Research Article

Zhongliang Wang, Han Xu and Kanglong Liu*

Lexical Complexity in Corporate Communication: A Corpus-Based Study of Translated and Non-Translated Chairman's Statements

https://doi.org/10.1515/csh-2024-0026
Received October 3, 2024; accepted October 25, 2024; published online November 14, 2024

Abstract: This study examines the lexical complexity of translated and non-translated chairman's statements in corporate annual reports through a corpus-based analysis. Drawing on a comprehensive corpus of English chairman's statements, including both English translations from Chinese and original English texts from the United States and Hong Kong, we analyze various dimensions of lexical complexity. Key measures include lexical diversity, such as the number of different words (NDW) and type-token ratio (TTR) variants, lexical density, and verb sophistication. The results present a nuanced view of lexical complexity in translated texts. Translated chairman's statements show lower complexity across NDW-related metrics, Corrected TTR, Root TTR, Bi-logarithmic TTR, Uber index, lexical word variation, verb variation, noun variation, and adverb variation. However, they display higher complexity in lexical density and verb sophistication compared to non-translated statements. These findings suggest that while translation may simplify certain aspects of vocabulary use, it can also produce more information-dense texts with more sophisticated verb usage. This paradox challenges traditional views on translation universals and underscores the intricate nature of translated corporate communications.

***Corresponding author: Kanglong Liu**, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China, E-mail: kl.liu@polyu.edu.hk
Zhongliang Wang, School of Foreign Languages and Commerce, Guangzhou Railway Polytechnic, Guangzhou, China, E-mail: wangzhongliang@gtxy.edu.cn
Han Xu, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China, E-mail: han12.xu@polyu.edu.hk

Keywords: lexical complexity; chairman's statements; corpus-based translation studies; translation universals

1 Introduction

The chairman's statement is a crucial component of a company's annual report, playing a vital role in conveying the company's performance, strategy, and outlook to stakeholders (Rutherford 2005). Given its importance, previous studies have explored the linguistic features of these statements (e.g., Wang, Li, and Cao 2012; Zhuang, Li, and Li 2024), though much of the research has focused on monolingual statements. This approach does not fully account for the increasingly multilingual nature of corporate communication, as multinational corporations frequently translate chairman's statements into multiple languages. Although studies such as Wang, Liu, and Moratto (2023) and Wang and Liu (2024) have begun to examine syntactic complexity and linguistic variations in translated versus untranslated statements, significant gaps remain, particularly regarding the lexical features. More research is needed to address these gaps and deepen our understanding of the linguistic complexity involved in translated corporate communications, which play a pivotal role in global business interactions (Jeanjean, Lesage, and Stolowy 2010).

Exploring the distinction between translated and untranslated texts has been a central focus in corpus-based translation studies (Baker 1993; Laviosa 2002; Wu, Lei, and Li 2024). This line of inquiry has contributed to the development of the translation universals theory, which suggests that translated texts exhibit certain linguistic features that set them apart from non-translated texts, irrespective of the languages involved (Chesterman 2004). Researchers have explored these potential universals through various approaches, including simplification, explicitation, normalization, and leveling out (Kwok, Laviosa, and Liu 2023; Liu, Liu, and Lei 2022; Liu, Yin, and Cheung 2024; Wang, Cheung, and Liu 2024a; Wang, Liu, and Liu 2024b). Among these approaches, a particular focus has been placed on lexical features, as they provide quantifiable insights into the nature of translated language (Xiao and Dai 2014). Lexical complexity, which encompasses various aspects of vocabulary use including diversity, density, and sophistication, is a critical factor in the readability and effectiveness of corporate communications (Loughran and McDonald 2014). It influences how investors, analysts, and other stakeholders perceive and interpret the information presented. In the context of translation, lexical complexity takes on additional significance, as it may be shaped by the translation process and the linguistic differences between source and target languages. Translation studies have long been concerned with how translated texts differ from non-translated texts in the same language, with "translation universals" being a prominent concept introduced by

Baker (1993). This concept proposes that translated texts may exhibit common features – such as simplification or explicitation – across different language pairs. However, these hypotheses remain subject to ongoing debate and empirical testing in various genres and language combinations (Chesterman 2004).

This study seeks to address the gap in research by conducting a corpus-based analysis of lexical complexity in translated and non-translated English chairman's statements. Including the number of different words (NDW) – related measures, type-token ratio (TTR) variants, lexical density, and verb sophistication, this study aims to identify patterns that reveal the linguistic nature of translated corporate communications. These findings will enhance understanding of translation effects within specialized discourses. The implications of this study extend to translation practices in corporate settings, the interpretation of financial communications by international stakeholders, and the broader field of Translation Studies. By providing empirical evidence on the lexical characteristics of translated chairman's statements, this research aims to contribute to more effective cross-linguistic corporate communication strategies and inform ongoing theoretical discussions about the nature of translated language.

2 Related Work

2.1 Simplification as a Translation Universal

The notion of translation universals was defined as "the features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker 1993, p. 243). Some of the most well-known translation universals include simplification, explicitation, and normalization. Explicitation refers to translation's tendency to "spell things out rather than leave them implicit" (Baker 1996, p. 180). It can be reflected both at the lexical or syntactic level. Normalization can also be termed "conventionalization" (Mauranen 2007) and means "the tendency to exaggerate features of the target language and to conform to its typical patterns" (Baker 1996, p. 183). Consequently, translated texts seem more normal than target texts (Xiao 2010). Among these proposed universals, simplification has received considerable attention in translation studies (Liu and Afzaal 2021). Simplification refers to the "tendency to simplify the language used in translation" (1996, p. 181).

Research on translation universals utilizes three primary methodological approaches, each of which requires specific types of corpora. The first is the comparable corpus approach, which compares translated texts with non-translated texts in the target language to identify universal features of translated language (Baker

1993). The second is the parallel corpus approach, which involves comparing source texts with their translations to examine translational patterns (Olohan 2004). The third is the multi-parallel corpus approach, which analyzes multiple translations of the same source text to explore translation variability (Granger 2003). Each of these approaches requires distinct combinations of corpora, including translated texts, original texts in the target language, and source texts. Regardless of the approach, certain essential criteria must be met for the corpora used: sufficient size for statistically significant results (Zanettin 2014), balanced representation of authors, topics, and source languages (Olohan 2004), and comparability in terms of genre, register, time period, and subject matter (Laviosa 2002).

Simplification has been explored at various linguistic levels, including the lexical level (e.g., Laviosa 1998) and the syntactic level (e.g., Wang, Liu, and Moratto 2023). For example, Blum-Kulka and Levenston (1983, p. 119) examined translations from Hebrew to English and identified lexical simplification, defined as "the process and/or result of making do with fewer words." Vanderauwera (1985) examined syntactic simplification, discovering that translated texts often replace finite clauses with non-finite ones to streamline complex syntactic structures. In addition to syntactic changes, Vanderauwera also investigated stylistic simplification in translated texts. However, the generalizability of these findings is limited due to the small sample sizes and the absence of robust statistical methods in these earlier studies.

Both lexical and syntactic indicators have been explored in studies of translation universals, with lexical features receiving more frequent attention. Among the key contributions in this area is Laviosa's (1998) study, which examined lexical density, sentence length, and the frequency of commonly used words in English translations of narrative prose. Using data from the English Comparable Corpus (ECC), the study identified four significant ways in which translated English differs from native English, highlighting distinct lexical patterns in translated texts. First, compared to original texts, translated texts exhibit a lower lexical density and a greater percentage of grammatical terms. Second, translated texts employ more high frequency terms. Third, the most frequently-used words are repeated more often in translated texts. Fourth, fewer lemmas are included in the most commonly-used terms in translated texts than in original texts (ibid.). Olohan (2004) used lexical diversity as a metric to compare translated and native English fiction, finding that translated works employed fewer color synonyms than their native counterparts. Similarly, Pastor et al. (2008) explored simplification by utilizing natural language processing tools, readability formulae, and other indices. Their study revealed that non-translated texts exhibited higher lexical density and richness compared to translated texts.

While simplification has been the focus of numerous studies, it remains a contentious topic among translation universals due to conflicting findings that

challenge its universality (Liu and Afzaal 2021). For instance, longer mean sentence lengths have been observed in translated texts (Laviosa 1998; Xiao and Yue 2009), which contradicts Malmkjær's (1997) assumptions. Jantunen (2004) found a higher frequency of modifiers in translated texts, offering no consistent evidence to support translation universals. Similarly, Mauranen (2000) studied collocations and found that multi-word patterns in untranslated texts were clearer and more stable, while translated texts displayed unusual collocations. Contradicting the simplification hypothesis, Ferraresi et al. (2018) found that translated texts were more complex and exhibited higher lexical density. It is also important to note that research on lexical simplification has predominantly focused on isolated or selectively chosen lexical features, lacking systematic and comprehensive analyses of broader lexical complexity indices. Moreover, most studies on simplification have concentrated on literary translation (Blum-Kulka and Levenston 1983; Laviosa 1998), leaving non-literary translation, such as financial texts, largely unexplored. To address this gap, it is crucial to investigate the lexical complexity of translated annual reports to determine whether simplification occurs in this genre. Following Baker's (1993) comparable corpus approach, this study compares translated chairman's statements from mainland Chinese companies listed on the Hong Kong Stock Exchange with non-translated statements from Hong Kong companies listed on the Hong Kong Stock Exchange and U.S. companies listed on the New York Stock Exchange and NASDAQ. This design provides a diverse representation of non-translated texts, enhancing the robustness of the investigation into translation universals in corporate communication. By analyzing the lexical complexity of translated and non-translated chairman's statements, we can deepen our understanding of whether simplification universals apply to this context. This research can contribute to the broader study of translation universals, highlighting similarities and differences between genres and registers in translated texts.

2.2 Lexical Complexity in Language-Related Research

Lexical complexity, also referred to as lexical richness, is a measure of how diverse and sophisticated the vocabulary is in a text or by a language user. It involves several key dimensions, including lexical density, lexical sophistication, and lexical variation (Bui 2021; Lu 2012; Shi and Lei 2022). Lexical density is defined as "the ratio of the number of lexical words (Nlex) to the number of words (N)" (Lu 2012, p. 196). It is a measure that determines the proportion of meaningful words or content words in a given text, relative to the total number of words (Ure 1971). Lexical diversity, also referred to as lexical variation or lexical range, assesses the range of different words employed in a text (Lu 2012). Lexical diversity can be easily measured using various methods, such as counting the number of unique words in a text or employing the

type-token ratio (TTR) (Jarvis 2013). However, TTR can be influenced by the length of the text (Engber 1995). To address this issue, scholars have put forward alternative calculations of TTR, such as corrected TTR (Carroll 1964) and root TTR (Guiraud 1960). Although the type-token ratio and its alternative calculations have some controversies, some previous studies used them in assessing lexical diversity of texts (Daller, Van Hout, and Treffers-Daller 2003; Lu 2012). Lexical sophistication refers to lexical rareness, which measures "the proportion of relatively unusual or advanced words" in the texts (Read 2000, p. 203). In contrast to lexical diversity, which focuses on the variety of words, lexical sophistication evaluates the complexity and rarity of the vocabulary. It is typically measured by calculating the proportion of less frequent words, as determined by reference frequency lists or corpora. Researchers compare the average word frequency in a text to that of a reference corpus (Lu and Hu 2022). These frequency-based measures indicate how much a text relies on uncommon or advanced vocabulary. Texts with a higher proportion of common words are generally considered less sophisticated (Crossley et al. 2007).

Lexical complexity has been extensively studied in various linguistic domains, including second language proficiency assessment, second language writing, academic writing, and translation studies. Researchers have explored its use in gauging second language learners' proficiency (e.g., Zareva, Schwanenflugel, and Nikolova, 2005) and examined its relationship with writing quality in language testing (Laufer and Nation 1995; Yu 2010). Lei and Yang (2020) investigated its role in academic writing across different levels of expertise, finding that academic proficiency had a greater impact on research article writing than nativeness. In translation studies, lexical density and variation have been analyzed to test the simplification hypothesis at the lexical level (Laviosa 1998; Pastor et al. 2008; Volansky, Ordan, and Wintner 2015). These diverse applications highlight the significance of lexical complexity as a linguistic measure across multiple areas of language research and assessment.

2.3 Research Gaps and Questions

Based on the preceding review, two research gaps emerge: (1) Previous studies have largely concentrated on isolated lexical features in translated texts, lacking systematic and comprehensive analyses of lexical complexity to investigate lexical simplification. (2) Research has primarily focused on simplification in literary translation, leaving non-literary translation, especially financial translation, underexplored. This study aims to address these gaps by applying a comprehensive set of lexical complexity measures to examine the differences between translated and untranslated chairman's statements. Specifically, the study seeks to explore the following research questions:

RQ1: In what ways do translated chairman's statements differ from untranslated ones in terms of lexical complexity?

RQ2: What factors might account for any identified similarities or differences between the two?

3 Methods and Procedures

3.1 Data Description

For this study, we compiled a corpus of chairman's statements (COCS) from three regions: Mainland China, Hong Kong, and the United States. The data were sourced from the annual reports of listed companies. The chairman's statements from Mainland Chinese and Hong Kong companies were obtained from companies listed on the Hong Kong Exchanges and Clearing Market (HKEX), while the statements from American companies were drawn from firms listed on either the New York Stock Exchange (NYSE) or the National Association of Securities Dealers Automated Quotations (NASDAQ). Unlike companies in Mainland China, which often produce annual reports first in Chinese and then translate them into English, companies in Hong Kong typically prepare their reports directly in English (Wang 2014). This longstanding practice is attributed to Hong Kong's bilingual environment and the strong influence of Western stock exchange systems (Setter, Wong, and Chan 2010). As a result, Hong Kong offers a unique perspective on a variety of English that reflects both Western and Eastern linguistic and cultural influences (Wang 2014). Hong Kong's position as a bridge between these two worlds provides a rich comparative framework, revealing subtleties in English usage within a bilingual and bicultural context. The inclusion of Hong Kong English texts enriches the diversity of non-translated corporate communications, aligning with Laviosa's (2002) emphasis on comparing translated texts with a broad spectrum of non-translated texts. By integrating both U.S. and Hong Kong English texts, this study captures a more comprehensive representation of original English corporate communications, thereby strengthening the robustness of comparisons with translated texts.

The compilation of the three corpora followed a two-phase process. In the first phase, companies were selected based on their sectors to ensure balance across the corpora. The sector distribution was aligned with that of stocks listed on the HKEX, with 100 companies selected from 10 sectors[1] for each corpus.

1 The ten sections include Finance, Consumer Services, Finance, Consumer Services, Industrial Services, Electronic Technology, Consumer Non-Durables, Commercial Services, Technology Services, Process Industries, Producer Manufacturing, Health Technology.

Table 1: Descriptive statistics of the corpus of chairman's statements (COCS).

Corpus	Number of texts	Total tokens	Year	Average tokens
USCS	100	183,851	2020	1,839
HKCS	100	113,818	2020	1,138
CMCS	100	154,965	2020	1,550

A corresponding number of companies from each sector were chosen for the Mainland Chinese (CMCS), Hong Kong (HKCS), and U.S. (USCS) corpora. This method ensured a balanced and comparable representation of industries across regions. Table 1 provides a detailed breakdown of the descriptive data of the compiled corpora. In addition, the original Chinese versions of the chairman's statements from Mainland Chinese companies were also collected as a reference, enabling a more comprehensive analysis of translation and its impact on the linguistic features of corporate communications.

3.2 Data Analysis

The chairman's statements in the three corpora were analyzed using the Lexical Complexity Analyzer, a computational tool developed by Lu (2012) to assess lexical complexity in written English. The tool offers 25 different measures of lexical density, variation, and sophistication, as detailed in Table 2. Its reliability depends on the accuracy of the POS tagging and lemmatization, which are managed by the Stanford tagger, known for its 99 % accuracy (Toutanova et al. 2003). Therefore, the overall reliability of the analyzer is strengthened by the strong performance of the tagger and lemmatizer (Jin et al. 2021).

This study employed statistical techniques to compare the means or mean ranks of the 25 lexical complexity measures across the three corpora in order to investigate potential differences. The three corpora were treated as the fixed factor, while the 25 lexical complexity measures served as the dependent variables. To assess the normality of the dependent variables, the Shapiro–Wilk test was conducted, revealing that most variables did not follow a normal distribution. Consequently, the Kruskal–Wallis test was applied to identify statistically significant differences among the three corpora. For pairwise comparisons, Mann–Whitney U tests were conducted to assess significant differences between groups. The initial significance level (alpha) was set at 0.05, and to account for the increased risk of Type I error due to multiple comparisons, a Bonferroni correction was applied. Since three Mann–Whitney U tests were performed, the adjusted alpha level was set at $0.05/3 = 0.0167$. The use of these statistical methods enabled the identification of significant differences in lexical complexity among the corpora. By applying

Table 2: 25 measures of lexical complexity (from Lu 2012, pp. 193–195).

Measure	Code	Measure	Code
Lexical density	LD	Root TTR	RTTR
Lexical sophistication-I	LS1	Bilogarithmic TTR	LogTTR
Lexical sophistication-II	LS2	Uber index	Uber
Verb sophistication-I	VS1	Lexical word variation	LV
Corrected VS1	CVS1	Verb variation-I	VV1
Verb sophistication-II	VS2	Squared VV1	SVV1
Number of different words	NDW	Corrected VV1	CVV1
NDW (first 50 words)	NDWZ	Verb variation-II	VV2
NDW (expected random 50)	NSWERZ	Noun variation	NV
NDW (expected sequence 50)	NEWESZ	Adjective variation	AdjV
Type-token ratio	TTR	Adverb variation	AdvV
Mean segmental TTR (50)	MSTTR	Modifier variation	ModV
Corrected TTR	CTTR		

the Bonferroni correction, we minimized the likelihood of Type I errors, thereby ensuring the reliability and validity of the study's findings.

4 Results

Table 3 presents the descriptive statistics for lexical complexity across the three corpora. Since 15 of the 25 dependent variables did not follow a normal distribution, non-parametric methods, specifically the Kruskal–Wallis test and Mann–Whitney U tests, were employed to compare the mean ranks between the USCS, CMCS, and HKCS corpora to explore their differences.

The results of the Kruskal–Wallis test reveal significant differences among USCS, CMCS, and HKCS in all 25 indices, as shown in Table 4. Mann–Whitney U tests were then conducted for pairwise comparisons, with the alpha level corrected to 0.0167. Table 5 demonstrates that USCS and CMCS differ significantly in 17 lexical complexity measures, including Lexical Density, Lexical Sophistication-I, NDW (across various ranges), TTR variants, Uber Index, Lexical Word Variation, Verb and Noun Variation, and Adverb Variation. Specifically, CMCS scores statistically lower than USCS in 15 measures, indicating that translated chairman's statements from Mainland China tend to use more limited and repetitive vocabulary compared to their American counterparts. However, CMCS scores higher in Lexical Density and Verb Sophistication-I, suggesting a focus on content-rich expressions and more sophisticated verb usage. Similarly, Table 6 shows that HKCS and CMCS differ in 16 lexical complexity measures. CMCS scores lower in 11 measures, such

Table 3: Descriptive statistics of lexical complexity of USCS, HKCS and CMCS.

Measure	USCS				HKCS				CMCS			
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
LD	0.56	0.03	0.50	0.63	0.56	0.03	0.49	0.62	0.57	0.03	0.51	0.63
LS1	0.59	0.03	0.49	0.66	0.56	0.04	0.44	0.66	0.57	0.05	0.43	0.67
LS2	0.53	0.05	0.36	0.62	0.51	0.04	0.36	0.60	0.54	0.05	0.39	0.62
VS1	0.47	0.08	0.26	0.68	0.52	0.07	0.38	0.71	0.52	0.09	0.34	0.83
VS2	47.18	22.09	6.56	148.58	33.15	13.95	7.00	61.07	43.48	18.91	7.41	97.93
CVS1	4.73	1.11	1.81	8.62	3.97	0.89	1.87	5.53	4.54	1.06	1.92	7.00
NDW	650.03	339.30	194.00	2,018.00	449.89	194.95	120.00	1,101.00	548.10	234.59	168.00	1,377.00
NDWZ	41.12	2.84	30.00	47.00	35.71	4.97	24.00	44.00	37.37	4.57	24.00	46.00
NDWERZ	41.80	1.25	38.80	44.30	40.72	1.49	37.70	43.90	40.43	1.38	37.40	44.30
NDWESZ	41.03	1.19	36.00	44.30	39.26	1.85	34.10	43.50	38.74	1.44	35.60	42.00
TTR	0.41	0.08	0.23	0.58	0.43	0.07	0.27	0.64	0.40	0.08	0.23	0.59
MSTTR	0.82	0.02	0.76	0.86	0.78	0.03	0.71	0.84	0.78	0.03	0.70	0.83
CTTR	10.92	1.83	7.36	17.15	9.44	1.51	5.75	12.93	9.95	1.46	6.05	13.74
RTTR	15.44	2.58	10.41	24.25	13.35	2.14	8.13	18.28	14.07	2.07	8.55	19.43
LOGTTR	0.88	0.02	0.83	0.91	0.88	0.02	0.84	0.92	0.87	0.02	0.82	0.91
UBER	25.75	2.11	21.08	33.36	24.33	2.26	19.57	30.66	23.98	2.13	16.21	28.32
LV	0.61	0.11	0.37	0.83	0.62	0.10	0.39	0.85	0.57	0.11	0.33	0.84
VV1	0.67	0.10	0.43	0.89	0.73	0.10	0.46	0.97	0.70	0.11	0.44	0.97
SVV1	95.53	40.63	32.36	269.36	62.53	23.20	17.29	117.88	76.46	28.64	17.55	159.16
CVV1	6.77	1.41	4.02	11.61	5.49	1.06	2.94	7.68	6.07	1.20	2.96	8.92
VV2	0.17	0.03	0.11	0.25	0.15	0.03	0.09	0.24	0.14	0.03	0.08	0.21
NV	0.57	0.11	0.33	0.83	0.57	0.10	0.35	0.82	0.52	0.11	0.30	0.79
ADJV	0.12	0.02	0.07	0.20	0.12	0.02	0.07	0.21	0.11	0.03	0.06	0.19
ADVV	0.02	0.01	0.01	0.04	0.02	0.01	0.01	0.06	0.02	0.01	0.01	0.05
MODV	0.14	0.03	0.08	0.24	0.15	0.03	0.08	0.25	0.13	0.03	0.06	0.24

as TTR variants, Lexical Word Variation, and various measures of verb and noun variation. However, CMCS scores higher in six key areas: Lexical Density, Lexical Sophistication-II, Corrected VS1, Verb Sophistication-II, and NDW. This suggests that while CMCS may use a less varied vocabulary compared to HKCS, it employs a higher proportion of content words, more sophisticated verbs, and more advanced lexical choices.

Table 7 summarizes the pairwise comparison results between CMCS and USCS, and between CMCS and HKCS for each measure. The comparison between CMCS and USCS reveals that CMCS exhibits lower complexity in most measures but scores higher in lexical density and verb sophistication. This indicates a simplification trend in translated corporate communication from Mainland China compared to original American texts, but with a focus on content-rich and action-oriented

Table 4: Results of the Kruskal–Wallis test for lexical complexity measures in chairman's statements.

	Chi-square	df	Asymp. sig.
LD	22.938	2	<0.001[b]
LS1	33.977	2	<0.001[b]
LS2	22.463	2	<0.001[b]
VS1	25.098	2	<0.001[b]
VS2	26.395	2	<0.001[b]
CVS1	26.303	2	<0.001[b]
NDW	21.027	2	<0.001[b]
NDWZ	73.258	2	<0.001[b]
NDWERZ	48.731	2	<0.001[b]
NDWESZ	99.216	2	<0.001[b]
TTR	10.125	2	0.006[a]
MSTTR	115.016	2	<0.001[b]
CTTR	32.963	2	<0.001[b]
RTTR	32.936	2	<0.001[b]
LOGTTR	10.919	2	0.004[a]
UBER	34.240	2	<0.001[b]
LV	13.482	2	0.001[a]
VV1	14.166	2	0.001[a]
SVV1	41.284	2	<0.001[b]
CVV1	41.273	2	<0.001[b]
VV2	38.329	2	<0.001[b]
NV	15.132	2	0.001[a]
ADJV	7.267	2	0.026[a]
ADVV	27.491	2	<0.001[b]
MODV	13.188	2	0.001[a]

[a]$p < 0.05$. [b]$p < 0.001$.

language. In contrast, the comparison between CMCS and HKCS shows fewer differences overall, with CMCS still scoring higher in similar key measures, suggesting that corporate communication in Hong Kong shares some features with Mainland China but also retains certain characteristics closer to Western norms.

These findings point to several key implications. The translation process from Chinese to English results in some simplification, particularly when compared to American English. However, the consistently higher lexical density in CMCS suggests a cultural preference for information-rich expressions. The use of sophisticated verbs, especially compared to HKCS, may reflect a strategic emphasis on action-oriented language in Chinese corporate communication. The closer alignment between HKCS and CMCS indicates that Hong Kong's bilingual environment bridges some linguistic differences between Mainland China and the West.

Table 5: Mann–Whitney U tests for lexical complexity measures between USCS and CMCS.

Measure code	Measure	USCS VS CMCS ($N = 200$)				
		Mean rank		U	z	p
		USCS	CMCS			
LD	Lexical density	83.88	117.13	3,337.5	−4.089	<0.001[b]
LS1	Lexical sophistication-I	118.80	82.20	3,170	−4.485	<0.001[b]
LS2	Lexical sophistication-II	98.44	102.57	4,793.5	−0.506	0.613
VS1	Verb sophistication-I	84.04	116.97	3,353.5	−4.026	<0.001[b]
VS2	Corrected VS1	103.61	97.40	4,689.5	−0.759	0.448
CVS1	Verb sophistication-II	103.62	97.38	4,688	−0.762	0.446
NDW	Number of different words	107.81	93.20	4,269.5	−1.785	0.074
NDWZ	NDW (first 50 words)	126.54	74.46	2,396	−6.387	<0.001[b]
NDWERZ	NDW (expected random 50)	128.22	72.79	2,228.5	−6.775	<0.001[b]
NDWESZ	NDW (expected sequence 50)	139.56	61.44	1,094	−9.547	<0.001[b]
TTR	Type-token ratio	106.46	94.54	4,404	−1.457	0.145
MSTTR	Mean segmental TTR (50)	142.08	58.93	842.5	−10.221	<0.001[b]
CTTR	Corrected TTR	115.82	85.18	3,468	−3.743	<0.001[b]
RTTR	Root TTR	115.80	85.20	3,470	−3.738	<0.001[b]
LOGTTR	Bilogarithmic TTR	111.80	89.20	3,870	−2.805	0.005[a]
UBER	Uber index	122.76	78.24	2,774	−5.439	<0.001[b]
LV	Lexical word variation	110.37	90.64	4,013.5	−2.412	0.016[a]
VV1	Verb variation-I	93.18	107.82	4,268	−1.790	0.074
SVV1	Squared VV1	113.43	87.57	3,707	−3.159	0.002[a]
CVV1	Corrected VV1	113.47	87.54	3,703.5	−3.168	0.002[a]
VV2	Verb variation-II	125.19	75.82	2,531.5	−6.060	<0.001[b]
NV	Noun variation	113.94	87.06	3,656	−3.285	0.001[a]
ADJV	Adjective variation	102.87	98.13	4,763	−0.584	0.559
ADVV	Adverb variation	116.29	84.72	3,421.5	−4.222	<0.001[b]
MODV	Modifier variation	108.76	92.24	4,174	−2.031	0.042

[a]$p < 0.0167$ after Bonferroni correction ($p < 0.05/3$). [b]$p < 0.001$.

5 Discussion

5.1 Strategic Simplification and Sophistication in Translation

The findings of this study reveal a dual trend in the translated chairman's statements (CMCS), challenging traditional assumptions about translation universals. On one hand, the CMCS texts exhibit lower complexity in most lexical measures compared to the original English chairman's statements (USCS and HKCS), including the Number of Different Words (NDW), Type-Token Ratio (TTR), and variations in

Table 6: Mann–Whitney U tests for lexical complexity measures between HKCS and CMCS.

Measure code	Measure	HKCS VS CMCS ($N = 200$)				
		Mean rank		*U*	*z*	*p*
		HKCS	CMCS			
LD	Lexical density	83.44	117.56	3,294	−4.200	<0.001[b]
LS1	Lexical sophistication-I	98.21	102.80	4,770.5	−0.562	0.574
LS2	Lexical sophistication-II	83.07	117.94	3,256.5	−4.273	<0.001[b]
VS1	Verb sophistication-I	101.11	99.89	4,939	−0.149	0.881
VS2	Corrected VS1	84.02	116.98	3,352	−4.027	<0.001[b]
CVS1	Verb sophistication-II	84.04	116.96	3,354	−4.022	<0.001[b]
NDW	Number of different words	88.53	112.48	3,802.5	−2.926	0.003[a]
NDWZ	NDW (first 50 words)	91.38	109.62	4,088	−2.235	0.025
NDWERZ	NDW (expected random 50)	106.77	94.23	4,373	−1.533	0.125
NDWESZ	NDW (expected sequence 50)	110.70	90.30	3,980	−2.493	0.013[a]
TTR	Type-token ratio	113.80	87.21	3,670.5	−3.251	0.001[a]
MSTTR	Mean segmental TTR (50)	106.76	94.25	4,374.5	−1.537	0.124
CTTR	Corrected TTR	91.19	109.81	4,069	−2.275	0.023
RTTR	Root TTR	91.20	109.81	4,069.5	−2.274	0.023
LOGTTR	Bilogarithmic TTR	112.21	88.80	3,829.5	−2.912	0.004[a]
UBER	Uber index	104.32	96.68	4,618	−0.933	0.351
LV	Lexical word variation	115.09	85.92	3,541.5	−3.566	<0.001[b]
VV1	Verb variation-I	108.93	92.07	4,157	−2.061	0.039
SVV1	Squared VV1	86.23	114.77	3,573	−3.487	<0.001[b]
CVV1	Corrected VV1	86.24	114.77	3,573.5	−3.486	<0.001[b]
VV2	Verb variation-II	110.54	90.46	3,996	−2.468	0.014[a]
NV	Noun variation	114.60	86.41	3,590.5	−3.446	0.001[a]
ADJV	Adjective variation	110.53	90.47	3,997	−2.470	0.014[a]
ADVV	Adverb variation	118.37	82.64	3,213.5	−4.806	<0.001[b]
MODV	Modifier variation	114.82	86.19	3,568.5	−3.520	<0.001[b]

[a]$p < 0.0167$ after Bonferroni correction ($p < 0.05/3$). [b]$p < 0.001$.

verbs, nouns, and adverbs. This reduced complexity suggests a simplification process, where the vocabulary becomes more limited and repetitive. These results align with the simplification hypothesis, a well-established concept in translation studies, which posits that translated texts tend to use more common words and reduce variability to maintain clarity (Baker 1993). Our findings align with previous studies (Laviosa 1998; Pastor et al. 2008) that have shown translated texts tend to be less complex than non-translated texts. The directionality of translation may influence these results. In our study, the chairman's statements from mainland Chinese companies were likely translated from Chinese (A language) into English (B language).

Table 7: Significant results between CMCS and USCS, between CMCS and HKCS.

Measure code	Measure	USCS versus CMCS	HKCS versus CMCS
LD	Lexical density	<	<
LS1	Lexical sophistication-I	>	–
LS2	Lexical sophistication-II	–	<
VS1	Verb sophistication-I	<	–
VS2	Corrected VS1	–	<
CVS1	Verb sophistication-II	–	<
NDW	Number of different words	–	<
NDWZ	NDW (first 50 words)	>	–
NDWERZ	NDW (expected random 50)	>	–
NDWESZ	NDW (expected sequence 50)	>	>
TTR	Type-token ratio	–	>
MSTTR	Mean segmental TTR (50)	>	–
CTTR	Corrected TTR	>	–
RTTR	Root TTR	>	–
LOGTTR	Bilogarithmic TTR	>	>
UBER	Uber index	>	–
LV	Lexical word variation	>	>
VV1	Verb variation-I	–	–
SVV1	Squared VV1	>	>
CVV1	Corrected VV1	>	>
VV2	Verb variation-II	>	>
NV	Noun variation	>	>
ADJV	Adjective variation	–	>
ADVV	Adverb variation	>	>
MODV	Modifier variation	–	>

> Indicates the former is statistically higher than the latter; < indicates the former is statistically lower than the latter; – indicates there are no significant differences.

Although translating into a B language is less conventional, it is increasingly common in international business communication (Pokorn 2005). The lower complexity observed in NDW-related metrics, TTR-related metrics, the Uber index, and various word class variations (lexical word, verb, noun, adverb) aligns with expectations for translation into a non-native language. Translators working into their B language often have a smaller active vocabulary and may choose more familiar, frequently used words to ensure clarity and avoid errors (Campbell 2014), which could explain the observed vocabulary simplification.

However, the translated texts also exhibit higher lexical density and verb sophistication, revealing a deliberate complexity in specific areas. This paradoxical finding can be attributed to two primary factors. First, the Chinese language

structure inherently employs more verbs to express actions and processes, which may result in a higher proportion of verbs in the translated texts (Liu 2010). Second, translators may opt for sophisticated word choices to accurately convey culturally specific or abstract concepts that exist in Chinese but lack direct English equivalents. This reflects Jakobson's (1959) concept of "equivalence in difference," which proposes that creative and nuanced linguistic solutions are essential when confronting the challenges of translating between languages with distinct grammatical and lexical structures. To maintain the original meaning and nuance, translators utilize more advanced vocabulary, contributing to an increase in both lexical density and verb sophistication. This complexity is not accidental but rather a manifestation of the translator's need to bridge the gap between the source and target languages while ensuring that the text remains comprehensible and preserves its original intent. The translator's choices are guided by the necessity to strike a balance between faithfully conveying the source text's content and making it accessible to the target readership. Consequently, while certain facets of the translation process lead to simplification, others require a more sophisticated approach to capture the nuances of the original text. This delicate equilibrium between simplification and sophistication illustrates that translation is not merely a cognitive activity occurring within the translator's mind but also a form of cross-cultural social interaction (House 2014). As a result, translators must carefully navigate these dual objectives, tailoring their strategies based on the specific genre and communicative purpose of the text (Nord 1997). Simplified language aligns with modern digital reading preferences (Liu 2008) and regulatory requirements (Securities and Exchange Commission 1998). Ultimately, the interplay between simplification and sophistication in translated chairman's statements reflects the translators' efforts to produce a text that is both faithful to the original and effective within its new cultural and linguistic context.

5.2 Cultural Influences on Lexical Complexity in Corporate Communication

Chairman's statements serve to communicate key information about a company's performance, opportunities, challenges, and future plans in non-quantitative terms (Hu and Tan 2020). However, cultural differences influence how this information is presented, especially in translated chairman's statements, which must align with their source texts in Chinese. According to Wang and Liu's study (2024) using Biber's multidimensional analysis, translated chairman's statements from Mainland Chinese companies are more informational, narrative, context-independent, and less spontaneous compared to those from American and Hong Kong companies.

This study's findings emphasize the significant role of cultural factors in shaping the lexical complexity of translated chairman's statements. The higher lexical

density observed in CMCS, compared to USCS and HKCS, reflects a cultural preference in Mainland China for information-rich communication. Mainland Chinese corporate discourse tends to favor more indirect, contextualized, and writer-oriented texts, with a preference for providing extensive background information and details (Zhu 2005). This approach contrasts with Western corporate communication norms, which prioritize directness, concision, and reader-orientation, emphasizing clarity and efficiency.

Hong Kong's business communication follows Western norms, reflecting its history as an international financial hub and former British colony. This is evident in the common practice of preparing annual reports in English rather than Chinese (Courtis and Hassan 2002). By contrast, many Chinese Mainland companies listed on the Hong Kong Exchange maintain a more authoritative and formal communication style, likely influenced by the political structure, where CEOs of state-owned enterprises often resemble government officials more than business professionals (Yang, Wang, and Nie 2013). The differences in lexical complexity between translated and non-translated chairman's statements, therefore, reflect not only linguistic but also political and cultural distinctions. Moreover, the similarities between CMCS and HKCS highlight Hong Kong's bilingual and bicultural influence on corporate communication. As a bridge between Eastern and Western business cultures, Hong Kong's corporate discourse seemingly blends Western linguistic norms with Eastern rhetorical traditions. This cross-cultural context results in corporate texts that balance simplification with complexity, aligning with both local and international expectations.

These findings underscore the importance of considering regional linguistic norms and cultural preferences when analyzing translated corporate texts. Cultural factors shape both the language used and the broader communicative objectives, influencing how translators balance clarity and simplicity with professionalism and credibility. Ultimately, the level of lexical complexity of translated chairman's statements reflects a nuanced interaction between strategic linguistic choices and cultural influences, requiring translators to adapt their approach to the specific cultural and communicative context of each text.

6 Conclusions

This study of lexical complexity in translated and non-translated chairman's statements presents a nuanced view of translation effects in corporate communication. While translated texts show lower complexity in most measures, such as NDW and TTR indices, they also exhibit higher lexical density and verb sophistication, reflecting the intricate balancing act in translation. Translators simplify certain areas to

enhance readability, but simultaneously introduce complexity to maintain a formal and sophisticated tone.

The implications of these findings are important for both Translation Studies and corporate communication. For Translation Studies, this research provides evidence that different aspects of lexical complexity – such as diversity, density, and sophistication – may follow distinct patterns in translated texts. This underscores the need for a more nuanced approach to understanding translation universals, recognizing that translation is influenced by both linguistic and cultural factors. For corporate communication practitioners and translators, recognizing that translated chairman's statements are conditioned by various factors, such as the source language and cultural context, is essential. Understanding how these factors shape lexical complexity can guide the creation of effective texts for international audiences. From a methodological perspective, this study highlights the value of corpus-based approaches in uncovering subtle linguistic differences between translated and non-translated texts. It also demonstrates the importance of examining multiple measures of lexical complexity to gain a comprehensive understanding of the language used in corporate communication.

While this study provides valuable insights, there are some limitations to consider. First, the focus on the Chinese-English language pair in one direction allowed for a detailed analysis but may limit the generalizability of the findings to other languages or translation contexts. In addition, the study examined only one genre, specifically chairman's statements in annual reports. Although this allowed for an in-depth exploration, the results may not fully extend to other types of corporate communication. Future research could expand on these findings by exploring how lexical complexity influences reader comprehension and stakeholder perceptions. Additionally, examining other language pairs and genres would offer a broader understanding of how translation impacts corporate communication across different contexts.

Research funding: This work was supported by Guangzhou Railway Polytechnic (Newly Introduced Talent Research Project, GTXYR2426) and The Hong Kong Polytechnic University (https://doi.org/10.13039/501100004377, P0051009).

References

Baker, M. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology. In Honor of John Sinclair*, edited by M. Baker, G. Francis, and E. Tognini-Bonelli, 233—50. Amsterdam: John Benjamins.

Baker, M. 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 J. C. Sager, and H. L. Somers, 44—54. Amsterdam: John Benjamins.

Blum-Kulka, S., and E. Levenston. 1983. “Universals of Lexical Simplification.” In *Strategies in Interlanguage Communication*, edited by C. Faerch, and G. Kasper, 119—39. London: Longman.

Bui, Gavin. 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—67.

Campbell, S. 2014. *Translation into the Second Language*. London: Routledge.

Carroll, J. B. 1964. *Language and Thought*. Englewood: Prentice-Hall.

Chesterman, A. 2004. “Hypotheses about Translation Universals.” *Benjamins Translation Library* 50: 1—14.

Courtis, J. K., and S. Hassan. 2002. “Reading Ease of Bilingual Annual Reports.” *Journal of Business Communication* 39 (4): 394—413.

Crossley, S. A., D. F. Dufty, P. M. McCarthy, and D. S. McNamara. 2007. “Toward a New Readability: A Mixed Model Approach.” *Proceedings of the Annual Meeting of the Cognitive Science Society* 29 (29): 197—202.

Daller, H., R. Van Hout, and J. Treffers-Daller. 2003. “Lexical Richness in the Spontaneous Speech of Bilinguals.” *Applied Linguistics* 24 (2): 197—222.

Engber, C. A. 1995. “The Relationship of Lexical Proficiency to the Quality of ESL Compositions.” *Journal of Second Language Writing* 4 (2): 139—55.

Ferraresi, A., S. Bernardini, M. M. Petrović, and M. A. Lefer. 2018. “Simplified or Not Simplified? the Different Guises of Mediated English at the European Parliament.” *Meta* 63 (3): 717—38.

Granger, S. 2003. “The Corpus Approach: A Common Way Forward for Contrastive Linguistics and Translation Studies?” In *Corpus-Based Approaches to Contrastive Linguistics and Translation Studies*, 17—29. Brill.

Guiraud, P. 1960. *Problémes et méthodes de la statistique linguistique [Problems and Methods of Statistical Linguistics]*. Dordrecht: D. Reidel.

House, J. 2014. “Translation Quality Assessment: Past and Present.” In *Translation: A Multidisciplinary Approach. Palgrave Advances in Language and Linguistics*, edited by J. House, 241—64. London: Palgrave Macmillan.

Hu, C., and J. Tan. 2020. “A Multi-Dimensional Analysis of Register Features of Letters to Shareholders by Chinese and American Enterprises.” *Foreign Languages and their Teaching* 6: 66—75.

Jakobson, R. 1959. “On Linguistic Aspects of Translation.” In *On Translation*, edited by R. Brower, 232—9. Cambridge: Harvard University Press.

Jantunen, J. 2004. “Untypical Patterns in Translations. Issues on Corpus Methodology and Synonymity.” In *Translation Universals: Do They Exist?* edited by A. Mauranen, and P. Kujamäki, 101—26. Amsterdam: John Benjamins.

Jarvis, S. 2013. “Defining and Measuring Lexical Diversity.” In *Vocabulary Knowledge: Human Ratings and Automated Measures*, edited by Scott Jarvis, and Michael Daller, 13—44. Amsterdam: John Benjamins.

Jeanjean, T., C. Lesage, and H. Stolowy. 2010. “Why Do You Speak English (In Your Annual Report)?” *The International Journal of Accounting* 45 (2): 200—23.

Jin, T., H. Duan, X. Lu, J. Ni, and K. Guo. 2021. “Do Research Articles with More Readable Abstracts Receive Higher Online Attention? Evidence from Science.” *Scientometrics* 126 (10): 8471—90.

Kwok, H. L., S. Laviosa, and K. Liu. 2023. “Lexical Simplification in Learner Translation: A Corpus-Based Approach.” *Research in Corpus Linguistics* 11 (2): 103—24.

Laufer, B., and P. Nation. 1995. “Vocabulary Size and Use: Lexical Richness in L2 Written Production.” *Applied Linguistics* 16 (3): 307—22.

Laviosa, S. 1998. “Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose.” *Meta* 43 (4): 557—70.

Laviosa, S. 2002. *Corpus-based Translation Studies: Theory, Findings, Applications*. Amsterdam: Rodopi.

Lei, S., and R. Yang. 2020. “Lexical Richness in Research Articles: Corpus-Based Comparative Study Among Advanced Chinese Learners of English, English Native Beginner Students and Experts.” *Journal of English for Academic Purposes* 47: 100894.

Liu, Z. 2008. *Paper to Digital: Documents in the Information Age*. Westport: Libraries Unlimited.

Liu, D. 2010. “Chinese as a Verby Language: On Typological Differences between Verby Languages and Nouny Languages.” *Chinese Teaching in the World* (1): 3—17.

Liu, K., and M. Afzaal. 2021. “Syntactic Complexity in Translated and Non-translated Texts: A Corpus-Based Study of Simplification.” *PLoS One* 16 (6): e0253454.

Liu, K., Z. Liu, and L. Lei. 2022. “Simplification in Translated Chinese: An Entropy-Based Approach.” *Lingua* 275: 103364.

Liu, K., H. Yin, and A. K. Cheung. 2024. “Interactional Metadiscourse in Translated and Non-translated Medical Research Article Abstracts: A Corpus-Assisted Study.” *Perspectives*: 1—21. https://doi.org/10.1080/0907676x.2024.2328756.

Loughran, T., and B. McDonald. 2014. “Measuring Readability in Financial Disclosures.” *The Journal of Finance* 69 (4): 1643—71.

Lu, X. 2012. “The Relationship of Lexical Richness to the Quality of ESL Learners’ Oral Narratives.” *The Modern Language Journal* 96 (2): 190—208.

Lu, X., and R. Hu. 2022. “Sense-aware Lexical Sophistication Indices and Their Relationship to Second Language Writing Quality.” *Behavior Research Methods* 54 (3): 1444—60.

Malmkjær, K. 1997. “Punctuation in Hans Christian Andersen’s Stories and Their Translations into English.” In *Nonverbal Communication and Translation: New Perspectives and Challenges in Literature, Interpretation and the Media*, edited by F. Poyatos, 151—62. Amsterdam: John Benjamins.

Mauranen, A. 2000. “Strange Strings in Translated Language: A Study on Corpora.” In *Intercultural Faultlines. Research Models in Translation Studies 1: Textual and Cognitive Aspects*, edited by M. Olohan, 119—41. Manchester: St. Jerome Publishing.

Mauranen, A. 2007. “Investigating English as a Lingua Franca with a Spoken Corpus.” *Spoken Corpora in Applied Linguistics* 51: 33.

Nord, C. 1997. *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester: St. Jerome Publishing.

Olohan, M. 2004. *Introducing Corpora in Translation Studies*. London: Routledge.

Pastor, G. C., R. Mitkov, N. Afzal, and V. Pekar. 2008. “Translation Universals: Do They Exist? A Corpus-Based NLP Study of Convergence and Simplification.” In *Proceedings of the 8th Conference of the Association for Machine Translation in the Americas: Research Papers*, 75—81.

Pokorn, N. K. 2005. *Challenging the Traditional Axioms: Translation into a Non-mother Tongue*. Amsterdam/Philadelphia: John Benjamins.

Read, J. 2000. *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Rutherford, B. A. 2005. “Genre Analysis of Corporate Annual Report Narratives: A Corpus Linguistics-Based Approach.” *Journal of Business Communication* 42 (4): 349—78.

Securities and Exchange Commission. 1998. A Plain English Handbook: How to Create Clear SEC Disclosure Documents. Also available at: http://www.sec.gov/pdf/handbook.pdf.

Setter, J., C. Wong, and B. Chan. 2010. *Hong Kong English*. Edinburgh: Edinburgh University Press.

Shi, Y., and L. Lei. 2022. “Lexical Richness and Text Length: An Entropy-Based Perspective.” *Journal of Quantitative Linguistics* 29 (1): 62—79.

Toutanova, K., D. Klein, C. D. Manning, and Y. Singer. 2003. “Feature-rich Part-Of-Speech Tagging with a Cyclic Dependency Network.” In *In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 252—9.

Ure, J. 1971. “Lexical Density and Register Differentiation.” *Applications of Linguistics* 23 (7): 443—5.

Vanderauwera, R. 1985. *Dutch Novels Translated into English: The Transformation of a “Minority” Literature*. Amsterdam: Rodopi.

Volansky, V., N. Ordan, and S. Wintner. 2015. “On the Features of Translationese.” *Digital Scholarship in the Humanities* 30 (1): 98—118.

Wang, P. 2014. “A Comparison between Chinese Version and English Version of Company Annual Reports.” *The Journal of Chinese Sociolinguistics* 2: 33—44.

Wang, Z., and K. 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, H., L. Li, and J. Cao. 2012. “Lexical Features in Corporate Annual Reports: A Corpus-Based Study.” *European Journal of Business and Social Sciences* 1 (9): 55—71.

Wang, Z., K. Liu, and R. 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—51.

Wang, Z., A. K. Cheung, and K. Liu. 2024a. “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, Z., M. Liu, and K. Liu. 2024b. “Utilizing Machine Learning Techniques for Classifying Translated and Non-translated Corporate Annual Reports.” *Applied Artificial Intelligence* 38 (1): 2340393.

Wu, K., V. L. Lei, and D. Li. 2024. “Charting the Trajectory of Corpus Translation Studies: Exploring Future Avenues for Advancement.” *Corpus-Based Studies Across Humanities*. https://doi.org/10.1515/csh-2024-0001.

Xiao, R. 2010. “How Different Is Translated Chinese from Native Chinese? A Corpus-Based Study of Translation Universals.” *International Journal of Corpus Linguistics* 15 (1): 5—35.

Xiao, R., and G. Dai. 2014. “Lexical and Grammatical Properties of Translational Chinese: Translation Universal Hypotheses Reevaluated from the Chinese Perspective.” *Corpus Linguistics and Linguistic Theory* 10 (1): 11—55.

Xiao, R., and M. Yue. 2009. “Using Corpora in Translation Studies: The State of the Art.” In *Contemporary Corpus Linguistics*, edited by P. Baker, 237—62. London: Continuum.

Yang, R., Y. Wang, and H. Nie. 2013. “The Political Promotion for Quasi-Government Officers: Evidence from Central State-Owned Enterprises in China.” *Management World* 3: 23—33.

Yu, G. 2010. “Lexical Diversity in Writing and Speaking Task Performances.” *Applied Linguistics* 31 (2): 236—59.

Zanettin, F. 2014. *Translation-driven Corpora: Corpus Resources for Descriptive and Applied Translation Studies*. London: Routledge.

Zareva, A., P. Schwanenflugel, and Y. Nikolova. 2005. “Relationship between Lexical Competence and Language Proficiency: Variable Sensitivity.” *Studies in Second Language Acquisition* 27 (4): 567—95.

Zhu, Y. 2005. *Written Communication across Cultures: A Sociocognitive Perspective on Business Genres*. Amsterdam: John Benjamins.

Zhuang, Q. Y., X. T. Li, and X. Q. Li. 2024. “A Corpus-Assisted Comparative Study of Chinese and Western CEO Statements in Annual Reports: Discourse-Historical Approach.” *Corpus-Based Studies Across Humanities* 1 (1): 217—43.