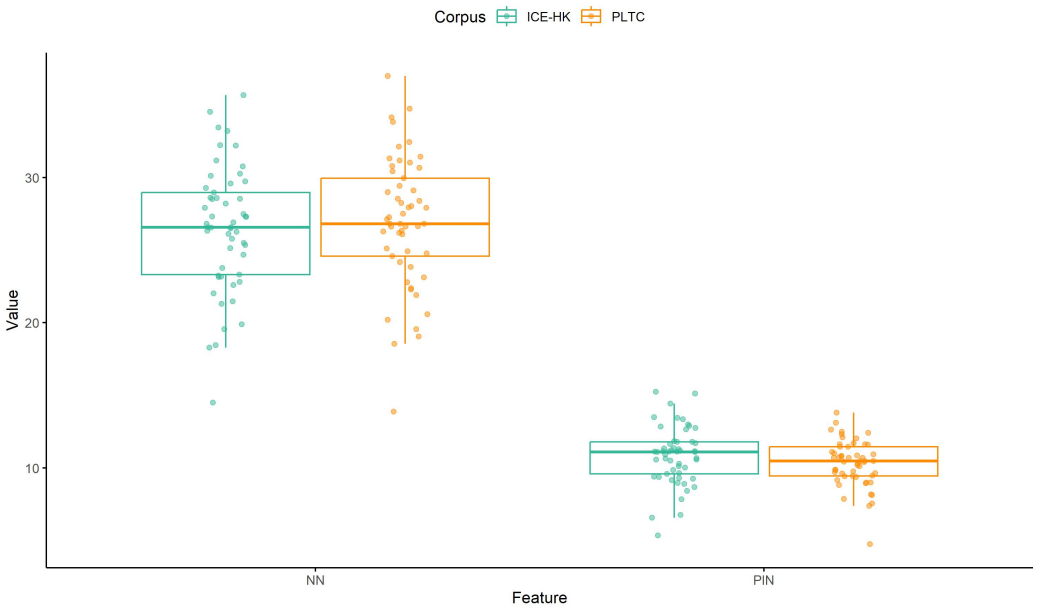


BEMA: "Be" as main verb; DEMP: Demonstrative pronouns;
POMD: Possibility modals; SERE: Sentence relatives; SPP2: Second person pronouns;
THATD: Subordinator "that" deletion; XX0: Analytic negation

Figure 2: Boxplots of the involved features exhibiting significant differences between the two corpora on Dimension 1

Specifically, *ICE-HK* is significantly lower than *PLTC* in the use of conditional adverbial subordinators and split auxiliaries. Conditional adverbial subordinators, such as "if" or "unless", are frequently used to express hypothetical or uncertain situations, which can be a strategy to introduce persuasive scenarios or arguments in translation. The lower frequency of these features in *ICE-HK* suggests that L2 writing tends to avoid direct hypothetical scenarios or conditional reasoning, which may be less persuasive in nature. Meanwhile, split auxiliaries—such as "do" in "do not" or "did not"—are often used in persuasive contexts to emphasize negation or to make a statement more forceful. The higher frequency of split auxiliaries in *PLTC* indicates that L2 learner translation may rely more on these structures to express negation and strengthen the persuasive impact of the text. In sum, the differences between *ICE-HK* and *PLTC* in their use of persuasive features suggest that L2 learner translation incorporates more overt strategies to influence the reader, possibly due to the nature of the task, where the translator's goal is to convey the source text's persuasive intent. In contrast, L2 writing, with its more measured use of persuasion, may prioritize clarity and objectivity over persuasion, reflecting a different communicative purpose.

Example 2 illustrates a distinctive pattern in the use of conditional structures between *ICE-HK* (L2 writers) and *PLTC* (learner translators). The learner translator in the *PLTC* text



NN: Total other nouns; PIN: Total prepositional phrases

Figure 3: Boxplots of the informational features exhibiting significant differences between the two corpora on Dimension 1

relies on a straightforward conditional structure, using the conditional adverbial subordinator ”if” twice, as in “if Hong Kong cannot catch up” and “if countries choose.” In contrast, the L2 writer in the *ICE-HK* text employs a more varied approach, using “if” once and incorporating the conditional “should” in a more formal expression, “should countries choose.” This shift from the formal, implicit conditional “should” in *ICE-HK* to the more explicit and repeated use of “if” in *PLTC* suggests a preference by learner translators for clear, accessible conditional markers. The *ICE-HK* text’s choice of “should” as an alternative conditional marker provides a nuanced, formal tone, aligning with informational discourse conventions, while *PLTC*’s repeated use of “if” creates a direct, easily interpretable structure. This pattern reflects the learner translator’s tendency toward explicit conditional relationships, which may enhance readability but with less syntactic variety than is found in the L2 writing.

Example 2: **If** Hong Kong does not keep pace with the green movement emerging in industrialised nations, trade performance here could decline should countries choose to use environmental issues to spark trade wars. (*ICE-HK*)

If Hong Kong cannot catch up on the Green Movement initiated by industrialized countries, Hong Kong could fail upon its trade performance **if** countries choose to use the environmental issues to trigger the trade war. (*PLTC*)

Dimension 6: *On-Line Information Elaboration*

The sixth dimension in Biber’s (1988) framework distinguishes between texts produced

Table 5: Comparison for linguistic features exhibiting significant differences between the two corpora on Dimension 1

Linguistic feature	***PLTC (n=53) VS ICE-HK (n=53)***						
	Negative Ranks		**Positive Ranks**		**Ties**	**Z**	***p***
	N	**Mean rank**	**N**	**Mean rank**	**N**		
Involved features							
Subordinator "that" deletion	29	28.53	19	18.34	5	-2.458	.014
Second person pronouns	9	13.94	20	15.48	24	-1.997	.046
Analytic negation	15	22.37	37	28.18	1	-3.222	.001
Demonstrative pronouns	19	19.63	31	29.10	3	-2.555	.011
"Be" as main verb	12	22.83	41	28.22	0	-3.910	<.001
Sentence relatives	13	15.38	28	23.61	12	-2.989	.003
Possibility modals	20	22.48	31	28.27	2	-2.002	.045
Informational features							
Total other nouns	17	21.79	35	28.79	1	-2.901	.004
Total prepositional phrases	37	32.32	16	14.69	0	-4.254	<.001

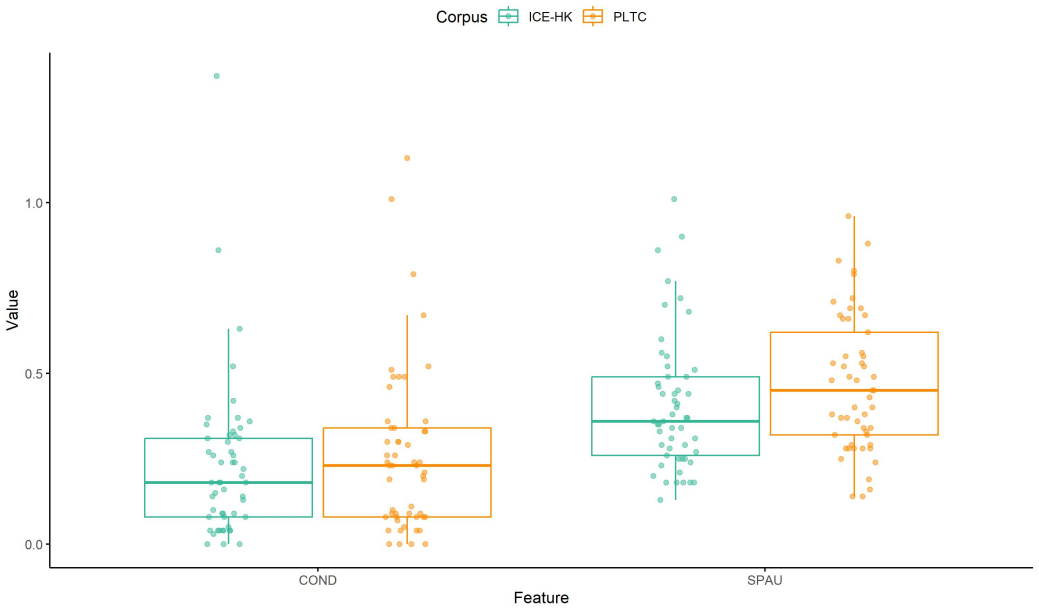
Note: Negative ranks indicate that scores for *PLTC* are lower than those for *ICE-HK*.
Positive ranks indicate that scores for *PLTC* are higher than those for *ICE-HK*.
Ties indicate no difference between *PLTC* and *ICE-HK* scores.

Table 6: Comparison for linguistic features exhibiting significant differences between the two corpora on Dimension 4

Linguistic feature	***PLTC (n=53) VS ICE-HK (n=53)***						
	Negative Ranks		**Positive Ranks**		**Ties**	**Z**	***p***
	N	**Mean rank**	**N**	**Mean rank**	**N**		
Conditional adverbial subordinators	12	21.63	29	20.74	12	-2.219	.026
Split auxiliaries	18	22.56	34	28.59	1	-2.578	.010

Note: Positive ranks indicate that scores for *PLTC* are higher than those for *ICE-HK*.
Ties indicate no difference between *PLTC* and *ICE-HK* scores.

under real-time conditions and those that are more carefully crafted. Texts scoring high on this dimension tend to feature an abundance of "that" clauses in various positions, signaling a more spontaneous, less planned production process. As Kruger and van Rooy (2016) suggest, this dimension reflects less cognitively demanding elaboration strategies, akin to casual conversation, rather than the complex, integrated structures typical of carefully planned texts. The analysis of *ICE-HK* and *PLTC* reveals negative mean scores on this dimension (*ICE-HK*: -1.37; *PLTC*: -0.56). These negative values indicate that both corpora exhibit characteristics of planned discourse, rather than impromptu speech. However, a significant difference emerges between them, with *ICE-HK* scoring notably lower than *PLTC* (see Table 7). This suggests that L2 writing (*ICE-HK*) contains fewer linguistic features typically associated with spontaneous or real-time communication than L2 learner translation (*PLTC*).



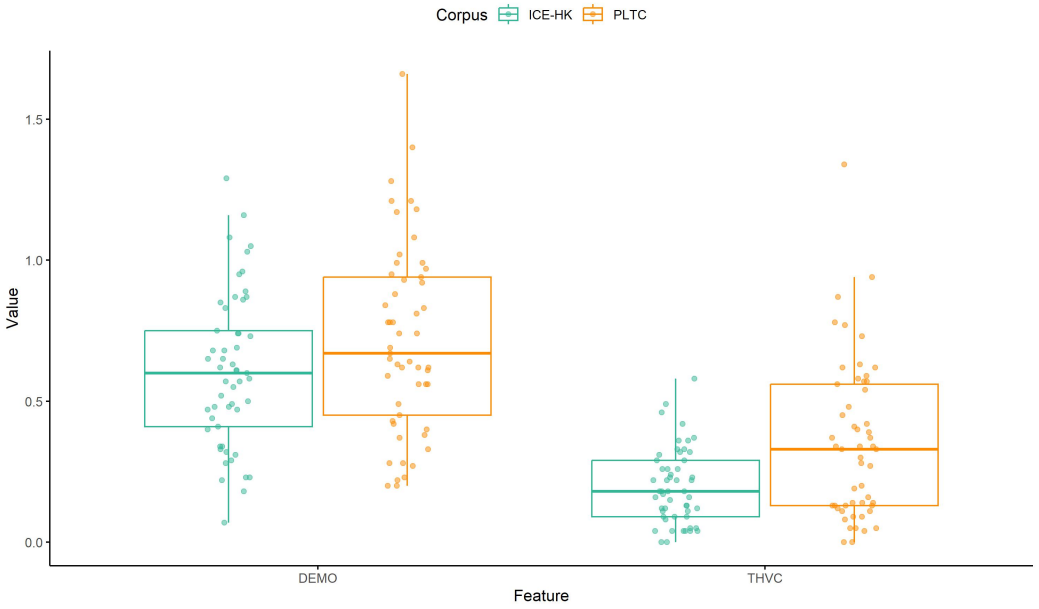
COND: Conditional adverbial subordinators; SPAU: Split auxiliaries

Figure 4: Boxplots of the linguistic features exhibiting significant differences between the two corpora on Dimension 4

ICE-HK and *PLTC* differ significantly in two out of four linguistic features on this dimension. Boxplots of these features are presented in Figure 6. Specifically, *ICE-HK* is significantly lower than *PLTC* in the use of "that" verb complements and demonstrative pronouns. The lower frequency of "that" verb complements in *ICE-HK* suggests that L2 writing tends to avoid constructions typical of spontaneous speech, such as those used in real-time conversation to clarify or specify actions. Similarly, the reduced use of demonstratives in *ICE-HK* further supports the idea that L2 writing is more formal and planned, prioritizing precision over immediacy. In sum, the significant differences between *ICE-HK* and *PLTC* highlight that L2 writing tends to feature more carefully structured, planned discourse, while L2 learner translation reflects a greater reliance on linguistic features indicative of spontaneous, less cognitively demanding production. Example 3 highlights differences in the use of referential language between *ICE-HK* (L2 writers) and *PLTC* (learner translators). In the *ICE-HK* text, the L2 writer uses the determiner "such" to refer to previously mentioned challenges, opting for the phrase "such potential hurdles." In contrast, the learner translator in the *PLTC* text selects the demonstrative "these" with "obstacles", resulting in "these potential obstacles". This choice of a demonstrative in *PLTC* reflects a tendency observed in texts influenced by real-time constraints, where demonstratives serve as explicit cohesive devices, creating clear, immediate textual references. Although the translation process itself is not real-time, it mirrors such constraints, favoring direct and accessible referential strategies.

Example 3: Such potential hurdles notwithstanding, an industrial strategy can be formulated without too great a difficulty. (*ICE-HK*)

Despite **these** potential obstacles, there would not be much difficulty in developing an industrial strategy. (*PLTC*)



DEMO: Demonstratives; THVC: "That" verb complements

Figure 5: Boxplots of linguistic features exhibiting significant differences between the two corpora on Dimension 6

Table 7: Comparison for linguistic features exhibiting significant differences between the two corpora on Dimension 6

Linguistic feature	***PLTC (n=53) VS ICE-HK (n=53)***						
	Negative Ranks		**Positive Ranks**		**Ties**	**Z**	***p***
	N	**Mean rank**	**N**	**Mean rank**	**N**		
"That" verb complements	7	16.36	41	25.89	5	-4.858	<.001
Demonstratives	20	17.65	31	31.39	2	-2.906	.004

Note: Positive ranks indicate that scores for *PLTC* are higher than those for *ICE-HK*. Ties indicate no difference between *PLTC* and *ICE-HK* scores.

6. Discussion

The present study employed Biber's (1988) Multidimensional Analysis to compare L2 learner translations and L2 writing in English. The results revealed that L2 writing is char-

acterized by greater informational density, less overt expression of viewpoints, and a more planned, rather than spontaneous, structure compared to L2 learner translations. However, it is important to emphasize that the significant differences observed between the two genres are specific to this study, which compared these two genres, and should be understood as relative along a continuum. These differences do not suggest that L2 writing is inherently more formal or sophisticated, nor do they imply that learner translations are deficient in other contexts. Rather, the differences reflect the relative characteristics of the two genres in this analysis and highlight their distinct features when compared to one another. The following discussion interprets these findings, examines potential underlying factors, and situates them within the broader domains of second language acquisition and translation studies.

The higher informational density observed in L2 writing, compared to translation, suggests that when composing original texts in English, writers tend to produce more content-rich and lexically dense language. This may be attributed to the academic contexts in which L2 writing is often situated, where concise expression and efficient information conveyance are emphasized (Hyland 2006). This finding is consistent with research indicating that advanced L2 writers typically aim to demonstrate linguistic sophistication through information-dense structures (Crossley and McNamara 2014). In contrast, L2 learner translations, produced by students working from Chinese versions of *ICE-HK* texts, exhibit lower informational density. This pattern aligns with the translation universal of simplification proposed by Baker (1996), which suggests that translated texts are generally simpler and less complex than their non-translated counterparts. Statistical evidence from this study supports Baker's simplification universal, showing reduced informational density in L2 learner translations. Moreover, L2 writing involves stages of planning, drafting, and revision, allowing writers to refine their ideas and enhance informational content (Manchón 2011). In translation, however, constraints from the source text and the need for semantic equivalence may limit opportunities for further elaboration (Dam-Jensen and Heine 2013).

The less explicit marking of points of view in professional L2 writing, compared to learner translation, may be influenced by factors such as explicitation in back-translation, proficiency level, and the particular demands of the translation process (Becher 2010). In this study, L2 English texts were first translated into Chinese by professional translators and then back-translated into English by learners, creating a parallel structure for comparing learner translations and professional L2 writing. This back-translation process can lead to an explicitation effect, which "spells things out rather than leaving them implicit" (Baker 1996, 180), as learners add markers like conditional adverbial subordinators and split auxiliaries to clarify logical relationships and convey the original intent more clearly. This explicitation reflects learners' tendency toward clarity and directness, as they may feel a need to use straightforward markers to capture the meaning of the source text accurately, potentially bridging gaps between languages (Blum-Kulk 1986). Learners' use of explicit markers, such as split auxiliaries and

overt conditionals, may also be influenced by their English proficiency level (Hinkel 2003). Less experienced writers might lean on clearer structures to express complex ideas, which can introduce elements of spoken language into their translations (Biber and Gray 2010). These features create a more cohesive and accessible style, though sometimes with less subtlety (Halliday and Hasan 1976). In contrast, professional L2 writing, created directly in English without a source text's influence, often takes a more nuanced and implicit approach to conveying relationships and viewpoints. Professional writers tend to favor a subtler style, aligning with formal writing conventions that emphasize informational density and minimize overt cohesion devices. This difference in approach suggests that learner translators prioritize clarity and accessibility, while professional L2 writers rely on more indirect and refined linguistic choices appropriate for a formal register.

L2 writing contains greater informational elaboration under lower real-time constraints compared to L2 learner translation, a distinction that becomes evident in the differing use of "that" verb complements and demonstratives. In learner translations, "that" verb complements are frequently used to clarify logical relationships, making the text more straightforward and enhancing readability, especially under the demands of translation, where learners must balance accuracy and clarity across languages (Hansen-Schirra et al. 2007). Demonstratives also appear more frequently in learner translations, serving as explicit cohesive devices that make references clear and accessible to the reader. These features reflect the cognitive constraints of translation, where learners often prioritize directness and explicit markers to ensure that the meaning of the source text is preserved and easily understood in the target language (Halverson 2003). In contrast, professional L2 writing in *ICE-HK* tends to minimize these explicit markers, favoring subtler cohesion techniques that create a more integrated and elaborative structure. Without the influence of a source text, L2 writers can plan and revise their work, allowing them to avoid overt markers like "that" complements and demonstratives. This controlled, carefully crafted approach aligns with the conventions of formal writing, where implicit cohesion and informational density are valued over directness (Hyland 2004). These differences underscore distinct cognitive constraints: learner translations, shaped by the immediate need to transfer meaning accurately between languages, often lean toward accessibility and clarity, while L2 writing benefits from a single-language focus that supports more nuanced and implicit cohesion techniques, resulting in a denser, more sophisticated text structure.

7. Conclusion

This study applied Biber's (1988) Multidimensional Analysis to compare English texts produced through L2 learner translation and direct L2 writing. The findings reveal significant differences between these two modes of L2 production: L2 writing exhibits higher information density, less explicit marking of point of view, and operates under lower real-time constraints

than L2 learner translation. These results offer valuable insights into the distinct nature of L2 language production across different tasks, with important implications for second language acquisition, translation studies, and language pedagogy. The observed differences suggest that direct L2 writing and L1 to L2 translation involve distinct cognitive processes, supporting theories of interlanguage variability and emphasizing the task-specific nature of language use in L2 production. This research contributes to our understanding of bilingual language processing, cross-linguistic influence, and the constraints imposed by various language production tasks. The unique linguistic profiles of L2 writing and learner translation underscore the concept of these tasks as forms of constrained communication, each imposing specific limitations on language production.

While this study offers valuable insights, some limitations should be acknowledged. First, the source texts for learner translation were based on professional translations of original L2 writing, which allowed for a direct comparison between the two modes. However, these professional translations may not fully capture the natural qualities of native English writing, as they still carry characteristics of translation. This aspect may limit the extent to which findings can generalize to naturally produced L2 writing. Second, this study focuses on the final written products by using a corpus-based approach to compare L2 writing and L2 learner translation, without examining the cognitive processes involved in producing these texts. Future research could explore these underlying cognitive processes more directly, employing methods such as think-aloud protocols or eye-tracking during writing and translation tasks to gain deeper insights into the mental activities and strategies that shape these outputs.

Funding information

This research was supported by Guangzhou Railway Polytechnic Newly Introduced Talent Research Project (Project Number: GTXYR2426).

References

Baker, Mona. 1993. “Corpus Linguistics and Translation Studies: Implications and Applications.” In *Text and Technology: In Honour of John Sinclair*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233–250. Amsterdam: John Benjamins.

Baker, Mona. 1996. “Corpus-based Translation Studies: The Challenges That Lie Ahead.” In *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, edited by Harold Somers, 175–186. Amsterdam: John Benjamins.

Becher, Viktor. 2010. “Abandoning the notion of ‘translation-inherent’ explicitation: Against a dogma of translation studies.” *Across languages and cultures* 11 (1): 1-28.

Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.

Biber, Douglas. 2006. *University language: A corpus-based study of spoken and written reg-*

isters. Amsterdam: John Benjamins Publishing Company.

Biber, Douglas, and Bethany Gray. 2010. "Challenging stereotypes about academic writing: Complexity, elaboration, explicitness." *Journal of English for Academic Purposes* 9 (1): 2-20.

Biber, Douglas, and Susan Conrad. 2001. *Variation in English multi-Dimensional studies*. Hoboken: Taylor and Francis.

Blum-Kulka, Shoshana. 1986. "Shifts of cohesion and coherence in translation." In *Interlingual and Intercultural Communication. Discourse and Cognition in Translation and Second Language Acquisition*, edited by Juliane House and Shoshana Blum-Kulka, 17-35. Tübingen: Gunter Narr.

Bolt, Philip, and Kingsley Bolton. 1996. "The International Corpus of English in Hong Kong." In *Comparing English Worldwide: The International Corpus of English* edited by Sidney Greenbaum, 197–214. Oxford: Clarendon Press.

Carl, Michael, and Barbara Dragsted. 2012. "Inside the monitor model: Processes of default and challenged translation production." In *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation*: TC3 II edited by Oliver Czulo and Silvia Hansen-Schirra, 5-30. Berlin: Language Science Press.

Chen, Jiaxin, Dechao Li, and Kanglong Liu. 2024. "Unraveling cognitive constraints in constrained languages: A comparative study of syntactic complexity in translated, EFL, and native varieties." *Language Sciences* 102: 101612.

Crossley, Scott A., and Danielle S. McNamara. 2012. "Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication." *Journal of Research in Reading* 35 (2): 115-135.

Crossley, Scott A., and Danielle S. McNamara. 2014. "Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners." *Journal of Second Language Writing* 26: 66-79.

Dam-Jensen, Helle, and Carmen Heine. 2013. "Writing and translation process research: Bridging the gap (Introduction)." *Journal of Writing Research* 5 (1): 89-101.

Gardner, Sheena, Hilary Nesi, and Douglas Biber. 2019. "Discipline, level, genre: Integrating situational perspectives in a new MD analysis of university student writing." *Applied Linguistics* 40 (4): 646-674.

Göpferich, Susanne, and Riitta Jääskeläinen. 2009. "Process research into the development of translation competence: Where are we, and where do we need to go?." *Across languages and cultures* 10 (2): 169-191.

Halliday, Michael Alexander Kirkwood, and Ruqaiya Hasan. 1976. *Cohesion in english*. London: Longman.

Halverson, Sandra L. 2003. "The cognitive basis of translation universals." *Target. International Journal of Translation Studies* 15 (2): 197-241.

Hansen-Schirra, Silvia, Stella Neumann, and Erich Steiner. 2007. "Cohesive explicitness and explicitation in an English-German translation corpus." *Languages in contrast* 7 (2): 241-265.

Hinkel, Eli. 2011. "What research on second language writing tells us and what it doesn't." In *Handbook of Research in Second Language Teaching and Learning* edited by Eli Hinkel 2, 523-538. New York/Oxon: Routledge.

Hinkel, Eli. 2003. "Simplicity without elegance: Features of sentences in L1 and L2 academic texts." *Tesol Quarterly* 37 (2): 275-301.

Hyland, Ken. 2004. *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.

Hyland, Ken. 2006. "The 'other' English: thoughts on EAP and academic writing." *The European English Messenger* 15 (2): 34-38.

Hu, Xianyao, Richard Xiao, and Andrew Hardie. 2016. "How do English translations differ from non-translated English writings? A multi-feature statistical model for linguistic variation analysis." *Corpus Linguistics and Linguistic Theory* 15 (2): 347-382.

Kormos, Judit. 2011. "Task complexity and linguistic and discourse features of narrative writing performance." *Journal of Second Language Writing* 20 (2): 148-161.

Kruger, Haidee, and Bertus van Rooy. 2016. "Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English." *English World-Wide* 37 (1): 26-57.

Kunilovskaya, Maria, Natalia Morgoun, and Alexey Pariy. 2018. "Learner vs. professional translations into Russian: Lexical profiles." *Translation & Interpreting: The International Journal of Translation and Interpreting Research* 10 (1): 33-52.

Kwok, Ho Ling, Sara Laviosa, and Kanglong Liu. 2023. "Lexical simplification in learner translation: A corpus-based approach." *Research in corpus linguistics* 11 (2): 103-124.

Lanstyák, István, and Pál Heltai. 2012. "Universals in language contact and translation." *Across Languages and Cultures* 13 (1): 99-121.

Liu, Jing, and Liang Xiao. 2022. "A multi-dimensional analysis of conclusions in research articles: Variation across disciplines." *English for Specific Purposes* 67: 46-61.

Liu, Kanglong, and Muhammad Afzaal. 2021. "Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification." *Plos One* 16 (6): e0253454.

Liu, Kanglong, Zhongzhu Liu, and Lei Lei. 2022. "Simplification in translated Chinese: An entropy-based approach." *Lingua* 275: 103364.

Manchón, Rosa. 2011. "Writing to learn the language: Issues in theory and research." In *Learning-to-write and writing-to-learn in an additional language* edited by Rosa Manchón, 61-82. Amsterdam: John Benjamins.

Manchón, Rosa. 2012. *L2 writing development: Multiple perspectives*. Boston, MA: De Gruyter Mouton.

McEnery, Tony, Richard Xiao, and Yukio Tono. 2006. *Corpus-based language studies: An advanced resource book*. Oxon: Routledge.

Nelson, Gerald. 2006. *The ICE Hong Kong Corpus: User Manual*. London: University College London.

Polio, Charlene. 2017. "Second language writing development: A research agenda." *Language Teaching* 50 (2): 261-275.

Pym, Anthony. 2015. "Translating as risk management." *Journal of pragmatics* 85: 67-80.

Ren, Chaowang, and Xiaofei Lu. 2021. "A multi-dimensional analysis of the Management's Discussion and Analysis narratives in Chinese and American corporate annual reports." *English for Specific Purposes* 62: 84-99.

Snape, Neal, Hironobu Hosoi, and Mari Umeda. 2023. "Second language processing of English definite noun phrases by Spanish speakers and Japanese speakers." *Journal of Second Language Studies* 6 (2): 319-347.

Su, Yanfang, and Kanglong Liu. 2022. "Orality in Translated and Non-Translated Fictional Dialogues." In *Advances in Corpus Applications in Literary and Translation Studies* edited by Riccardo Morrato, and Defeng Li, 119-137. London: Taylor and Francis Ltd.

Toury, Gideon. 2012. *Descriptive Translation Studies – and Beyond*. Amsterdam: John Benjamins.

Vo, Sonca. 2019. "Use of lexical features in non-native academic writing." *Journal of Second Language Writing* 44: 1-12.

Wang, Zhongliang, Andrew KF Cheung, and Kanglong Liu. 2024. "Entropy-based syntactic tree analysis for text classification: a novel approach to distinguishing between original and translated Chinese texts." *Digital Scholarship in the Humanities* fqae030.

Wang, Zhongliang, Kanglong Liu, and Riccardo Moratto. 2023. "A corpus-based study of syntactic complexity of translated and non-translated chairman's statements." *Translation & Interpreting: The International Journal of Translation and Interpreting Research* 15 (1): 135-151.

Wang, Zhongliang, Ming Liu, and Kanglong Liu. 2024. "Utilizing machine learning techniques for classifying translated and non-translated corporate annual reports." *Applied Artificial Intelligence* 38 (1), 2340393.

Wang, Zhongliang, and Kanglong Liu. 2024. "Linguistic variations between translated and non-translated English chairman's statements in corporate annual reports: A multidimensional analysis." *SAGE Open* 14 (2): 1-17.

Wang, Wenyu, and Qiufang Wen. 2002. "L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers." *Journal of second language writing* 11 (3): 225-246.

Wetzel, Mathis, Sandrine Zufferey, and Pascal Gygax. 2024. "The role of complexity in the ability to notice incoherent uses of connectives for L2." *Journal of Second Language Studies* 7 (1): 44-74.

Wu, Shiyu, Dilin Liu, and Shaoqiang Huang. 2023. "The Effects of Over-and Under-Specified Linguistic Input on L2 Online Processing of Referring Expressions." *Journal of Psycholinguistic Research* 52 (1): 283-305.

Xiao, Richard. 2009. "Multidimensional analysis and the study of world Englishes." *World Englishes* 28 (4): 421-450.

Xu, Cui, and Dechao Li. 2024. "More spoken or more translated? Exploring the known unknowns of simultaneous interpreting from a multidimensional analysis perspective." *Target* 36 (3): 445-480.

Yao, Yao., Dechao Li, Yingqi Huang, Y and Zhonggang Sang. 2024. "Linguistic variation in mediated diplomatic communication: a full multi-dimensional analysis of interpreted language in Chinese Regular Press Conferences." *Humanities and Social Sciences Communications* 11 (1): 1-12.

Appendix: Genres in MDA (Biber 1988, 67)

	Written genres		**Spoken genres**
1	Press reportage	1	Face-to-face conversation
2	Editorials	2	Telephone conversation
3	Press reviews	3	Public conversations, debates, and Interviews
4	Religion	4	Broadcast
5	Skills and hobbies	5	Spontaneous speeches
6	Popular lore	6	Planned speeches
7	Biographies		
8	Official documents		
9	Academic Prose		
10	General fiction		
11	Mystery fiction		
12	Science fiction		
13	Adventure fiction		
14	Romantic fiction		
15	Humor		
16	Personal letters		
17	Professional letters		