Style in speech and narration of two English translations of *Hongloumeng*

A corpus-based multidimensional study

Isabelle Chou and Kanglong Liu
University of Electronic Science and Technology of China | Hong Kong Polytechnic University

This study examines the style of two English translations of *Hongloumeng*, by David Hawkes, and Xianyi Yang and Gladys Yang. It makes use of multidimensional analysis to identify how the two translations differ in their sub-registers (narration and fictional speech). The results reveal that the Yangs' translation of narration is relatively more narrative and context-independent, whereas Hawkes' is more active and context-bound. Similarly, Hawkes' translation of fictional speech is more conversational and interactional and tends more towards the orality scale with a strong emphasis on the involvement of fictional characters. In contrast, the Yangs' translation of fictional speech tends to be more informational and explicit. These stylistic differences reflect the translators' conscious and/or unconscious choices, which are attributable to their language backgrounds, translation strategies, and cultural stances. By taking sub-register variation and the functions of linguistic features into consideration, the article outlines a new approach to investigating translation style.

Keywords: translator style, multidimensional analysis, *Hongloumeng*, narration, fictional speech

1. Introduction

Acclaimed as one of the four classical masterpieces in China, *Hongloumeng* (Cao and Gao 2005; hereafter HLM) has enjoyed enduring fame for its vivid depiction of almost every facet of eighteenth-century China. The first eighty chapters of the novel were authored by Xueqin Cao (1715–1763) while the remaining forty chapters were authored by E. Gao (1738–1885) following Cao's death. The novel has been translated into more than twenty languages across the centuries (Shengyu

https://doi.org/10.1075/target.22020.cho | Published online: 11 December 2023
Target 36:1 (2024), pp. 76–111. ISSN 0924-1884 | E-ISSN 1569-9986

and Minford 2017). Among dozens of its English translations, two have been widely read and studied, namely, *A Dream of Red Mansions* (Cao 1978a, 1978b, 1980a) by Xianyi Yang (also spelled as Hsien-Yi Yang) and Gladys Yang (YT hereafter), and *The Story of the Stone*, translated by David Hawkes (Cao 1973, 1977, 1980b) and John Minford (Cao 1982, 1986) (HT hereafter). Since their publication, many comparative and contrastive studies have been conducted on various linguistic aspects of these two translations, including metaphor translation (Met and Li 1997; Lian 2014), colour terms (Ke 1995), hedges and boosters (Liu, Kwok, and Moratto 2022), and even idioms (Su 2021). These lines of inquiry have yielded some interesting insights into the translators' tactics and strategies for handling translation challenges in the novel.

With the rise of Corpus-Based Translation Studies (Baker 1993, 1995), researchers have made use of corpus methods to examine the style of the translations of HLM (Liu 2008; Li, Zhang, and Liu 2011; Liu and Afzaal 2021). The focus of such research is generally on the distinct manner in which a writer presents the creation, including the choice of words, sentence structure, use of imagery, and rhythm (Gillespie 2008, 9). Similarly, research focusing on translator style attempts to capture a specific translator's characteristic use of language and linguistic habits, as opposed to that of other translators (Baker 2000, 245). As stated by Baker (2000, 244), "it is as impossible to produce a stretch of language in a totally impersonal way as it is to handle an object without leaving one's fingerprints on it." Based on this assertion, there has been a plethora of insightful corpus-based investigations into the translation style of the HLM translations investigated in this study.

However, like other research on translation style in fiction, previous studies of the translations of HLM have tended to view the novel as a single register without considering its internal variation. Fiction, which has traditionally been treated as one single register, is notoriously "one of the most complicated registers" (Biber and Conrad 2009, 132). However, in recent years there has been growing interest in treating fiction as a hybrid register consisting of fictional speech as well as narration: "Fictional speech is a crucial component of novels for the representation of spoken interaction between characters. Narration on the other hand presents a different situational context for the telling of the story" (Egbert and Mahlberg 2020, 74). Research has also shown that the proportion between fictional speech and narration can be an important factor in affecting the language profiling and syntactic complexity of fiction texts (Liu and Afzaal 2021). Egbert and Mahlberg (2020) find that fictional speech and narration are two contrasting registers, with the former positioned between conversation and spoken interviews, and the latter taking a middle position somewhere between general fiction and biographies. In the field of translation style research, little scholarly attention has been paid to this

internal register variation to show how fictional speech and narration are handled differently by translators. As far as HLM is concerned, Wong (2002) adopts a qualitative approach to investigate the register-related problems across translations of HLM (e.g., religious, scientific registers) in different languages (English, French, German, Italian); and Liu (2008) explores the fictional dialogues of HLM. However, so far, no research has investigated the translation style of HLM by examining fictional speech and narration separately.

In order to address this gap, we adopt corpus-based methods to systematically compare how the two groups of translators, the Yangs and Hawkes, handled the two sub-registers of HLM in their translations. It is believed that such an approach can yield more insights into the style of the two translations than would an examination of the novel as one single register.

2. Studies of translation style in English translations of *Hongloumeng*

2.1 Style in translation

'Style' is an umbrella term that can refer to language style or author's style in literary studies, and many linguists have offered useful definitions and approaches to its analysis. According to Crystal (1999, 323), style is "any situationally distinctive use of language, and of the choices made by individuals and social groups in their use of language." Wales (2001, 371) defines style as "the perceived distinctive manner of expression in writing or speaking." Leech and Short (1981) list four main categories for the analysis of style in literary works: lexical, grammatical, figures of speech, and cohesion and context. They regard literary stylistics as the study of language and artistic function (13).

The concept of style has always been central to Translation Studies. The view that translators need to make proper linguistic choices to reproduce the source text style was dominant in the early phases of Translation Studies (Nida and Taber 1969; Vinay and Darbelnet [1958] 1995). It was believed that apart from retaining the linguistic style of the source text, the rhetorical elements of the source text should also be transferred to the translation so that target readers can experience the same effect when reading the translation as the source readers reading the source text (Newmark 1988). However, this approach has tended to subscribe to the stereotypical notion of conceiving translation as a derivative and secondary activity by reinforcing the dominant position of the source text. In challenging this, Baker's (2000) proposal to study translation style has revolutionised this research field, acknowledging translation as a unique and creative

activity rather than a process of mechanically reproducing the meaning and style of the source text.

In this line of enquiry, researchers mainly follow two different approaches, focusing on translator style or translation style. The first approach is based on comparative studies of the oeuvre of different translators to identify the patterned clustering of linguistic features attributable to specific translators (e.g., Olohan 2004; Bosseaux 2007; Saldanha 2011; Huang and Chu 2014). The second approach explores translations of the same source text by different translators and compares how one translator differs from another in terms of a range of linguistic features (e.g., Winters 2009; Mastropierro 2018). Some scholars, such as Baker (2000) and Saldanha (2011), favour the first approach because the second approach might be practically unfeasible for larger literary works where there are fewer translations of the same original available for such research. However, the translator style approach can be methodologically complicated as there are a plethora of confounding variables that contribute to stylistic differences, including diverse styles of the source text authors, divergences in the stylistic features of the original texts, and different influences of the source languages (Mastropierro 2018, 242). The lack of availability of multiple translations is more problematic for language pairs other than Chinese and English, where the multiple translations and retranslations that exist provide fertile ground for translation style research. Notwithstanding terminological inconsistencies, an investigation of style in the translations of HLM is feasible because the availability of two full-length translations makes it possible to isolate and analyse each translator's style by comparing how the two translation versions differ, while controlling for the various confounding variables.

2.2 Previous studies on style in English translations of *Hongloumeng*

As one of the four classic masterpieces in Chinese literature, HLM and its English translations have been a popular research subject in Translation Studies. Among the various research topics, translator style has attracted much attention from translation scholars. There are two main approaches to exploring style in HLM translations: Some scholars have focused on a particular translator of HLM (Liu and Shi 2018; Li and Lu 2022), whereas others have adopted comparative studies to examine multiple translations of HLM (Li, Zhang, and Liu 2011; Liu, Liu, and Zhu 2011; Liu and Afzaal 2021). In the following, we summarise the main studies reflecting the two approaches and identify potential limitations to provide the necessary context for our own study.

Liu (2008) approaches the topic of the translator's style in relation to three HLM translations (Hawkes, Joly,[1] and the Yangs) by studying how titles and honorifics are translated into English. He finds that different translators adopt different translation strategies, tending more towards domestication (Hawkes) or foreignisation (the Yangs and Joly). Liu and Afzaal (2021) investigate the use of lexical bundles in HLM and find that Hawkes tends to use a higher number and a wider variety of lexical bundles compared to the Yangs. They further argue that the difference can be accounted for by the translators' varying language backgrounds, the target readership, and the socio-political milieu in which the translations were produced. Other research on the translation style of HLM has focused on nominalisation (Hou 2013), vocabulary richness (Fang and Liu 2015), and metaphorical idioms (Su 2021). This line of research thus tends to adopt isolated linguistic features instead of carrying out a holistic examination of the translations of HLM.

In addition to the narrow scope of translation style, research on translations of HLM seems to yield conflicting findings. For example, while several researchers have found that the two translations by Hawkes and the Yangs are not significantly different from each other in the richness of their vocabulary (Fang and Liu 2015), degree of readability (Liu 2014), and lexical density and average word length (Liu, Liu, and Zhu 2011), some studies have found that Hawkes' translation has a higher token count but lower type count than that of the Yangs (Li, Zhang, and Liu 2011). After carefully reviewing these studies, we believe that these conflicting findings can be attributed to the different samples that the various studies are based on. For example, some studies are based on all 120 chapters of HLM (e.g., Fang and Liu 2015), while some extracted only a portion of the chapters (e.g., Li, Zhang, and Liu [2011], who examined the first fifteen chapters of HLM).

Our review reveals that there are two major limitations to current research on stylistic variation in translations of HLM. To begin with, most existing research has relied heavily on isolated linguistic features to distinguish between style in the two translations by Hawkes and the Yangs. This cherry-picking of features may not provide an accurate representation of the major differences between the two translations. In addition, these studies have all sampled HLM as a single register without taking into account its internal variation. By treating fictional speech and narration as a single register, corpus-based research on HLM

1. H. Bencraft Joly, who served as the Vice-Consul of Macao during his translation of *The Dream of the Red Chamber*, aimed to promote the understanding of Chinese literature among Western scholars. His translation, completed in 1892–1893, is accessible on the Gutenberg website, consisting of 56 chapters out of the total 120. It should be noted that Joly passed away before completing the entire translation.

may fail to characterise the major differences between the two translation versions. These, we believe, are two of the major reasons why many studies have concluded that the two translations by Hawkes and the Yangs are not significantly different in terms of style (e.g., Liu, Liu, and Zhu 2011; Fang and Liu 2015).

To address these two limitations, we utilise a multidimensional framework consisting of an array of linguistic features (Biber 1986, 1988) to compare co-occurrences of linguistic features in the two HLM translations by considering sub-register variation in fictional speech and narration. We therefore attempt to answer the following research questions:

1. In what dimensions do the two HLM translations differ from each other in fictional speech and narration?
2. What are the individual linguistic features that characterise the differences in speech and narration in both translations?
3. What are the possible reasons for such differences?

3. Methodology

3.1 Corpus composition

The corpus consists of the two most authoritative translations of HLM, *A Dream of Red Mansions* by Xianyi Yang and Gladys Yang, and *The Story of the Stone* by David Hawkes and John Minford. It should be noted that while the Yangs collaborated on the translation of HLM, Hawkes translated the first eighty chapters and Minford the remaining forty chapters separately. In order to exclude the confounding variable of multiple translators, only the first eighty chapters of each translation were used in the current study.

We first compiled a Python program to automatically extract the dialogues using speech punctuation (quotation marks) in order to separate fictional speech from narration in the two translations. We divided each translation into two distinct sub-corpora, one dedicated to dialogue and the other to narration. Consequently, our compilation resulted in a total of four sub-corpora originating from the two translations. Next, we manually proofread the two sets of texts and removed the incomplete sentences in the narration components, which consisted mainly of reporting clauses such as "he said" and "said Baoyu." This process resulted in 320 files derived from the first 80 chapters of HT and YT. Details are provided in Table 1.

Table 1. Descriptive statistics for narration and fictional speech in the HLM Corpus

Fictional speech				
Version	**Types**	**Tokens**	**TTR**	**STTR**
YT (n=80)	9 801	219 473	4.47	42.14
HT (n=80)	10 730	280 716	3.82	39.28
Fictional narration				
Version	**Types**	**Tokens**	**TTR**	**STTR**
YT (n=80)	11 082	193 903	5.72	43.73
HT (n=80)	14 683	279 361	5.26	43.66

3.2 Multidimensional analysis (MDA)

Pioneered by Biber (1986, 1988), MDA is "a methodological approach that employs multivariate statistical techniques (particularly factor analysis and cluster analysis) to investigate register variation in a language" (Biber 2004, 15). The main motivation of Biber's MDA is that successful register identification should be based on variation in linguistic co-occurrence patterns (termed 'dimensions' in Biber's approach) rather than the distribution of individual linguistic features that tend to be numerous and unstable. As argued by Biber (1988, 20), "comparison of texts with respect to any single dimension gives an incomplete, and sometimes misleading, picture." By analysing spoken and written registers in contemporary British English, Biber (1988) identifies six dimensions of variation by examining how sixty-seven linguistic features co-occur in these registers. The dimensions are interpreted in terms of communicative functions which are assumed to be closely related to the co-occurrence of linguistic features. These six dimensions, each describing an opposition of two types of discourses, may include a number of positive-loading and negative-loading linguistic features. For example, Dimension 1, which represents involved discourse versus informational discourse, contains both linguistic features with positive loadings (e.g., private verbs, *that* deletion, contractions) as well as features with negative loadings (e.g., nouns, word length, prepositional phrases). Dimension 2, which describes narrative versus non-narrative registers, contains only linguistic features with positive loadings. It should be noted that the dimensions are not dichotomies (though the two poles are given opposite labels), but continuous scales along which texts can be quantitatively measured. Each text is allocated a dimension score on the six dimensions: "The dimension score of a text is computed by adding together the factor score of each feature with a positive loading, and then subtracting the factor score of each feature, if any, with a negative loading" (McEnery, Xiao, and Tono

2006, 288). Since Biber's original study, MDA has been adopted to study a wide array of text types including written and spoken registers (Biber et al. 2002) and conversations (Biber 2004), and has been used in comparative studies of press, general prose, academic prose, and fiction (Zhang 2016), and studies of world Englishes (Xiao 2009).

Two types of MDA have been used in the field of Translation Studies. The first type, which can be referred to as self-designed MDA, occurs when researchers choose a range of linguistic features based on the nature of the textual data under investigation, conduct a factor analysis, identify and interpret the dimensions based on the statistical results, and finally calculate and analyse the dimension scores. The second type, which can be referred to as standard MDA, occurs when researchers simply make use of the dimensions identified by previous MDA studies (e.g., Biber's MDA) and calculate the scores based on these established dimensions to examine the textual variations. Self-designed MDA has been used to compare translated and non-translated language (Hu, Xiao, and Hardie 2016), written contact varieties of English (both non-native English and translated English) (Kruger and Van Rooy 2018), and interpreted language (Xu 2021). In standard MDA, researchers mainly make use of certain dimensions of Biber's MDA according to their own research purposes. For example, Kruger and Van Rooy (2016) use Dimensions 1, 3, 5, and 6 to study translated and non-native indigenised varieties of English from the perspective of constrained language. Calzada-Pérez and Sánchez Ramos (2021) use Dimensions 1, 2, and 6 to examine translated and non-translated parliamentary discourse. According to Biber (1988, 200), stylistic comparison is one of the typical applications of MDA in language research. For instance, Biber and Finegan (1988) apply standard MDA to investigate stylistic shifts in fiction, essays, and letters across three centuries. As aptly pointed out by Biber (1988, 203), one major advantage of standard MDA is that by "considering particular authors, and particular works, relative to all six dimensions, we achieve a macroscopic analysis of an author's stylistic distinctiveness." In other words, all six dimensions of Biber's MDA are relevant to the stylistic analysis of English texts. In this study, we attempt to provide a macroscopic examination of translators' stylistic traits by evaluating their specific works in relation to the established dimensions, and therefore regard the use of Biber's standard MDA as appropriate for our study.

To the best of our knowledge, no research to date has employed MDA to investigate the intratextual variations of translated fiction. As a framework that has been proven to work effectively to uncover the uniqueness of various registers and literary styles (Ek and Wirén 2019; Egbert and Mahlberg 2020), MDA can serve as a robust approach for examining the unique style in English translations of HLM.

3.3 Data collection and processing

We used the Multidimensional Analysis Tagger (MAT) to extract the linguistic features in HT and YT respectively for speech and narration. MAT was developed by Nini (2019) to automatically retrieve all sixty-seven linguistic features used in Biber's MDA (1988). Following Biber's approach, the scores for each dimension are calculated by summing all the feature loadings on individual dimensions. The tool has reported high accuracy and efficiency/effectiveness in the classification of different English registers (Nini 2019). As with most corpus software tools, the accuracy needs to be tested and verified with the actual texts. For this reason, we selected twenty texts (ten texts of fictional dialogue and ten texts of narration) for manual checking. The twenty tagged files generated by MAT were checked for mistags to ensure there was no serious skewing of the tagging results. The task was done by two independent raters (interrater reliability: 93%) and a third rater resolved the differences in a three-way discussion between all three raters. It was found that most of the mistags occurred in the categories of attributive adjectives, conjuncts, and other adverbial subordinators. After manual checking was done, the dimension scores for all twenty texts were calculated, using both the manually verified and the unverified datasets. We found that significant differences exist between the verified and unverified samples for Dimension 5, which is not surprising, considering that this dimension contains only six linguistic features (i.e., conjuncts, agentless passives, past participial clauses, *by*-passives, past participial WHIZ deletions, and other adverbial subordinators), and two of these features were incorrectly tagged by the software to a certain extent. On the other hand, Dimension 1 contains twenty-eight linguistic features and the mistags of certain features did not affect the overall dimension scores. Besides, as is noted by Biber (1988), Dimensions 1, 3, and 5 are most relevant for distinguishing 'oral' and 'literate' registers. In comparison to Dimension 1 and 3, Dimension 5 is relatively less robust in this discrimination task (R^2 value is the lowest among the three). In view of the reasons stated above, our analysis thus leaves out Biber's Dimension 5 and focuses on the remaining five dimensions, which we believe are most pertinent for stylistic comparisons of the two translations of HLM.

As the dimension scores and feature scores are not normally distributed, we used a non-parametric test, the Mann-Whitney *U* test, to examine if YT and HT differ in the two sub-registers across the five dimensions. If significant differences were identified in a particular dimension, the features in that dimension were further examined. In the discussion in Section 4, typical translation examples from YT and HT concerning particular linguistic features are presented for further qualitative analysis.

4. Findings

MDA aims to capture subtle and gradual relations between registers. Thus, our first research question can be answered by examining whether significant differences exist in each dimension by comparing the dimension scores of the two HLM translations. We calculated the dimension scores by treating each chapter of each sub-register as a data point; thus, the analysis is based on 320 data points. The means and standard deviations of the dimension scores of YT and HT across the two different sub-registers (narration and fictional speech) are given in Tables 2 and 3. Appendix 1 and 2 provide the complete list of linguistic features for the five dimensions used in this study, and the full results of the MDA analyses for translated fictional speech and narration, respectively; and the statistical comparisons between YT and HT for the two sub-registers.

Table 2. Dimension scores in narration in YT and HT

	YT (*n*=80)				HT (*n*=80)			
Dimensions	**Min.**	**Max.**	**Mean**	***SD***	**Min.**	**Max.**	**Mean**	***SD***
Dimension 1 *Involved vs. informational production*	−19.63	−5.96	−12.62	3.093	−20.3	−5.8	−12.48	2.850
Dimension 2 *Narrative vs. non-narrative*	0.94	11.57	6.49	2.173	0.95	10.37	5.7	1.941
Dimension 3 *Explicit vs. situation-dependent reference*	−2.13	4.85	1.65	1.411	−2.86	4.86	0.7	1.470
Dimension 4 *Overt vs. covert persuasion*	−4.37	3.73	−0.06	1.673	−3.89	3.33	0.01	1.536
Dimension 6 *On-line information elaboration*	−2.9	1.69	−0.58	0.812	−2.02	1	−0.47	0.709

Table 3. Dimension scores in fictional speech in YT and HT

	YT (n=80)				HT (n=80)			
Dimensions	**Min.**	**Max.**	**Mean**	**SD**	**Min.**	**Max.**	**Mean**	**SD**
Dimension 1 *Involved vs. informational production*	−2.77	22.93	13.76	4.891	−0.16	26.02	17.27	5.268
Dimension 2 *Narrative vs. non-narrative*	−2.03	4.08	0.42	1.156	−2.17	4.6	0.35	1.312
Dimension 3 *Explicit vs. situation-dependent reference*	−4.65	3.73	−1.28	1.683	−4.66	2.5	−1.97	1.429
Dimension 4 *Overt vs. covert persuasion*	−0.14	11.64	7.60	2.446	0.96	11.82	8.25	2.342
Dimension 6 *On-line information elaboration*	−1.72	1.52	−0.13	0.664	−1.48	2.31	0.52	0.707

A Mann-Whitney U test shows that there is a significant difference between YT and HT in narration in Dimension 2 (U=2384, p=.005) and Dimension 3 (U=2035, p<.001). In fictional speech, significant differences between YT and HT are found for Dimension 1 (U=1810, p<.001), Dimension 3 (U=2393, p=.006), and Dimension 6 (U=1575, p<.001).

As the functions of each dimension are assessed through the co-occurrence of an array of linguistic features, we will discuss the stylistic differences between YT and HT by referring to the features related to dimensions that demonstrate a statistically significant difference in the dimension score across the two sub-registers. Thus, only Dimensions 1, 2, 3, and 6 are discussed with respect to specific findings (see Table 4 to 7). Given that more dimensions and linguistic features are found to be significantly different between YT and HT in fictional speech than in narration, we focus more on fictional speech in our qualitative analysis.

4.1 Stylistic difference in narration

As shown in Table 2, in terms of narration, YT and HT are significantly different in Dimensions 2 and 3. In Biber's MDA framework, Dimension 2 distinguishes narrative discourse from other types of discourse. High values for the mean rank and sum of ranks for features with positive loadings reflect discourse that is more narrative in nature, whereas high scores for features with negative loadings represent

other types of discourse. Dimension 3 demonstrates the writer's (in this case, the translator's) degree of assumption that the reader is aware of the immediate context of the text: from explicit and context-independent referencing on the positive end of the scale to situation-dependent referencing on the negative end (Biber 1988, 110). For Dimension 3, high values for the mean rank and sum of ranks for features with positive loadings characterise texts with explicit and context-independent features, while situation-dependent reference is reflected in high scores for features with negative loadings.

Table 4. Linguistic features with statistically different frequencies on Dimensions 2 and 3 for narration in YT and HT

	YT (n=80)		HT (n=80)		
Linguistic features	**Mean rank**	**Sum of ranks**	**Mean rank**	**Sum of ranks**	**Mann-Whitney pairwise comparisons**
Dimension 2					
Features with positive loadings					
Present participial clauses	90.68	7254.50	70.32	5625.50	$p<.005$
Features with negative loadings					
Synthetic negation	93.41	7472.50	67.59	5407.50	$p<.001$
Dimension 3					
Features with positive loadings					
Nominalisations	48.90	3912.00	112.10	8968.00	$p<.001$
Pied-piping relative clauses	57.13	4570.00	103.88	8310.00	$p<.001$
Phrasal coordination	96.41	7712.50	64.59	5167.50	$p<.001$
Features with negative loadings					
Place adverbials	64.77	5181.50	96.23	7698.50	$p<.001$
Time adverbials	66.04	5283.00	94.96	7597.00	$p<.001$

As shown in Table 4, two features occur at significantly different frequencies in narration in YT and HT in Dimension 2, one with a positive loading (present participial clauses) and one with a negative loading (synthetic negation). According to the principles of MDA, present participial clauses and synthetic negation are two conflicting features on Dimension 2. This means that a higher score on

this dimension should be reflected by a higher score for the features with positive loadings and a lower score for the features with negative loadings. However, the results in Table 4 demonstrate that YT is characterised by a higher score for both features. This result can be related to the overall distribution of all ten features that load onto Dimension 2. In order to better characterise the differences between HT and YT in their presentation of narration in this dimension, Figure 1 presents the Dimension 2 scores of HT and YT alongside those for five other fictional registers identified by Biber (1988), including romance, mystery, adventure, general fiction, and science fiction. It is evident that, although HT and YT differ significantly in Dimension 2, they still fall within the range of fictional registers. To a large extent, they still demonstrate clear narrative styles that are very close to other types of fiction.

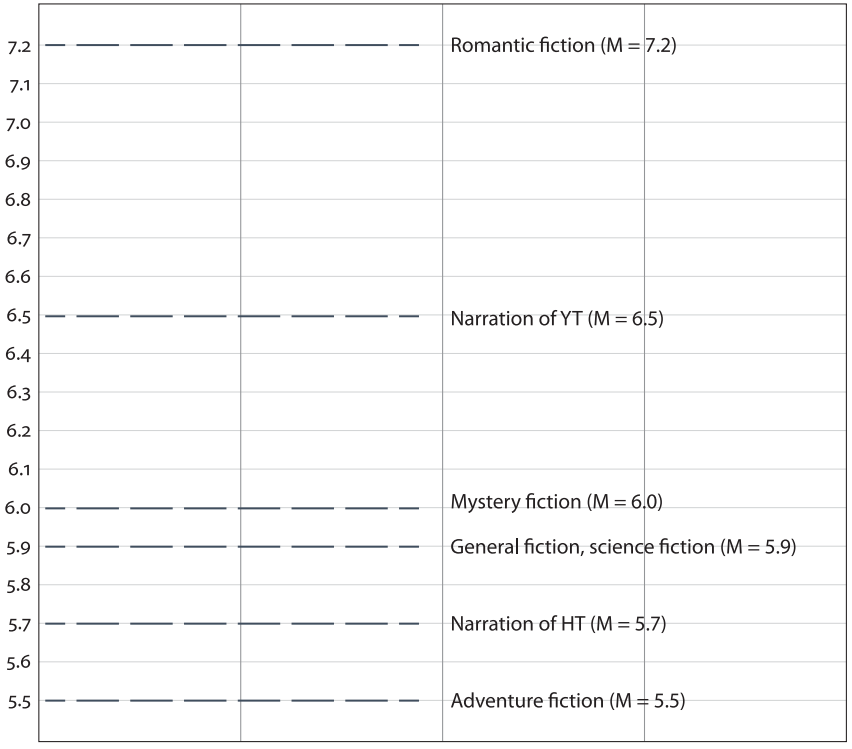

Figure 1. Distribution of narration in HT and YT along Dimension 2 compared to fictional registers from Biber (1988); M = mean dimension score

As for Dimension 3, the results show that YT is more explicit and context-independent in its referencing (i.e., the text is more self-explanatory and self-contained), whereas HT tends to be less explicit and context-dependent (i.e.,

dependent on the situation outside of the text for understanding of its meanings). This difference is also in line with the results for fictional speech translation (see Section 4.2.2). YT has a positive score of 1.67, which is slightly higher than the score of 1.47 for HT for Dimension 3 (see Table 3). In examining the specific features, YT has one positive-loading feature with a higher mean rank (phrasal coordination), and two negative-loading features with much lower mean ranks than HT (place and time adverbials) (see Appendix 2). Together they constitute a more explicit style for YT and a more context-dependent style for HT in fictional narration. Examples (1) to (3) demonstrate how the distinctive styles of the two translations are produced.

(1) ST: 彼时贾珍带着贾蓉来到诸长辈前让坐轿上马...
(Cao and Gao 2005, 193)
Bishi Jia Zhen daizhe Jia Rong laidao zhu zhangbei qian rang zuojiao shangma...
'At that time, Jia Zhen took Jia Rong to the elders and asked them to sit on the sedan chairs or mount the horses ...'
YT: Now Chia Chen and Chia Jung urged their elders to mount their chairs or horses. (Cao 1978a, 201)
HT: Cousin Zhen went round with Jia Rong to the senior men among the mourners and invited them to proceed from there onwards by the transport provided. (Cao 1973, 290)

In Example (1), a nominal phrasal coordination is used as the subject in YT whereas the singular nominal phrase "Cousin Zhen" functions as subject followed by an adverbial of accompaniment ("with Jia Rong"). In YT, another nominal coordination is used as the object in the infinitive ("their chairs or horses"), while HT uses an adverbial of manner ("by the transport provided"). The use of phrasal coordination contributes to a context-independent writing style (Biber 1988).

In contrast, HT is found to have more place and time adverbials than YT, which together create a more context-dependent style. Examples (2) and (3) show the effect of time and place adverbials (underlined) in HT to establish a contextual connection. According to Bosseaux (2007, 179), the use of temporal references can arouse an effect through its "deictic anchorage." HT is more deictically anchored than YT as it more explicitly characterises the spatio-temporal situation of the events. The space and time adverbs serve to anchor the characters and guide the readers in the fictional world. The use of deictic devices enables the narrator(s) to strengthen the "position within the situation they are talking about and also emphasise that the actions are taking place during the unfolding of the [narrator(s)'] utterances" (Bosseaux 2007, 180).

(2) ST: 贾母等让贾妃归坐，又逐次一一见过，又不免哭泣一番。
(Cao and Gao 2005, 240)
Jiamu deng rang Jiafei guizuo, you zhuci yiyi jianguo, you bumian kuqi yifan.
'Grandmother Jia and others asked Jia Concubine to have a seat, (Jia Concubine) then met people one by one, and couldn't help but weep once again.'

YT: … and the Lady Dowager asked her to take a seat, after which she exchanged courtesies with each in turn and more tears were shed.
(Cao 1978a, 254)

HT: Grandmother Jia made her sit down while the members of the family came forward one at a time to greet her and say a few words. This was an occasion for further tears. (Cao 1973, 361)

(3) ST: 因看房内瑶琴、宝鼎、古画、新诗，无所不有，更喜窗下亦有唾绒，奁间时渍粉污。 (Cao and Gao 2005, 80)
Yin kan fangnei yaoqin, baoding, guhua, xinshi, wusuobuyou, geng xi chuangxia yiyou tuorong, lian jian shi zifen wu.
'Looking at the room, (there were) stringed musical instruments, precious tripods, ancient paintings, new poems, and everything was there. (He was) also delighted by some rouge by the window and spilt powder in the women's toilet.'

YT: He saw jasper lutes, rare bronze tripods, ancient paintings, new volumes of verse nothing was lacking. But what delighted him most was the rouge by the window and the spilt powder left from a lady's toilet.
(Cao 1978a, 78)

HT: Looking around the room he noticed various musical instruments, antique bronzes, paintings by old masters, poems by new poets, and other hallmarks of gracious living. He was particularly delighted to observe some rouge-stained pieces of cotton-wool lying on the window-sill – evidently the aftermath of some fairy-woman's toilet.
(Cao 1973, 138)

4.2 Stylistic difference in fictional speech

Regarding fictional speech, HT and YT are significantly different in three dimensions: involved vs. informational production (Dimension 1), explicit vs. situation-dependent reference (Dimension 3), and on-line information elaboration (Dimension 6) (see Table 3).

4.2.1 *Dimension 1: Involved vs. informational production*

In the MDA framework, Dimension 1 is likely to signal clear differences between the two sub-registers, with fictional speech being much more involved and fictional narration being much more informational. As Dimension 1 consists of the largest number of features, it is claimed to be the most robust dimension and has often been the focus of research on register variation (Egbert and Mahlberg 2020). In this dimension, a higher score is observed for HT than for YT, indicating that fictional speech in HT contains more "affective, interactional, and generalised content" (Biber 1988, 107) than YT. In this regard, HT demonstrates a style characterised by a higher degree of personal involvement than YT. The eighteen features demonstrating significant differences in frequency between the two translations in Dimension 1 are listed in Table 5.

According to Biber (1988, 104), "the interpretation of the factors is based on the theoretical assumption that these co-occurrence patterns indicate an underlying communicative function shared by the features." Among these features, some have positive loadings (i.e., the higher their ranks, the more the text is inclined to be of an involved style). Some have negative loadings, which means the higher their frequencies, the more the text tends to have a more informational style. Of the thirty-four linguistic features loading on Dimension 1, eighteen features demonstrate significant differences in frequency between HT and YT. For example, discourse particles (e.g., *well, anyway*), which make no contribution to the propositional content, are commonly considered as indexical devices for spoken discourse (Diewald 2006). Table 5 shows that HT has nearly twice the mean rank score of discourse particles than YT. Together with other features, they reflect a more involved style in the fictional speech of HT and a more informational style in YT. The involved subset is comprised of features which are "verbal, interactional, affective, fragmented, reduced in form, and generalized in content" (Biber 1988, 105). As Table 5 shows, a number of features reflect a higher degree of interpersonal involvement in HT compared to YT. For example, amplifiers can function to signal solidarity with the listener in addition to marking certainty or conviction towards the proposition (258), and "[t]hat clauses, WH-clauses, and adverbial subordinators co-occur frequently with interpersonal and reduced-content features such as first and second person pronouns, questions, contractions, hedges, and emphatics" (230). These features are also a hallmark of spoken registers. Biber (106) further points out that "*BE* as main verb is typically used to modify a noun with a predicative expression, instead of integrating the information into the noun phrase itself," and thus the use of this feature can result in a low informational density. The use of private verbs (e.g., *think, feel*) create a more involved 'feel' by expressing private attitudes, thoughts, and emotions. These fea-

Table 5. Linguistic features with statistically different frequencies on Dimensions 1 for fictional speech in YT and HT

Dimension 1					
	YT (n=80)		HT (n=80)		
Linguistic features	Mean rank	Sum of ranks	Mean rank	Sum of ranks	Mann-Whitney pairwise comparisons
Features with positive loadings					
Amplifiers	47.74	3819.50	113.26	9060.50	p<.001
Causative adverbial subordinators	64.70	5176.00	96.30	7704.00	p<.001
Discourse particles	65.45	5236.00	95.55	7644.00	p<.001
Subordinator *that* deletion	54.58	4366.50	106.42	8513.50	p<.001
WH-clauses	66.93	5354.50	94.07	7525.50	p<.001
Pronoun *it*	61.05	4884.00	99.95	7996.00	p<.001
BE as main verb	69.09	5527.00	91.91	7353.00	p<.002
Private verbs	58.50	4680.00	102.50	8200.00	p<.001
Pro-verb DO	66.00	5280.00	95.00	7600.00	p<.001
Attributive adjectives	98.80	7904.00	62.20	4976.00	p<.001
Demonstrative pronouns	92.05	7364.00	68.95	5516.00	p<.001
Indefinite pronouns	87.78	7022.50	73.22	5857.50	p<.047
Mean word length	90.50	7240.00	70.50	5640.00	p<.006
Direct WH-questions	94.32	7545.50	66.68	5334.50	p<.001
Contractions	100.61	8049.00	60.39	4831.00	p<.001
Features with negative loadings					
Total other nouns	103.41	8273.00	57.59	4607.00	p<.001
Total prepositional phrases	72.29	5783.50	88.71	7096.50	p<.025
Type token ratio	100.49	8039.50	60.51	4840.50	p<.001

tures all have a heavy positive weighting on Dimension 1 and are found to be more frequently used in HT than in YT. In Examples (4) and (5), the use of private verbs "think" and "feel" in HT helps create a more interactional relation in the dialogues. It is evident that the use of private verbs can make the dialogue feel more

conversational, lively, and interactive. HT has a more involved style for translating Chinese dialogues into English than YT. This corroborates the findings by Liu, Cheung, and Moratto (2022) that the dialogue translation of HT has a more conversational tone compared to YT.

(4) ST: "若问此物，倒有一面之缘。" (Cao and Gao 2005, 10)
"Ruo wen ci wu, daoyou yimianzhiyuan."
'"If you ask what it is, you are actually destined to have a look at it."'
YT: "If you want to know, you are destined in your life to meet with it." (Cao 1978a, 9)
HT: "Oh, as for that," "I think it is on the cards for you to have a look at him." (Cao 1973, 54)

(5) ST: "我的五脏都碎了，你还只是哭。" (Cao and Gao 2005, 409)
"Wo de wuzang dou suile, ni hai zhishi ku."
'"All my five internal organs were shattered, and you just go on crying."'
YT: "You're breaking my heart with your weeping." (Cao 1978b, 441)
HT: "I don't know why you go on crying. I feel as if all my insides were shattered." (Cao 1977, 96)

Another distinctive linguistic feature with a positive loading on Dimension 1 is DO as pro-verb. A higher rank (HT: 95) for this feature indicates that the text is more interactive, while a lower rank (YT: 66) typically reflects a more informational style. Examples (6) and (7) show that the use of DO gives a more conversational style to the dialogues in HT. In contrast, YT makes greater use of more descriptive and specific (and often more formal) verbs instead of the verb DO. For these reasons, the style of HT tends to be more conversational and straightforward than that of YT.

(6) ST: "明儿也替我写个匾。" (Cao and Gao 2005, 126)
"Ming'er ye ti wo xie ge bian."
'"Please help me write an inscription tomorrow."'
YT: "You must write an inscription for me some time too." (Cao 1978a, 128)
HT: "You must do one for me some time!" (Cao 1973, 197)

(7) ST: "明日闲了，我一个人作出十二首来。" (Cao and Gao 2005, 516)
"Mingri xian le, wo yige ren zuochu shi'er shou lai."
'"When I have time tomorrow, I will write out twelve of them."'
YT: "Tomorrow when I've time, I mean to write on all twelve themes." (Cao 1978b, 564)
HT: "Perhaps tomorrow or the day after, if I've got the time, I'll try to do all twelve of them again on my own." (Cao 1977, 256)

The use of the pronoun *it* is another feature with a positive loading on Dimension 1 which adds to the interactivity of the text. Biber (1988, 226) assumes that the pronoun *it* is used more frequently in spoken language than in informational contexts because it can be used to replace nouns, phrases, and even entire clauses. HT (rank: 99.95) is characterised by the more frequent use of pronoun *it* compared to YT (rank: 61.05), indicating a more involved and oral style in the former translation than the latter, as illustrated in Example (8).

(8) ST: "明儿叫上屋里听见，可又是不好。" (Cao and Gao 2005, 350)
"Ming'er jiaoshang wuli tingjian, ke youshi buhao."
'"If the ones in the upper room hear about it tomorrow, this will not be good."'
YT: "When word of this gets to the Master, there'll be trouble." (Cao 1978a, 375)
HT: "Let's hope Their Ladyships don't find out about it. There'll be trouble if they do." (Cao 1973, 510)

Biber (1988, 104) states that Dimension 1 "is rather an extremely powerful factor representing a very basic dimension of variation among spoken and written texts in English." From the above quantitative and qualitative analysis, it is evident that HT makes use of a more dynamic, interactional, and involved style for translating the fictional dialogues into English, while YT's speech translation is more formal and information-oriented.

4.2.2 *Dimension 3: Explicit vs. situation-dependent reference*

According to Biber (1988, 110), Dimension 3 functions to distinguish between highly explicit, context-independent reference and nonspecific, situation-dependent reference in a text. As indicated in Table 3, both HT and YT have negative scores for this dimension, indicating that both tend to be situation-dependent in translating dialogue, which is not unusual considering the nature of fictional dialogue, which reflects situated interaction. However, HT is more situation-dependent in its reference than YT, as evident in its higher negative score (HT: −1.97 vs. YT: −1.28). This finding is consistent with the finding for narration in which HT is relatively more context-bound than YT (see Section 4.1). Altogether, there are four features that differ significantly in frequency between HT and YT in this dimension (see Table 6).

As shown in Table 6, HT has one feature with a higher negative loading (place adverbials) than YT, which adds to the higher degree of situation-dependent reference in this translation. In contrast, YT has two features with higher positive loadings (phrasal coordination and WH relative clauses on subject position) than HT, which renders the referencing in this translation more context independent.

Table 6. Linguistic features with statistically different frequencies on Dimension 3 for fictional speech in YT and HT

Dimension 3					
	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean rank**	**Sum of ranks**	**Mean rank**	**Sum of ranks**	**Mann-Whitney pairwise comparisons**
Features with positive loadings					
Phrasal coordination	98.96	7916.50	62.04	4963.50	$p<.001$
Pied-piping relative clauses	71.29	5703.50	89.71	7176.50	$p=.007$
WH relative clauses on subject position	88.01	7041.00	72.99	5839.00	$p=.040$
Features with negative loadings					
Place adverbials	69.93	5594.00	91.08	7286.00	$p=.004$

Overall, the combination of these features contributes to a more explicit and elaborated style in YT than in HT. As an example, WH relative clauses are used to "specify the identity of referents within a text in an explicit and elaborated manner, so that the addressee will have no doubt as to the intended referent" (Biber 1988, 110). Examples (9) and (10) demonstrate that YT tends to use WH relative clauses in subject position, whereas HT renders these using a simple sentence. Our findings for Dimension 3 corroborate our findings for Dimension 1 in that YT is more explicit and informational, while HT is more situation-dependent and involved in style.

(9) ST: "若说我不是真心，暂且拿话支吾，日后再图别的，这天地鬼神，日头月亮照着嗓子，从嗓子里头长疔烂了出来，烂化成酱在这里！" (Cao and Gao 2005, 625)
"Ruo shuo wo bushi zhenxin, zanqie na hua zhiwu, rihou zai tu bie de, zhe tian di gui shen, ritou yueliang zhao zhe sangzi, cong sangzi litou zhang ding lan le chulai, lan hua cheng jiang zai zheli!"
'"If I am not sincere and am only saying it in order to get a different deal in the future, may heaven, earth, ghosts, gods, the Sun and moon be my witnesses, I would have a quinsy in my throat that will rot into pulp."'
YT: "If you think I'm not in earnest and this is just empty talk which I'll go back on later, may Heaven, Earth, all the deities and the Sun and Moon who are my witnesses choke me with an ulcer in my throat so that I rot away into a pulp!" (Cao 1978b, 90)

HT: "... I call on heaven and earth and all the gods and the sun and moon to be my witness: if I don't honestly and sincerely mean every word I say, may I be struck with a quinsy this very moment and matter burst out of my mouth!" (Cao 1977, 424)

(10) ST: "...再有几家，或十斤、八斤、三斤、五斤的不等，也少不得要替他点。" (Cao and Gao 2005, 338)
"Zai you jijia, huo shijin, bajin, sanjin, wujin de budeng, ye shaobude yao ti ta dian."
'"... And some families (donate) ten *jin*, eight *jin*, three *jin*, or five *jin*, (we also) have to keep the lights burning for them."'

YT: "Some poor families who can't afford so much may just donate a quarter or half a catty, but we keep a lamp burning for them just the same." (Cao 1978a, 361)

HT: "Oh, and there's some pays for ten pounds a day, some for eight pounds, three pounds, five pounds – all sorts. All of them I keep their seas of light burning for them, back at my house." (Cao 1973, 494)

4.2.3 *Dimension 6: On-line information elaboration*

Dimension 6 distinguishes between discourse that is informational and produced under real-time conditions compared to informational discourse produced under other, less pressurised conditions, which is typically highly integrated (Biber 1988, 114). In this dimension, YT has a lower score than HT, indicating YT is more integrated and denser than HT (see Table 7).

Table 7. Linguistic features with statistically different frequencies on Dimension 6 for fictional speech in YT and HT

Dimension 6					
	YT (n=80)		HT (n=80)		
Linguistic features	**Mean rank**	**Sum of ranks**	**Mean rank**	**Sum of ranks**	**Mann-Whitney pairwise comparisons**
That relative clauses on object position	64.64	5171.50	96.36	7708.50	p<.001
That adjective complements	61.29	4903.00	99.71	7977.00	p<.001
That verb complements	51.46	4116.50	109.54	8763.50	p<.001

As can be seen from Table 7, *that* relative clauses in object position and *that* complements occur more frequently in HT compared to YT. According to Biber

(1988, 158), *that* complements to verbs and adjectives and *that* relatives are used for informational elaboration in such a way that each additional piece of information is tacked on rather than integrated tightly into the text. This is typical of linguistic production taking place under real-time conditions instead of being prepared beforehand. Hawkes seems to deliberately adopt this strategy to imitate real-life dialogues in his translation. Example (11) shows that HT adopts a *that* adjective complement ("You'd be always finding that …"), while YT uses a rather succinct translation ("Comparisons are so invidious").

(11) ST: "依我说，里头也不用归账，这个多了，那个少了，倒多了事。" (Cao and Gao 2005, 769)
"*Yi wo shuo, litou ye buyong guizhang, zhege duole, nage shaole, dao duo liao shi.*"
'"In my opinion, no more account settling is needed. Otherwise more trouble would occur as one may have more and the other may have less."'
YT: "I don't think accounts need be turned in at all. Comparisons are so invidious." (Cao 1980a, 256)
HT: "If you ask me, I don't think there should be any settling of accounts at all. You'd always be finding that this one had too much and that one too little. It would only be a lot of extra trouble." (Cao 1980b, 75)

Thus, the features grouped on this dimension enable a direct encoding of attitude or stance in addition to their use for informational elaboration. Kruger and Van Rooy (2016, 43) contend that the co-occurrence of these features on Dimension 6 reflects the selection of "elaboration structures that are not densely integrated or cognitively complex, but rather constitute a cognitively less demanding 'add-on' strategy to elaboration typical of spoken conversation." In this sense, the findings for Dimensions 1, 3, and 6 are consistent, and together reflect a more natural conversational style in HT than YT.

5. Discussion

As noted by Biber and Egbert (2018), the existence of hybrid registers and intratextual register variation has often been ignored. This has hampered not only the development of register studies, but also stylistic studies such as research on translation style. Fiction has traditionally been treated as an independent register in its own right (Biber 1988; Biber and Conrad 2009) as a result of the traditional notion "that texts are nested within registers, but registers are not nested within texts" (Egbert and Mahlberg 2020, 97). However, fiction has posed great challenges for researchers because the real-world context used to explain the textual

functions of a particular register, such as telephone conversations or academic prose, is of little relevance to fiction (see Biber and Conrad 2009, 132). As noted by Egbert and Mahlberg (2020, 74), a variety of imaginary situational contexts can be developed in fiction and thus it is natural to treat fiction as a register with internal sub-registers. By separating fiction into dialogue and narration, Egbert and Mahlberg (2020) determine that the novels of Charles Dickens differ from other nineteenth-century fiction in various linguistic dimensions in terms of the presentation of dialogue and narration. They further contend that the separation of fiction into the two sub-registers (i.e., fictional dialogue and narration) "does not only contribute to a more systematic understanding of the features that build fictional worlds, but also to further development of approaches in corpus and register studies" (98).

Likewise, research on translation style has traditionally tended to treat fiction as one single register without distinguishing between its internal variations (e.g., Bosseaux 2007; Li, Zhang, and Liu 2011). Such an approach actually runs the risk of producing "an 'average' of different register features" (Egbert and Mahlberg 2020, 76). Our study shows that, by separating fictional speech and narration in HLM and its translations, we can gain more insights into the translators' style than we would by examining the novel as a whole. The use of this new methodology could contribute to stylistic research in Translation Studies as it provides a more nuanced perspective on translators' strategies and idiosyncrasies.

Previous studies on translator style are often limited by the selection of isolated linguistic features. Such indicators, despite their strengths, only reveal one particular aspect of the translator's style. In fact, the investigation of isolated features might lead to contradictory findings. MDA, on the other hand, can overcome such methodological limitations as the dimensions are made up of an array of linguistic features, which help to identify general trends with greater certainty. Following the MDA approach, the current study has found that HT and YT differ in their translation of fictional speech and narration in a number of dimensions. The tendency of HT towards greater orality in fictional speech is reflected in its scores for Dimension 1, and also corroborated by the scores for Dimension 3 (more context-dependent reference) and Dimension 6 (less integrated and more improvised). According to Egbert and Mahlberg (2020), fictional speech in the novels of Charles Dickens for Dimension 1 returns a mean score of 24.5 and a reference corpus of other nineteenth-century fiction, a mean score of 20.1, both being less involved than face-to-face conversations (35.3) but slightly more involved than spontaneous speeches (18.2), as identified by Biber (1988). Based on Dimension 1, it is obvious that Hawkes' translation of fictional speech (17.27) is closer to the style of original English novels than that of the Yangs (13.76). According to Wong (2014, 308–309), "the original dialogues in HLM exhibit a

wide range of styles exploited to the full to achieve artistic effects and subtleties of characterization." From the reader's perspective, Hawkes's translation of the fictional dialogues in HLM may present those subtleties by using a style more familiar to English readers. In comparison, the translation of fictional speech by the Yangs is more informational than conversational, which has been shown elsewhere through the analysis of linguistic features such as sentence length (see Li, Zhang, and Liu 2011). In the current study, we have also demonstrated that MDA can serve as a robust method for quantitatively characterising translator's style in a holistic manner. MDA is clearly a more comprehensive approach than using a limited number of randomly chosen linguistic features.

As argued by Li, Zhang, and Liu (2011, 164), "corpus-assisted translation research can go beyond proving the obvious or the already known as long as meta- or para-texts are available for analysis." We agree that translation style research should not be limited to uncovering the stylistic differences; instead, it should go one step further to explain such differences by relating these to the social, cultural, and ideological backgrounds of the translators. This is also the case with the current study. The two HLM translations exhibit stylistic differences which can be traced back to the translator's language background, the translation purpose and readership awareness, and the socio-political and ideological milieu in which the translations were produced (see Li, Zhang, and Liu [2011]; Liu and Afzaal [2021], and Liu, Cheung, and Moratto [2022] for a detailed discussion). Of all the factors, we believe that the translator's purpose in translating the novel is the key factor. Hawkes translated HLM out of personal interest, whereas the Yangs were largely compelled by the political task of disseminating Chinese culture and literature to the west (see Li, Zhang, and Liu 2011). These different motivations clearly prompted the use of different translation strategies. Hawkes rendered the novel, notably its fictional dialogue, in a style that resembles similar registers in the target language, whereas the Yangs' translation deviates from reader expectations in this regard. Hawkes (1973, 46) states that he would not have lived in vain if he could "convey to the reader even a fraction of the pleasure of this Chinese novel." In so doing, he has adopted a more domesticated approach to his translation of HLM's dialogue and narrative. As mentioned, the Yangs' translation task was to a large extent imposed by the publisher's goal of promoting Chinese culture to the Anglophone world (Li, Zhang, and Liu 2011, 159), which resulted in their approach of drawing the readers of the translation into the expectations of the source text. In the current study, the use of corpus-assisted MDA has enabled us to report on the influence of these factors on the translators' style in a more objective and balanced manner.

6. Conclusion

The aim of this article has been to explore the stylistic differences between two HLM translations using MDA to systematically study two register-internal variations: narration and fictional speech. While we have revealed some important but often overlooked aspects of research on HLM translation style, it should be noted that this study is by no means exhaustive as fiction is highly complex and intricate. That said, we hold that MDA can serve as an effective approach to research on translator style. Our study represents an effort in this direction and future studies with other works are needed to test the validity of the MDA framework by taking the internal register variations of fiction into consideration. Furthermore, we believe that more paratexts (such as prefaces, commentaries, and personal letters) should be collected to help provide a contextualised explanation for corpus-derived findings in research on HLM translation style.

Funding

This work was supported by a startup fund "A Parallel and Comparable Investigation of Translator's Style in *Hongloumeng* Translations" (Ref No: P0000242) awarded to Kanglong Liu by The Hong Kong Polytechnic University.

Acknowledgements

The authors would like to extend their sincere gratitude to the editors and reviewers for their comments and suggestions.

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. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 7 (2): 223–243.

Baker, Mona. 2000. "Towards a Methodology for Investigating the Style of a Literary Translator." *Target* 12 (2): 241–266.

Biber, Douglas, and Edward Finegan. 1988. "Drift in Three English Genres from the 18th to the 20th Centuries: A Multidimensional Approach." In *Corpus Linguistics, Hard and Soft: Proceedings of the Eighth International Conference on English Language Research on Computerized Corpora*, edited by Merja Kytö, Ossi Ihalainen, and Matti Rissanen, 83–101. Amsterdam: Rodopi.

Biber, Douglas, and Jesse Egbert. 2018. *Register Variation Online*. Cambridge: Cambridge University Press.

Biber, Douglas, and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.

Biber, Douglas, Susan Conrad, Randi Reppen, Pat Byrd, and Marie Helt. 2002. "Speaking and Writing in the University: A Multidimensional Comparison." *TESOL Quarterly* 36 (1): 9–48.

Biber, Douglas. 1986. "Spoken and Written Textual Dimensions in English: Resolving the Contradictory Findings." *Language* 62 (2): 384–414.

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

Biber, Douglas. 2004. "Conversation Text Types: A Multi-Dimensional Analysis." In *Le Poids des Mots: Proceedings of the 7ième Journées internationales d'Analyse statistique des Données Textuelles* [The weight of words: Proceedings of the 7th International Conference on the Statistical Analysis of Textual Data], edited by Gérald Purnelle, Cédrick Fairon, and Anne Dister, 15–31. Louvain: Presses Universitaires de Louvain.

Bosseaux, Charlotte. 2007. *How Does It Feel? Point of View in Translation: The Case of Virginia Woolf into French*. Amsterdam: Rodopi.

Calzada-Pérez, María, and María del Mar Sánchez Ramos. 2021. "MDA Analysis of Translated and Non-Translated Parliamentary Discourse." In *Corpus Exploration of Lexis and Discourse in Translation*, edited by Meng Ji and Michael P. Oakes, 26–55. Abingdon: Routledge.

Cao, Xueqin, and E. Gao. 2005. *Hongloumeng* [红楼梦]. Beijing: People's Literature Publishing House.

Cao, Xueqin. 1973. *The Story of the Stone, Vol. 1: The Golden Days* [红楼梦第一卷：枉入红尘]. Translated by David Hawkes. Harmondsworth: Penguin.

Cao, Xueqin. 1977. *The Story of the Stone, Vol. 2: The Crab-Flower Club* [红楼梦第二卷：海棠诗社]. Translated by David Hawkes. Harmondsworth: Penguin.

Cao, Xueqin. 1978a. *A Dream of Red Mansions, Vol. I* [红楼梦第一卷]. Translated by Xianyi Yang and Gladys Yang. Beijing: Foreign Languages Press.

Cao, Xueqin. 1978b. *A Dream of Red Mansions, Vol. II* [红楼梦第二卷]. Translated by Xianyi Yang and Gladys Yang. Beijing: Foreign Languages Press.

Cao, Xueqin. 1980a. *A Dream of Red Mansions, Vol. III* [红楼梦第三卷]. Translated by Xianyi Yang and Gladys Yang. Beijing: Foreign Languages Press.

Cao, Xueqin. 1980b. *The Story of the Stone, Vol. 3: The Warning Voice* [红楼梦第三卷：异兆悲音]. Translated by David Hawkes. Harmondsworth: Penguin.

Cao, Xueqin. 1982. *The Story of the Stone, Vol. 4: The Debt of Tears* [红楼梦第四卷：绛珠还泪]. Translated by John Minford. Harmondsworth: Penguin.

Cao, Xueqin. 1986. *The Story of the Stone, Vol. 5: The Dreamer Wakes* [红楼梦第五卷：万境归空]. Translated by John Minford. Harmondsworth: Penguin.

Crystal, David. 1999. *The Penguin Dictionary of Language*. London: Penguin.

Diewald, Gabriele. 2006. "Discourse Particles and Modal Particles as Grammatical Elements." In *Approaches to Discourse Particles*, edited by Kerstin Fischer, 403–426. Amsterdam: Elsevier.

Egbert, Jesse, and Michaela Mahlberg. 2020. "Fiction – One Register or Two? Speech and Narration in Novels." *Register Studies* 2 (1): 72–101.

Ek, Adam, and Mats Wirén. 2019. "Distinguishing Narration and Speech in Prose Fiction Dialogues." In *Digital Humanities in the Nordic Countries 4th Conference (DHN), Copenhagen, Denmark, March 5–8*, edited by Costanza Navarretta, Manex Agirrezabal, and Bente Maegaard, 124–132. CEUR-WS.org

Fang, Yu, and Haitao Liu. 2015. "Comparison of Vocabulary Richness in Two Translated *Hongloumeng*." *Glottometrics* 31: 54–75.

Gillespie, J. Thomas. 2008. *Historical Fiction for Young Readers (Grades 4–8): An Introduction*. London: Libraries Unlimited.

Hawkes, David. 1973. "Introduction." In *The Story of the Stone: A Chinese Novel by Cao Xueqin in Five Volumes, Vol. 1: The Golden Days* [红楼梦第一卷：枉入红尘]. Translated by David Hawkes, 15–46. Harmondsworth: Penguin.

Hou, Yu. 2013. "A Corpus-Based Study of Nominalization as a Feature of Translator's Style (Based on the English Versions of *Hongloumeng*)." *Meta* 58 (3): 556–573.

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.

Huang, Libo, and Chiyu Chu. 2014. "Translator's Style or Translational Style? A Corpus-Based Study of Style in Translated Chinese Novels." *Asia Pacific Translation and Intercultural Studies* 1 (2): 122–141.

Ke, Wen-Li. 1995. "Translation and Culture: With Special Reference to the Translation of *Hong Lou Meng*." *Babel* 41 (1): 24–35.

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.

Kruger, Haidee, and Bertus van Rooy. 2018. "Register Variation in Written Contact Varieties of English: A Multidimensional Analysis." *English World-Wide* 39 (2): 214–242.

Leech, Geoffrey N., and Mick Short. 1981. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. London: Longman.

Li, Defeng, Chunling Zhang, and Kanglong Liu. 2011. "Translation Style and Ideology: A Corpus-Assisted Analysis of Two English Translations of *Hongloumeng*." *Literary and Linguistic Computing* 26 (2): 153–166.

Li, Ping, and Dongli Lu. 2022. "A Study of Lin Yutang's Edited Translation of *Hongloumeng* and His Explicitation Strategies." In *Dream of the Red Chamber: Literary and Translation Perspectives*, edited by Riccardo Moratto, Kanglong Liu, and Di-kai Chao, 152–163. Abingdon: Routledge.

Lian, Zhangjun. 2014. "A Parallel Corpus-Based Study of Interpersonal Metaphors in *Hong Lou Meng* and Their Translations." *Theory and Practice in Language Studies* 4 (6): 1155–1161.

Liu, Chaohui. 2014. "On the Acceptability of Two English Translations of *Hongloumeng*." [评《红楼梦》两个英译本的可接受性——以美国亚利桑那州立大学学生的抽样调查为例]. *Chinese Translators Journal* 35 (1): 82–87.

Liu, Kanglong, and Muhammad Afzaal. 2021. "Translator's Style through Lexical Bundles: A Corpus-Driven Analysis of Two English Translations of *Hongloumeng*." *Frontiers in Psychology* 12: 633422.

Liu, Kanglong, Ho Ling Kwok, and Riccardo Moratto. 2022. "Hedges and Boosters as Indicators of Translation Style: With Reference to Fictional Dialogues in *Hongloumeng* Translations." In *Dream of the Red Chamber: Literary and Translation Perspectives*, edited by Riccardo Moratto, Kanglong Liu, and Di-kai Chao, 182–199. Abingdon: Routledge.

Liu, Kanglong, Joyce Oiwun Cheung, and Riccardo Moratto. 2022. "Lexical Bundles in the Fictional Dialogues of Two *Hongloumeng* Translations: A Corpus-Assisted Approach." In *Advances in Corpus Applications in Literary and Translation Studies*, edited by Riccardo Moratto and Defeng Li, 229–253. Abingdon: Routledge.

Liu, Zequan, and Gaoyuan Shi. 2018. "On the Narrative Construction of Lin Yutang's Abridged Translation." *Studies on 'A Dream of Red Mansions'* 2: 231–259.

Liu, Zequan, Chaopeng Liu, and Hong Zhu. 2011. "An Exploration of the Translators' Styles in Four English Translations of Hongloumeng: A Corpus-Based Statistical Analysis." [四个英译本的译者风格初探——基于语料库的统计与分析]. *Chinese Translators Journal* 32 (1): 60–64.

Liu, Zequan. 2008. "Translating Tenor: With Reference to the English Versions of Hong Lou Meng." *Meta* 53 (3): 528–548.

Mastropierro, Lorenzo. 2018. "Key Clusters as Indicators of Translator Style." *Target* 30 (2): 240–259.

McEnery, Tony, Richard Xiao, and Yukio Tono. 2006. *Corpus-Based Language Studies: An Advanced Resource Book*. Abingdon: Routledge.

Met, Cheng, and Li Bingbing. 1997. "Cultural Transplantation and Transformation: Metaphor Translation of *A Dream of Red Mansions*." *Perspectives: Studies in Translatology* 5 (2): 227–235.

Newmark, Peter. 1988. *A Textbook of Translation*. New York: Prentice Hall.

Nida, Eugene A., and Charles Taber. 1969. *The Theory and Practice of Translation*. Leiden: Brill.

Nini, Andrea. 2019. "The Multi-Dimensional Analysis Tagger." In *Multi-Dimensional Analysis: Research Methods and Current Issues*, edited by Tony Berber Sardinha and Marcia Veirano Pinto, 67–94. London: Bloomsbury Academic.

Olohan, Meave. 2004. *Introducing Corpora in Translation Studies*. Abingdon: Routledge.

Saldanha, Gabriela. 2011. "Translator Style: Methodological Considerations." *The Translator* 17 (1): 25–50.

Shengyu, Fan, and John Minford. 2017. "The Story of the Stone's Journey to the West." In *The Routledge Handbook of Chinese Translation*, edited by Chris Shei and Zhao-Ming Gao, 374–387. Abingdon: Routledge.

Su, Ke. 2021. "Translation of Metaphorical Idioms: A Case Study of Two English Versions of *Hongloumeng*." *Babel* 67 (3): 332–354.

Vinay, Jean-Paul, and Jean Darbelnet. (1958) 1995. *Comparative Stylistics of French and English: A Methodology for Translation* [orig. *Stylistique comparée du français et de l'anglais, méthode de traduction*]. Translated and edited by Juan C. Sager and Marie-Josée Hamel. Amsterdam: John Benjamins.

Wales, Katie. 2001. *A Dictionary of Stylistics*. 2nd ed. London: Longman.

Winters, Marion. 2009. "Modal Particles Explained: How Modal Particles Creep into Translations and Reveal Translators' Styles." *Target* 21 (1): 74–97.

Wong, Laurence Kwok Pun. 2002. "Translating Register: With Reference to English, French, German, and Italian Versions of the *Hong lou meng*." *Babel* 48 (3): 247–266.

Wong, Laurence Kwok Pun. 2014. *Dreaming across Languages and Cultures: A Study of the Literary Translations of the Hong lou meng*. Cambridge: Cambridge Scholars.

Xiao, Richard. 2009. "Multidimensional Analysis and the Study of World Englishes." *World Englishes* 28 (4): 421–450.

Xu, Cui. 2021. *Identification of L2 Interpretese: A Corpus-Based, Intermodal, and Multidimensional Analysis*. PhD diss. The Hong Kong Polytechnic University.

Zhang, Man. 2016. "A Multidimensional Analysis of Metadiscourse Markers across Written Registers." *Discourse Studies* 18 (2): 204–222.

Appendix 1. MDA results of HLM translated fictional speech

	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean score**	**Mean rank**	**Mean score**	**Mean rank**	**Mann-Whitney pairwise comparisons**
Dimension 1					
Features with positive loadings					
Private verbs	1.79	58.50	2.15	102.50	$p<.001$
Subordinator *that* deletion	0.35	54.58	0.54	106.42	$p<.001$
Contractions	4.46	100.61	3.59	60.39	$p<.001$
Present tense	7.18	76.40	7.31	84.60	$p=.263$
Second person pronouns	4.02	76.59	4.18	84.41	$p=.285$
Pro-verb DO	0.24	66.00	0.33	95.00	$p<.001$
Analytic negation	1.96	86.53	1.85	74.48	$p=.100$
Demonstrative pronouns	0.87	92.05	0.76	68.95	$p<.010$
Emphatics	1.12	83.43	1.07	77.57	$p=.423$
First person pronouns	5.25	79.68	5.26	81.33	$p=.822$
Pronoun *it*	1.53	61.05	1.94	99.95	$p<.001$
BE as main verb	2.35	69.09	2.52	91.91	$p<.010$
Causative adverbial subordinators	0.11	64.70	0.15	96.30	$p<.001$
Discourse particles	0.18	65.45	0.23	95.55	$p<.001$
Indefinite pronouns	0.16	87.78	0.13	73.22	$p<.05$
Hedges	0.01	75.42	0.02	85.58	$p=.089$

Appendix 1. *(continued)*

	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean score**	**Mean rank**	**Mean score**	**Mean rank**	**Mann-Whitney pairwise comparisons**
Amplifiers	0.12	47.74	0.34	113.26	$p<.001$
Sentence relatives	0.03	84.28	0.03	76.72	$p=.285$
Direct WH-questions	0.45	94.32	0.33	66.68	$p<.001$
Possibility modals	0.95	81.72	0.94	79.28	$p=.739$
Phrasal coordination	0.64	98.96	0.47	62.04	$p<.001$
WH-clauses	0.19	66.93	0.25	94.07	$p<.001$
Stranded preposition	0.28	76.61	0.3	84.39	$p=.288$
Total adverbs	5.2	85.83	5.03	75.18	$p=.146$
Conditional adverbial subordinators	0.84	86.28	0.78	74.73	$p=.115$
Features with negative loadings					
Total other nouns	15.66	103.41	13.86	57.59	$p<.001$
Word length	3.94	90.50	3.9	70.50	$p<.010$
Total prepositional phrases	7.68	72.29	8.01	88.71	$p<.050$
Type–token ratio	209.82	100.49	200.31	60.51	$p<.001$
Attributive adjectives	4.89	98.80	4.29	62.20	$p<.001$
Place adverbials	0.32	69.93	0.39	91.08	$p<.010$
Agentless passives	0.69	88.50	0.6	72.50	$p<.050$
Past participial WHIZ deletion relatives	0.05	82.69	0.05	78.31	$p=.542$
Present participial WHIZ deletion relatives	0.14	78.73	0.15	82.28	$p=.627$
Dimension 2					
Features with positive loadings					
Past tense	2.65	77.42	2.74	83.58	$p=.400$
Third person pronouns	3.73	77.06	3.89	83.94	$p=.347$
Perfect aspect	0.55	55.11	0.85	105.89	$p<.001$
Public verbs	0.64	73.70	0.7	87.30	$p=.063$
Synthetic negation	0.38	99.93	0.27	61.08	$p<.001$
Present participial clauses	0.11	94.39	0.08	66.61	$p<.001$

Appendix 1. *(continued)*

	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean score**	**Mean rank**	**Mean score**	**Mean rank**	**Mann-Whitney pairwise comparisons**
Features with negative loadings					
Present tense	7.18	76.40	7.31	84.60	$p=.263$
Attributive adjectives	4.89	98.80	4.29	62.20	$p<.001$
Past participial WHIZ deletion relatives	0.05	82.69	0.05	78.31	$p=.542$
Word length	3.94	90.50	3.9	70.50	$p<.010$
Dimension 3					
Features with positive loadings					
WH relative clauses on object position	0.02	78.61	0.02	82.39	$p=.566$
Pied-piping relative clauses	0.02	71.29	0.03	89.71	$p<.010$
WH relative clauses on subject position	0.11	88.01	0.09	72.99	$p<.050$
Phrasal coordination	0.64	98.96	0.47	62.04	$p<.001$
Nominalisations	0.6	73.81	0.69	87.19	$p=.068$
Features with negative loadings					
Time adverbials	0.86	83.96	0.82	77.04	$p=.345$
Place adverbials	0.32	69.93	0.39	91.08	$p<.010$
Total adverbs	5.2	85.83	5.03	75.18	$p=.146$
Dimension 4					
Features with positive loadings					
Infinitives	2.18	56.59	2.54	104.41	$p<.001$
Predictive modals	1.55	86.56	1.44	74.44	$p=.098$
Suasive verbs	0.44	87.13	0.38	73.88	$p=.070$
Conditional adverbial subordinators	0.84	86.28	0.78	74.73	$p=.115$
Necessity modals	0.52	71.98	0.57	89.03	$p<.050$
Split auxiliaries	0.4	63.52	0.51	97.48	$p<.001$
Possibility modals	0.95	81.72	0.94	79.28	$p=.739$

Appendix 1. *(continued)*

Linguistic features	YT ($n=80$) Mean score	YT ($n=80$) Mean rank	HT ($n=80$) Mean score	HT ($n=80$) Mean rank	Mann-Whitney pairwise comparisons
Dimension 6					
Features with positive loadings					
That verb complements	0.14	51.46	0.3	109.54	$p<.001$
Demonstratives	1.21	76.76	1.26	84.24	$p=.307$
That relative clauses on object position	0.05	64.64	0.08	96.36	$p<.001$
That adjective complements	0.04	61.29	0.07	99.71	$p<.001$
Stranded prepositions	0.28	76.61	0.3	84.39	$p=.288$
Existential *there*	0.29	64.36	0.38	96.64	$p<.001$
Demonstrative pronouns	0.87	92.05	0.76	68.95	$p<.010$
WH relative clauses on object position	0.02	78.61	0.02	82.39	$p=.566$
Features with negative loadings					
Phrasal coordination	0.64	98.96	0.47	62.04	$p<.001$

Appendix 2. MDA results of HLM translated narration

Linguistic features	YT ($n=80$) Mean score	YT ($n=80$) Mean rank	HT ($n=80$) Mean score	HT ($n=80$) Mean rank	Mann-Whitney pairwise comparisons
Dimension 1					
Features with positive loadings					
Private verbs	1.51	84.18	1.43	76.82	$p=.315$
Subordinator *that* deletion	0.19	86.59	0.16	74.41	$p=.096$
Contractions	0.07	81.89	0.05	79.11	$p=.690$
Present tense	0.8	80.03	0.72	80.97	$p=.898$
Second person pronouns	0.09	74.84	0.12	86.16	$p=.115$
Pro-verb DO	0.06	70.72	0.09	90.28	$p<.010$

Appendix 2. *(continued)*

	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean score**	**Mean rank**	**Mean score**	**Mean rank**	**Mann-Whitney pairwise comparisons**
Analytic negation	0.47	93.78	0.37	67.22	$p<.001$
Demonstrative pronouns	0.41	100.73	0.29	60.27	$p<.001$
Emphatics	0.53	84.76	0.49	76.24	$p=.244$
First person pronouns	0.19	75.89	0.21	85.11	$p=.206$
Pronoun *it*	0.68	57.97	0.97	103.03	$p<.001$
BE as main verb	0.93	64.05	1.15	96.95	$p<.001$
Causative adverbial subordinators	0.04	75.98	0.05	85.02	$p=.206$
Discourse particles	0.1	103.88	0.04	57.12	$p<.001$
Indefinite pronouns	0.1	87.41	0.07	73.59	$p=.059$
Hedges	0	75.70	0.01	85.30	$p=.012$
Amplifiers	0.08	50.90	0.2	110.10	$p<.001$
Sentence relatives	0.04	74.47	0.05	86.53	$p=.093$
Direct WH-questions	0.01	81.19	0.01	79.81	$p=.815$
Possibility modals	0.3	78.43	0.31	82.57	$p=.572$
Phrasal coordination	1.09	96.41	0.89	64.59	$p<.001$
WH-clauses	0.14	88.29	0.1	72.71	$p<.050$
Stranded prepositions	0.27	91.86	0.2	69.14	$p<.010$
Total adverbs	3.89	85.85	3.76	75.15	$p=.144$
Conditional adverbial subordinators	0.12	78.66	0.12	82.34	$p=.614$
Features with negative loadings					
Total other nouns	23.36	99.39	21.16	61.61	$p<.001$
Word length	4.46	65.59	4.51	95.41	$p<.001$
Total prepositional phrases	10.18	48.55	11.63	112.45	$p<.001$
Type–token ratio	222.46	87.53	218.54	73.47	$p=.055$
Attributive adjectives	6.5	83.46	6.36	77.54	$p=.419$
Place adverbials	0.52	64.77	0.66	96.23	$p<.001$
Agentless passives	0.78	78.08	0.8	82.93	$p=.508$

Appendix 2. *(continued)*

	YT ($n=80$)		HT ($n=80$)		
Linguistic features	**Mean score**	**Mean rank**	**Mean score**	**Mean rank**	**Mann-Whitney pairwise comparisons**
Past participial WHIZ deletion relatives	0.17	98.26	0.11	62.74	$p<.001$
Present participial WHIZ deletion relatives	0.23	76.50	0.25	84.5	$p=.275$
Dimension 2					
Features with positive loadings					
Past tense	9.45	91.32	8.99	69.68	$p<.010$
Third person pronouns	7.96	81.99	7.84	79.01	$p=.685$
Perfect aspect	1.16	73.56	1.23	87.44	$p=.058$
Public verbs	0.55	78.47	0.56	82.53	$p=.580$
Synthetic negation	0.25	93.41	0.19	67.59	$p<.001$
Present participial clauses	0.41	90.68	0.35	70.32	$p<.010$
Features with negative loadings					
Present tense	0.8	80.03	0.72	80.97	$p=.898$
Attributive adjectives	6.5	83.46	6.36	77.54	$p=.419$
Past participial WHIZ deletion relatives	0.17	98.26	0.11	62.74	$p<.001$
Word length	4.46	65.59	4.51	95.41	$p<.001$
Dimension 3					
Features with positive loadings					
WH relative clauses on object position	0.04	75.71	0.05	85.29	$p=.181$
Pied-piping relative clauses	0.07	57.13	0.14	103.88	$p<.001$
WH relative clauses on subject position	0.13	75.68	0.15	85.32	$p=.188$
Phrasal coordination	1.09	96.41	0.89	64.59	$p<.001$
Nominalisations	0.88	48.90	1.38	112.1	$p<.001$
Features with negative loadings					
Time adverbials	0.63	66.04	0.74	94.96	$p<.001$
Place adverbials	0.52	64.77	0.66	96.23	$p<.001$
Total adverbs	3.89	85.85	3.76	75.15	$p=.144$

Appendix 2. *(continued)*

Linguistic features	YT ($n=80$) Mean score	YT ($n=80$) Mean rank	HT ($n=80$) Mean score	HT ($n=80$) Mean rank	Mann-Whitney pairwise comparisons
Dimension 4					
Features with positive loadings					
Infinitives	2.36	83.54	2.28	77.46	$p=.406$
Predictive modals	0.24	70.36	0.3	90.64	$p=.006$
Suasive verbs	0.6	94.36	0.47	66.64	$p<.001$
Conditional adverbial subordinators	0.12	78.66	0.12	82.34	$p=.614$
Necessity modals	0.1	66.31	0.15	94.69	$p<.001$
Split auxiliaries	0.36	68.90	0.44	92.10	$p=.002$
Possibility modals	0.3	78.43	0.31	82.57	$p=.572$
Dimension 6					
Features with positive loadings					
That verb complements	0.25	63.68	0.35	97.32	$p<.001$
Demonstratives	0.87	91.06	0.76	69.94	$p<.010$
That relative clauses on object position	0.09	63.10	0.15	97.90	$p<.001$
That adjective complements	0.05	66.47	0.08	94.53	$p<.001$
Stranded prepositions	0.27	91.86	0.2	69.14	$p<.010$
Existential *there*	0.12	64.84	0.17	96.16	$p<.001$
Demonstrative pronouns	0.41	100.73	0.29	60.27	$p<.001$
WH relative clauses on object position	0.04	75.71	0.05	85.29	$p=.181$
Features with negative loadings					
Phrasal coordination	1.09	96.41	0.89	64.59	$p<.001$

Address for correspondence

Kanglong Liu
The Hong Kong Polytechnic University
AG518, Core A, CBS
Hung Hom, Kowloon
Hong Kong
klliu@polyu.edu.hk
https://orcid.org/0000-0003-3962-2563

Co-author information

Isabelle Chou
University of Electronic Science and Technology of China
isabellecchou@uestc.edu.cn
https://orcid.org/0000-0003-1739-431X

Publication history

Date received: 29 January 2022
Date accepted: 16 November 2023
Published online: 11 December 2023