

13 Lexical Bundles in the Fictional Dialogues of Two *Honglouloumeng* Translations

A Corpus-Assisted Approach

*Kanglong Liu, Joyce Oiwan Cheung, and
Riccardo Moratto*

13.1 Introduction

Acclaimed¹ as one of China's four great classical novels, the Chinese classic *Dream of the Red Chamber*, or in Chinese, *Honglouloumeng* (hereinafter HLM), has drawn attention from both literary and translation researchers over decades. The work is widely acknowledged as one of the greatest Chinese fictions for it paints a vivid picture of the aristocratic families against the broad social background of the late Qing Dynasty (1644–1911). The first 80 chapters of this 120-chapter chronicle were composed by the Qing writer Cao Xueqin, and the Qing scholar Gao E completed the remaining 40 chapters after Cao's death (Cao and Gao 1982).

As a renowned Chinese literary work, the novel has been translated numerous times, hence providing scholars with a good source for comparative translation analysis. From 1979 to 2013, over 1,300 HLM research articles were published, with a majority focusing on the English translations of this classic (Ran and Yang 2013). There are three full-length versions, namely, *The Story of the Stone*, translated by David Hawkes and his son-in-law, John Minford; *A Dream of Red Mansions*, by Xianyi Yang and his wife, Gladys Yang; and *The Red Chamber Dream*, by B. S. Bonsall. The Bonsall version has never been officially published but is currently archived in the University of Hong Kong Library (Bonsall 2004), whereas the first two versions have been read by many people across the globe. Hawkes translated the first 80 chapters, and Minford finished the remaining 40, which parallels the division of labor between the two HLM writers, Cao Xueqin and Gao E. On the other hand, Xianyi Yang seemed to be the major translator of HLM, while his wife, Gladys Yang, served an assisting role. As stated by their daughter Chi Yang (cited in Li et al. 2011, 163):

When he [Xianyi Yang] was translating at his top speed, he didn't write, but simply rendered orally while my mother would type the translation on a typewriter. While she was typing the text, she also polished or edited it. So the translation was ready when all this was done.

Wang (2016) comments that Hawkes and Minford's HLM translation is extremely popular among the broad reading public in comparison to the Yangs' version. Such a difference in popularity has led to a number of studies exploring the various linguistic features between these two versions and translation strategies employed by respective translators.

The advances in corpus-based translation studies initiated by Baker (1993) have provided an impetus for translation/translator style research. According to 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." Thus, similar to the research on translation universals, researchers have made use of various language indicators, such as type-token ratio, sentence length, lexical density, which are believed to be the translators' "characteristic use of language and linguistic habits" (Baker 2000, 245), to examine how translators or translations differ. So far as HLM translations are concerned, researchers have compiled parallel and comparable corpora to examine how translations differ in a range of the aforementioned indicators. For example, previous research on HLM translations has identified that Hawkes diverged from the Yangs in various stylistic features (Li et al. 2011; Liu 2008; Liu and Afzaal 2021). In particular, based on the first 15 chapters of the two translation versions, Li et al. (2011) found that Hawkes's version contained more tokens and used longer sentences than did the Yangs', whereas the latter used a wider range of words, as reflected in a higher type-token ratio. Other linguistic indicators that have been used to study HLM translations include nominalization (Hou 2013), vocabulary richness (Fang and Liu 2015), and even idioms (Su 2021). To a large extent, researchers are largely confined to the use of word-level indicators to approach the style of HLM translations. As argued by Mastropierro (2018), the use of lexical bundles (LBs), or key clusters, can serve as a reliable indicator of translator's style, as they can reveal the translators' idiosyncrasies beyond the use of words. Following Mastropierro, the current study will make use of lexical bundles as a linguistic indicator to examine the fictional dialogues of the first 80 HLM chapters respectively translated by David Hawkes and Xianyi Yang and Gladys Yang.

13.2 Literature Review

13.2.1 Translation Style Research

In order to properly define "translation style," we must know the definition of *style* in the field of literary studies. Crystal (1999, 323) stated that style is "any situationally distinctive use of language, and of the choices made by individuals and social groups in their use of language." Leech and Short (1981) specifically proposed four main categories for style analysis in literary works, including lexical category, grammatical category, figures of speech, as well as cohesion and context. Style research in the field of translation studies, to a large extent, borrows heavily from similar research in literary studies. With the rise of descriptive

translation studies (DTS), which aims at studying translation in its own right and situating it within the target social-cultural background, translation style research has attracted considerable scholarly attention from researchers working in corpus-based translation studies. The traditional prescriptive notion that translation should be faithful to the source text has largely lost its appeal due to the shift toward DTS. Generally speaking, style research mainly falls into two major strands: translator style and translation style. The first one concerns the use of a comparable corpus (Bosseaux 2007; Saldanha 2011) to study the oeuvre of a translator as opposed to the other by capturing “the translator’s characteristic use of language, his or her individual profile of linguistic habits, compared to other translators” (Baker 2000, 245). On the other hand, translation style research is often conducted based on a parallel corpus to examine how two or more translations of a particular work diverge from each other in certain linguistic indicators or features (Li et al. 2011; Mastropierro 2018). However, the two terms are sometimes used interchangeably, as it is practically impossible to examine all the translated works of a translator. Similar to the translation universals (TUs) research, which has benefited from the use of corpus tools, translation style research has also benefited from the methodology of TUs research, including the use of linguistic indicators and analytical frameworks. In the case of HLM, the two full-length translations, which were done at roughly the same time (i.e., 1970s–1980s), have provided a good source for the current study to examine how they differ in style.

13.2.2 Previous Studies on Style in English Translations of HLM

Over the years, HLM and its translations have attracted much attention from translation scholars. As a monumental literary work, HLM has multiple translations, including some partial and complete translations. So far, most research efforts have been devoted to comparing the two full-length translation versions, namely, the one translated by David Hawkes and John Minford, and the other by Xianyi Yang and Gladys Yang. Early research on HLM translations is mainly based on qualitative deliberations. According to Yan’s (2005) systematic review of 50 research articles on HLM translations, a majority have adopted comparative methods to study a wide range of topics, ranging from poems to rhetoric devices. Some of the most frequently investigated topics in HLM translations include culture-specific items, book titles, idioms, character names, rhetoric devices, and history of translation. More recent publications also investigated how social terms (Tsao 2020) and material culture-loaded words (Yu 2020) are translated in HLM translations. Some other recent works also scrutinized letters exchanged between translators to discuss the commissioners behind HLM translations (Tong and Morgan 2021). Qualitative HLM research in general has studied a wide range of issues related to HLM translations in a descriptive yet case-by-case manner.²

With the rise of corpus linguistics in the field of translation studies, corpus methods have also been adopted to systematically analyze styles in the HLM translations. To this end, researchers often compiled parallel corpus consisting of the Chinese source text and the English translations. For example, Liu (2008)

compared how titles and honorifics were handled in HLM translations. Ji and Oakes (2018) studied earlier HLM translations produced in the eighteenth century using corpus methods and found that Edward Bowra used more conjunctions and genitives while H. Bencraft Joly used more determiners which largely characterized Joly's translator style. Joly's version was also compared with the Yangs' in Hou (2013), which revealed that nominalization construed formality in Joly's version but conciseness in the Yangs' version. In two doctoral theses, Hawkes's and the Yangs' HLM translations have been studied in detail: Mu (2012) found that Hawkes's style emphasized events and feelings by following the Western narrative convention; on the other hand, the Yangs' style was found to be non-event-oriented and less direct. Wu (2021) further used Biber's multidimensional analysis to analyze the acceptability of Hawkes's and the Yangs' versions respectively. From the development of corpus research on HLM translations, we can see the use of various linguistic indicators – from tokens and lexical types in Li et al. (2011), sTTR and lambda in Fang and Liu (2015), to metaphorical idioms in Su (2021) and lexical bundles in Liu and Afzaal (2021).

13.2.3 Lexical Bundles as an Indicator in Translation Style Research

Lexical bundles (LBs), also known as multiword expressions (MWEs), ngrams, and formulaic sequences, mean recurring lexical sequences in a register (Biber et al. 2004). In the field of second language acquisition, the use of LBs has been found to be one of the features distinguishing native from non-native English (e.g., Chen and Baker 2010, Wei 2007); recently, LBs have also been affirmed an effective indicator for investigating translator's style as well. Mastropierro (2018) compared LBs in two English-Italian translations of a thriller and found that one translator used significantly more bundles than the other. While acknowledging the merits of using LBs in translation style research, Mastropierro (2018) proposed that LBs can be categorized into groups which may disclose a translator's linguistic patterns and habits. As noted by Mahlberg et al. (2019), LBs are sometimes marked features of a specific character; thus, the use of different LBs can help construct characters with its various functions of "negotiation of information, turn-taking, politeness, and first-person narration" (Mahlberg and Hoey 2012, 76). In terms of translation, translators' use of LBs not only shows their linguistic preferences and characterization of the fictional characters but also impacts on the readability of their translations. Shrefler (2011) argued that Martin Luther's German translation of the Bible is more reader-friendly because of his frequent use of verb-related LBs. Accordingly, the use of LBs is closely connected with translation style research.

As a matter of fact, LBs have been used in *Hongloulmeng* translation research. Based on the first 15 chapters of HLM translations, Liu and Afzaal (2021) demonstrated that Hawkes's translation is embedded with a greater number and variety of LBs than the Yangs' version. Although their study has shown major differences in the use of LBs between the two HLM translations, it is believed that a study taking all 80 chapters into consideration should yield more rigorous results. Moreover, responding to Axelsson's (2008) call on treating fictional dialogue and

narration as two separate genres, a study on examining the use of LBs in HLM fictional dialogues can yield some new insights into HLM translation style research. Therefore, the current study focused on the dialogue part of both translations (all 80 chapters) to examine how the two diverge in translation style. The representation of LBs in respective translations serves as a departure point for the identification of the “the specific translator’s idiosyncrasies and conscious interpretive or unconscious idiolectal choices” (Munday 2012, 144).

13.2.4 Research Questions

Based on the foregoing review, we can see that lexical bundles can be used as a reliable indicator for translation style research. Though such an indicator has been used to explore some parts of HLM translations (Liu and Afzaal 2011), no research has been conducted to systematically examine all 80 chapters translated by Hawkes and the Yangs. Besides, no research has so far attempted to separate HLM into fictional dialogues and narration. Thus, we believe that a study aiming at examining how lexical bundles are represented in HLM fictional dialogues can provide novel insights into this line of research. In this study, we aim at addressing the following three research questions:

- (1) Do the two *Honglouloumeng* translations differ in style as represented by the frequency and types of lexical bundles?
- (2) If such differences are identified, do they diverge in terms of the structural and functional categories of the key lexical bundles?
- (3) What are the possible factors contributing to the different use of lexical bundles in the two *Honglouloumeng* translations?

13.3 Data and Procedure

13.3.1 Corpus

The current study made use of the English-Chinese Parallel Corpus of *Honglouloumeng*, which was built by Li et al. (2011). The corpus was compiled by either scanning hard copies or downloading soft copies from the internet. It consists of three parts running in parallel, namely, the original Chinese texts, the translation by Hawkes and Minford, and the translation by the Yangs. The current research is based on the first 80 chapters of the two translations. In other words, the part translated by Minford is not included in our study.

A self-written Python program was utilized to automatically extract the dialogues using punctuation (in this case, quotation marks) to separate fictional dialogues from narrations. The data were then manually proofread to ensure accuracy, as some quotation marks are used to mark titles or emphasize certain details instead of indicating dialogues. Upon completion, we have compiled two corpora, namely, the Yangs Dialogue Corpus (YD) and the Hawkes Dialogue Corpus (HD). YD consists of 219,478 tokens (i.e., the total number of orthographic words separated by

Table 13.1 Descriptive Statistics of Fictional Dialogues in HD and YD

<i>Measures</i>	<i>HD</i>	<i>YD</i>
Tokens	280,716	219,768
Types	10,730	9,801
TTR ¹	3.82	4.47
STTR ²	39.28	42.14

Source: ¹TTR = type-token ratio.
²sTTR = standardized type-token ratio.

spaces and punctuations) and 9,801 types (i.e., the number of distinct words in the corpus), while HD has 280,682 tokens and 10,734 types (see Table 13.1). Although Hawkes used more words to translate the first 80 chapters, by dividing the number of types by tokens (i.e., type-token ratio or TTR) we can see a higher TTR in YD, showing that the Yangs used a wider range of distinct words. As YD and HD differ in size, standardized TTR (sTTR) of the two corpora were also calculated by working out the average of all the TTRs per 1,000 words. YD has a higher sTTR than HD, confirming that the Yangs indeed used more distinct words than Hawkes did.

13.3.2 Analytical Framework

In order to identify the representative LBs used by Hawkes and the Yangs, we used WordSmith 8.0 (Scott 2020) to firstly turn both corpora into index files, which were then used to generate lists of three-word and four-word LBs with their corresponding frequencies. Most studies have opted for a frequency threshold for retrieving LBs, ranging from 10 (Biber et al. 1999), 20 (Cortes 2004; Hyland 2008), to 40 times (Biber et al. 2004; Pan et al. 2016) per million words (pmw). In view of the corpus size and the purpose of the current study, we have opted for a threshold of three times to retrieve the three-word and four-word LBs. Details of the retrieved LBs can be seen in Table 13.2. Based on the statistics, YD contains fewer tokens and types of both three-word and four-word LBs than HD. This is normal, considering the relatively smaller size of YD compared to HD. Further comparison of the TTRs reveal that YD has higher TTRs in both three-word and four-word LBs.

Table 13.2 Types and Tokens of 3-Word and 4-Word LBs in HD and YD

<i>Measures</i>	<i>HD</i>	<i>YD</i>
Tokens of 3word LBs	60,538	32,692
Types of 3word LBs	10,498	6,235
TTR of 3word LBs	17.34	19.07
Tokens of 4word LBs	12,867	5,972
Types of 4word LBs	2,931	1,413
TTR of 4word LBs	22.78	23.66

Based on the two lists of LBs, we further adopted the structural and functional classifications framework proposed by Biber et al. (2004) to investigate how Hawkes and the Yangs used LBs differently. Structural classification is a system which broadly categorizes expressions into different groups based on their part of speech (POS) information. For LBs which contain at least one verb component, they are classified as verbphrase-based (VPbased). For the LBs which do not have any verb components, they are classified as nounphrase-based (NPbased) if a noun component comes before prepositions or other POS components. In case a preposition comes before nouns, the expression is then classified as prepositional phrase-based (PPbased). As for those without any verbs, nouns, or prepositions, they are classified as others. While structural classification is useful in differentiating the structural patterns of LBs preferred by respective translators, functional classification enables a comparison of the LBs in terms of their communicative goals. The LBs can be broadly categorized into stance, discourse markers, referential, and special conversational functions, depending on their use in the context. Sometimes an expression may perform more than one function. For example, *I want to* can be a discourse marker which introduces a topic; alternatively, *I want to* can also be used to express desire. To decide on the major function of an expression, we employed a context-based annotation. In other words, the LBs were studied in the context before we ultimately annotated the expression with its key function.

In this study, we conducted two rounds of Key-LBs analysis. In the first round, we compared the YD LBs against the HD LBs as the reference corpus to identify the Key-LBs used in YD. In the second round, the two lists were reversed in order to identify the Key-LBs in HD. LBs having passed the keyness tests in the analyses (i.e., loglikelihood > 6.63) would be considered Key-LBs, meaning, that these LBs have an unusually high frequency in their respective corpus.³ Among these LBs, some content expressions, mainly, character and place-names, such as *Our Old Lady*, which are irrelevant for the analysis were redacted, leaving us with 57 and 139 LBs types in YD and HD, respectively. We applied the structural classification (i.e., NPbased, VPbased, PPbased, and others) and functional classification (i.e., stance, discourse organizers, referential, and special conversational functions) (Biber et al. 2004) to classify the Key-LBs, with the ultimate aim to identify how HD and YD diverge in style represented by the use of LBs.

13.4 Results

13.4.1 Structural Patterns

Although YD yielded a higher TTR of LBs than HD, we only identified 57 Key-LBs in YD; HD, on the other hand, showed a lower TTR of LBs but recorded 139 Key-LBs (see Table 13.3). This reveals that TTR might not be a reliable indicator if we are comparing two LBs lists that differ in length. We found that HD and YD differ not only in the number of Key-LBs but also in structures and functions.

While both Key-LBs in HD and YD are mostly VPbased (i.e., consisting of a verb component), HD has a higher proportion of VPbased Key-LBs (75.54%) than that of YD (61.40%). The result shows that HD is closer to Conrad and Biber's (2005) finding that 90% of the LBs used in spoken British English involve verb components. On the other hand, a higher proportion of PPbased Key-LBs (i.e., bundles starting with a preposition) is found in YD (17.54%) than HD (7.91%).

Since the majority of Key-LBs in HD and YD are VPbased, which involve at least one verb component, thus we proceeded to study their subpatterns (see Table 13.4). Our findings revealed that 40.95% of Key-LBs in HD started with a personal pronoun (e.g., *I, you, she*), 29.52% started with a verb (e.g., *be, do, have*, modal, or other verbs), and 20.95% started with either a conjunction or linking words, such as *that* and *to* (see Table 13.3). We further categorized the VP-based Key-LBs for their subcategories (see Table 13.4). Likewise, the PP-based Key-LBs were also further categorized for their subcategories (see Table 13.5).

Table 13.3 Structural Classifications of Key-LBs in HD and YD

Structural Classifications	HD		YD	
	Key-LBs	%	Key-LB	%
NPbased	21	15.11	9	15.79
VPbased	105	75.54	35	61.4
PPbased	11	7.91	10	17.54
Others	2	1.44	3	5.26
Total	139	100	57	100

Table 13.4 Statistics of VP-Based Key-LBs in HD and YD

VP-Based Key-LBs	Types in HD	%	Types in YD	%
Starting with personal pronouns	43	40.95	5	14.29
Starting with verbs (including <i>be, do, have</i> , modal verbs, and other verbs)	31	29.52	15	42.86
Starting with conjunctions, <i>that, to</i> , or <i>not to</i>	22	20.95	6	17.14
Starting with <i>wh</i> words	5	4.76	8	22.86
Starting with existential markers (including <i>there</i> and <i>this</i>)	2	1.9	1	2.86
Starting with an adjective	2	1.9	0	0
Total	105	100	35	100

Table 13.5 Statistics of PP-Based Key-LBs in YD

PP-Based Key-LBs	HD	%	YD	%
Starting with a preposition and a determiner	5	45%	6	60.00
Starting with two prepositions	0	0%	1	10.00
Starting with conjunction	3	27%	3	30.00
Total	3	27%	10	100

13.4.2 Contextual Use of Key VP-Based and PP-Based LBs

In this section, the two most common types of VP-based Key-LBs (i.e., those starting with personal pronouns and those starting with verbs) and the PP-based Key-LBs will be further discussed in relation to some examples extracted from HD and YD.

Many of Hawkes's VP-based Key-LBs are headed by a personal pronoun. *I think you* is the LB that is most significantly different between HD and YD (LL: 49.70), showing a clear overrepresentation in HD. This phrase usually appears at the beginning of a sentence and manifests the subject prominence in English. As we can see in excerpt 1, the suggestion of paying someone a visit is expressed in the form *I think you should* (i.e., first personal pronoun + verb base + second personal pronoun) in HD. Meanwhile, such subjectpredicate relation is absent in YD, which simply used the directive *Go* to express the character's permission of the visit, which is a topic that has already been introduced in the previous dialogue exchange. YD prioritized the topic (*Go*), whereas HD adhered to the English convention of subject prominence (e.g., *She is, I think you*). As can be seen, Hawkes tended to use subjectpredicate structures (e.g., personal pronoun + verbs), whereas such structures are less found in the Yangs' version.

Excerpt 1

“你看看就過去罷，那是侄兒媳婦。” [Source] (Chapter 11)

“Yes,” “she is your nephew's wife. I think you should. Just look in for a moment, though, and then join the rest of us.” [Hawkes]

“Go if you want, but don't be long,” “Remember she's your nephew's wife.” [Yangs]

Similar contrast is also observed in Key-LBs which begin with a verb. *Ought to be* is the most significant KeyLB in HD (LL: 36.02), which starts with a verb component. As we can see in excerpt 2, *ought to be* follows the subject *you* in HD. In his rendition, Hawkes translated the invitation 請, *qing* (literal translation: please), using a subject (*you*) and its predicate (*ought to be getting back . . .*). The Yangs, on the other hand, did not use the subjectpredicate structure but instead retained the semantic meaning (*please*) of 請, *qing*, in the source text. Since *please* is a near equivalent of 請, *qing*, the Yangs used literal translation by following the same sentence order as that of the source text. Subject is again omitted in the Yangs' version. Excerpts 1 and 2 are just two of the many examples contrasting Hawkes's and the Yangs' preferences for subjectpredicate and topic-comment structures, respectively. Overall, we can safely conclude that Hawkes's Key-LBs follow the spoken English convention in which most of the LBs involve verb components (Conrad and Biber 2005) structured in the form of personal pronouns + verb (Biber 2009).

Excerpt 2

“如今來回老祖宗，債主已去，不用躲了。已預備下希嫩的野鷄，請吃晚飯去，再遲一會子就老了。” [Source] (Chapter 50)

“So now your creditors have gone, you can come out of hiding. You ought to be getting back now in any case. You’ve got some nice, tender pheasant for dinner and if you leave it much longer it will spoil.” [Hawkes]

“Now I’ve come to report to our Old Ancestress: Your duns have gone, you can come out of hiding. I’ve some very tender pheasant ready. So please come back for dinner. If you leave it any later, it’ll be overcooked.” [Yang]

However, this is not the case in YD. Although more than half of the Yangs’ Key-LBs are still VPbased, this proportion is still fewer than that of HD because 17.54% belong to PPbased LBs. Meanwhile, only 7.91% of Hawkes’s Key-LBs are PPbased. This indicates that YD has used more PPbased LBs which were significantly underused by Hawkes when translating *Honglouloumeng* (see Table 13.4). Yip (1995, 78) pinpointed that bare noun phrases are often placed in the beginning of a Chinese sentence to refer to a topic due to topic prominence, but such a syntactic structure (i.e., sentences beginning with a bare noun) is not really natural in English. Hence, Yip believed that Chinese speakers strategically use prepositional phrases to encapsulate a bare noun phrase when they need to first talk about a topic.

Based on the results, it can be seen that using a prepositional noun phrase to start a sentence is more prevalent in YD than HD. For example, the Yangs used *If not for* (LL: 29.64) significantly more frequently than Hawkes did. *If not for* is a typical prepositional phrase which consists of the conjunction *if*, the adverb *not*, and the preposition *for*. In excerpt 3, we can see that the source text in Chinese is structured as 要不是 (*if not*) and 我 (*me*), which the Yangs directly translated into *If not for me*. As the focus is on the speaker holding back the other one from attacking people, the Yangs kept this topic in the translation and used the prepositional phrase *If not for* to topicalize the object *me*. The syntactic order of *If not for me* is almost an equivalent to the dependent clause 要不是我 (literally: *if not me*) in the Chinese source text. Conversely, Hawkes followed the subjectprominent convention by using a verb phrase to start the sentence. He used the verbpronounverb clause *Suppose I hadn’t been here* to describe a condition that is contrary to fact.

Excerpt 3

“要不是我，你要傷了他的命，這會子可怎麼樣？” [Source] (Chapter 44)

“If not for me you might have killed her. What do you intend to do now?” [Yang]

“Suppose I hadn’t been here to protect her and you really had done her an injury, what would you have had to say for yourself then, I wonder?” [Hawkes]

The Yangs also used prepositional phrases at the end of sentences. For example, they extensively used *for no reason* to express the absurdity of a situation. *For no reason* is one of the Key-LBs in YD consisting of a preposition, a determiner, and a noun, which yielded a very high keyness value (LL: 31.29), meaning, that it is overrepresented in YD than HD. PPbased LBs like *for no reason*, when placed at the end of a sentence, often serve as an adverbial. From excerpt 4 we can see that the Yangs used this prepositional phrase to describe the unlikeliness that someone would offend those people. The Yangs not only used prepositional phrases to make noun phrase topics grammatically well-formed (e.g., excerpt 3) but also used them to describe actions. However, no such substantial use of prepositional phrases was found in Hawkes's dialogue translation. Hawkes used a variety of linguistic choices to achieve the same purpose; in this case, he used the adverb *possibly* to express the unlikeliness of the event. So far, our study has found that there are more unique VPbased LBs in HD and more distinctive PPbased LBs in YD. Our findings revealed that the Yangs seemed to prefer using prepositions to introduce noun topics, while Hawkes used more verb phrases to express subject-predicate relations.

Excerpt 4

“誰可好好的得罪着他?” [Source] (Chapter 78)

“Why should anyone offend them for no reason.” [Yang]

“Who could possibly have offended her?” [Hawkes]

13.4.3 Functional Classifications

After manual classification, it was found that 47.48% of Hawkes's Key-LBs mainly expressed stances, while 36.84% of the Yangs' Key-LBs mainly served as referential bundles (see Table 13.6). This means almost half of Hawkes's unique LBs come from his use of stance markers. Thus, these two functional categories were further examined in detail. In order to show how HD diverged from the YD in the use of stance markers, we further categorized the stance markers for their subpatterns (see Table 13.7). Likewise, we also further categorized the referential Key-LBs in HD and YD for their subpatterns (see Table 13.8).

Table 13.6 Functional Classifications of Key-LBs in HD and YD

Functional Classifications	HD	%	YD	%
Stance	66	47.48	10	17.54
Discourse organizers	31	22.3	10	17.54
Referential	37	26.62	21	36.84
Special conversational functions	5	3.6	16	28.07
Total	139	100	57	100

Table 13.7 Statistics of Stance Key-LBs in HD and YD

<i>Stance Functions</i>	<i>HD</i>	<i>%</i>	<i>YD</i>	<i>%</i>
Epistemic stance	20	30.30	3	30
Overall attitudinal/modality stance	4	6.06	0	0
Desire	4	6.06	0	0
Obligation/directive	19	28.79	4	40
Intention/prediction	13	19.70	1	10
Ability	6	9.09	2	20
Total	66	100.00	10	100

Table 13.8 Statistics of Referential Key-LBs in HD and YD

<i>Referential Functions</i>	<i>HD</i>	<i>%</i>	<i>YD</i>	<i>%</i>
Identification/focus	7	18.92	4	19.05
Imprecision	6	16.22	1	4.76
Quantity/specification	9	24.32	5	23.81
Intangible framing attributes	6	16.21	4	19.05
Place reference	1	2.70	1	4.76
Time reference	3	14.29	3	14.29
Multifunctional reference	1	2.70	3	14.29
Total	37	100	21	100

13.4.4 Contextual Use of Key Stance and Referential LBs

According to Biber and Barbieri (2007), the predominant function of LBs in all spoken registers (i.e., teaching, class management, office, study groups, and service encounters) is to express stance. It seems that Hawkes tended to use stance markers to translate the fictional dialogue. Among Hawkes's Key-LBs which are classified as stance, 30.30% construe an epistemic stance, while 28.79% convey obligations/directives (see Table 13.6). The rest are distributed among intentions/predications, desire, ability, etc. This means most of Hawkes's Key-LBs perform either an epistemic or a directive function. For instance, one of Hawkes's KeyLB, *I think I* (LL: 33.52), is a very common epistemic marker in conversational English. It indicates personal opinions and sometimes functions as a hedge to soften the illocutionary force of an assertion. In excerpt 5, Hawkes added *I think I* to express the speaker's decision to stay overnight. This use of hedging in decision-making is, however, not found in the source text. It is solely Hawkes's interpretation that a certain degree of hedging might be required in this context. Such stance markers are found neither in the source text nor in YD. The Yangs used *shan't*, the contraction form of *shall not*, to keep the formality and courtesy conveyed in the source text. On the other hand, the Yangs literally rendered the source text without adding any epistemic stances in relation to the context.

Excerpt 5

“有的是炕，只管睡。我是二爺使我送月銀的，交給了奶奶，我也不回去了。” [Source] (Chapter 65)

“There’s plenty of room here for you to sleep. Make yourselves at home. Actually, I came here to bring the mistress her monthly allowance. Now that I’ve given it to her, I think I shall spend the night here as well.” [Hawkes]

“Well, there’s plenty of room on the kang, just lie down as you like. Second Master sent me to bring the monthly allowance to the mistress, so I shan’t be going back either.” [Yang]

Apart from epistemic stances, Hawkes used significantly more LBs to perform a kind of speech act directives. Among his stance Key-LBs, 28.79% assert obligation/directives. *You ought to* (LL: 28.64) is one of the LBs with a high keyness value which is used by the speaker to imply that the listener has a sense of duty or morality to undertake a certain task. Clearly, HD contains more expressions conveying obligations and directives than YD. Take a translation pair as an example (see excerpt 6): the source text 你細想去 (literal translation: you carefully think about) does not contain any sense of obligation. However, Hawkes used *you ought to be able to* in his translation, which signaled an obligation for the listener to work things out by themselves. Such an obligation sense was not found (at least literally) in the source text, so the Yangs simply used the adverb *just* to begin the subjectless command *work it out yourself*. In view of the fact that there are more stance Key-LBs (66) in HD compared to YD (10), it can be postulated that Hawkes tended to add stance LBs in his translation while the Yangs used stance LBs to a lesser degree. Among these stance Key-LBs, Hawkes mainly used them to convey epistemic stances or obligation/directives, as has been exemplified in excerpts 5 and 6.

Excerpt 6

“非也。我哥哥已經相准了，只等來年就下定了，也不必提出人來，我方才說你認不得娘，你細想去。” [Source] (Chapter 57)

“No, that’s not the reason. It’s because someone has already been chosen for my brother. We are only waiting for him to come home to make it public. I don’t need to name names. If I tell you that you can’t possibly become Mamma’s god daughter, you ought to be able to work it out for yourself.” [Hawkes]

“No, it’s because my brother has already set his mind on someone, and it’ll be fixed up as soon as he returns. I needn’t name any names. Why did I say you couldn’t take her as your mother? Just work it out for yourself!” [Yang]

Unlike Hawkes, many of the Key-LBs in YD are referential markers. Results show that 36.84% of the frequently occurring LBs in YD were used to refer to different attributes. The referential Key-LBs in YD are distributed across many subfunctions, including identification/focus, imprecision, quantity/specification, intangible framing attributes, place, time, and multifunctional reference (see Table 13.8). Since the Yangs’ referential Key-LBs are evenly distributed across all subfunctions, we have selected two referential Key-LBs for detailed analysis

based on the two LBs' exceptionally high keyness values. The first one is *this is just* (LL: 24.70), which functions as an identification/focus marker. The Yangs used *this is just* significantly more frequently than Hawkes did (see excerpt 7). *This is just what* (YD) differs from *this way of carrying on* (HD), as the former just refers to a vague subject matter which readers can by no means infer from the literal meaning, but the latter identifies the exact misbehavior. In the source Chinese text 正爲勸你這些 (literal translation: just persuading you these), the word 這, *zhe* (literal translation: this), is exactly an identifier in Chinese. By starting a sentence with the identifier 這, *zhe*, Chinese speakers can easily follow the topic, which need not be reintroduced repeatedly. Largely a literal translation approach, the Yangs used identifiers (e.g., *this*) in their translation by adhering closely to the source text. We assume that the overuse of identification LBs in YD is thus probably a result of direct translation of Chinese identifier 這, *zhe* (i.e., *this*), which is a more economical way of introducing a mutually known topic. Hawkes, on the other hand, felt the need to explicate the topic clearly.

Excerpt 7

“好好的，正爲勸你這些，更說的狠了。” [Source] (Chapter 19)

“This is just what I wanted to warn you against, yet here you go, talking more wildly than ever.” [Yang]

“It's precisely this way of carrying on that I was going to talk to you about, and here you go, ranting away worse than ever!” [Hawkes]

Another function of the Key-LBs in YD is the use of express imprecision. *On like this* is one KeyLB in this subcategory with a high keyness value (LL: 32.94) and overused in YD than HD. This LB does not specify what qualities it is referring to. Instead, it makes the circumstances off the record and leaves readers some room for imagination. For example, in excerpt 8, the Yangs used *on like this* to refer to the girl's poor situation, which is not explicitly mentioned in the corresponding source text. The source text 這個形景 (literal translation: this situation) does not specify clearly what situation the girl is in. On the contrary, Hawkes did not use the imprecise LB *on like this* like the Yangs did but instead used the noun phrase *her outward behavior*. Again, Hawkes has given his own personal interpretation of the expression 這個形景 (i.e., *this situation*).

Excerpt 8

“這女孩子一定有什麼話說不出來的大心事，才這麼個形景。外面既是這個形景，心裏不知怎麼熬煎。看他的模樣兒這般單薄，心裏那裏還攔的住熬煎。可恨我不能替你分些過來。” [Source] (Chapter 30)

“She must have some secret anxiety preying on her mind to carry on like this, yet she looks too delicate to stand much anxiety. I wish I could share her troubles.” [Yang]

“One can see from her outward behaviour how much she must be suffering inwardly. And she looks so frail. Too frail for suffering. I wish I could bear some of it for you, my dear!” [Hawkes]

13.5 Discussion

This chapter has applied keyword analysis to identify the three-word and four-word lexical bundles (LBs) which are significantly more frequent in each of the *Honglouloumeng* translations compared to meaningful LBs of other lengths. It is found that many of Hawkes’s Key-LBs (i.e., lexical bundles unusually frequent in Hawkes’s dialogue translation but infrequent in the Yangs’ dialogue translation) are verb phrases, while many of the Yangs’ Key-LBs (i.e., bundles unusually frequent in the Yangs’ dialogue translation but infrequent in Hawkes’s dialogue translation) are prepositional phrases. We have also found that almost half of Hawkes’s Key-LBs function as stance markers, while the largest proportion of the Yangs’ Key-LBs are referential markers. In this section, Hawkes’s and the Yangs’ use of LBs will be discussed with reference to their language backgrounds, life experiences, and respective translation purposes.

13.5.1 Language Backgrounds

David Hawkes is a native English speaker, while Xianyi Yang is a native Chinese speaker. Although his wife, Gladys Yang, is a native English speaker, she mainly typed “the translation on a typewriter. While she was typing the text, she also polished or edited it” (Li et al. 2011, 163). In our study, it is found that Hawkes used more VP-based LBs, which is in line with Biber’s (2009) finding that 50% of the LBs used in native spoken English are structured as “personal pronoun + verb components.” This shows that Hawkes’s translation of fictional dialogues is largely in line with the norm of spoken English in this respect. On the contrary, Xianyi Yang, as a native Chinese speaker, is found to have used more PP-based LBs. This is also consistent with some findings that L2 speakers (e.g., native Chinese speakers) tend to overuse certain LBs which native English speakers seldom use (Chen and Baker 2010) and that Chinese speakers use more prepositions to construct lexical bundles than did their native English counterparts (Wei 2007; Chen and Baker 2010). As Chinese is a topicprominent language (Yip 1995), it is not surprising that Chinese speakers adhere to the topicprominence convention by using prepositions combined with a bare noun phrase in the topic position to ensure grammaticality in English. On the other hand, English is a subjectprominent language which often structures sentences in a subjectpredicate relation (ibid.); thus, half of the LBs in spoken English are made up of “pronoun + verb” (Biber 2009). Hawkes’s VP-based Key-LBs, such as *I think you* and *ought to be*, are manifestations of subject prominence in English; the Yangs’ PP-based Key-LBs, such as *if not for* and *for no reason*, may be influenced by topic prominence, in which preposition phrases often serve as adverbials in Chinese. This supports previous research (e.g., Yip 1995; Biber and Barbieri 2007, 2009; Conrad and Biber 2005)

that LBs in spoken English are mostly verb phrases, and Chinese speakers tend to use prepositional phrases to topicalize the bare nouns or noun phrases when they speak English.

As for the functional aspects, Hawkes's Key-LBs, such as *I think I* and *you ought to*, also resonate with the convention that the most prominent function of LBs in spoken English is stancemaking: to assert epistemic stance and give directives (Biber and Barbieri 2007). Meanwhile, the Yangs' less frequent use of stance bundles might be related to the fact that Chinese speakers often underuse participant-oriented LBs (Wei and Lei 2011; Pan and Liu 2019). The Yangs' overuse of LBs such as *this is just* and *on like this* reflects Chinese speakers' frequent use of identifiers to express mutually known topics. Hence, Hawkes's frequent use of verb phrases and stance LBs, as well as the Yangs' frequent use of prepositional phrases and referential LBs, reveal that divergent translation styles can be attributed to the different language backgrounds of the respective translators.

13.5.2 Life Experiences

David Hawkes went to China and received postgraduate education in Beijing in 1948, while Xianyi Yang started his university education at Oxford University in 1936. According to Minford's foreword to Xianyi Yang's (2002) autobiography *White Tiger*, Xianyi and Gladys Yang would visit David and Jean Hawkes and the couples knew each other well. David Hawkes and Xianyi Yang were intellectuals in pretty much the same historical time, and they published their translations of *Honglouloumeng* about the same time as well (i.e., both finished their translations by 1980). On the other hand, Hawkes and Xianyi Yang contrast in their walks of life. David Hawkes was a sinologist who first encountered *Honglouloumeng* when he studied at Peking University. He read the novel under the guidance of a Chinese-speaking "laoxiansheng," 老先生 (translation: *old scholar*), who was a former government clerk from the Hebei province. Hawkes described the reading journey as "direct method gone mad" in a sense that he barely understood what the teacher said. Perhaps due to his unpleasant experience, Hawkes preferred a more fluent approach in rendering the fictional dialogues (more VP-based LBs) which sound as if they were naturally spoken to the readers in English. Out of his passion for the novel, Hawkes resigned from his chair professorship at Oxford in 1971 to be fully devoted to his translation of *Honglouloumeng* (Minford 2012). At that time, Hawkes was already an established scholar who had a research fellowship to live on. He did not translate for money's sake but for his sheer joy. Contrary to David Hawkes, Xianyi Yang did not have the luxury of spending years on polishing his translated work. After he and his wife joined the official translation bureau in 1943, and subsequently the Foreign Languages Press in 1952, the couple was in charge of translating literary works in new China. In the 1950s, Xianyi Yang was drained by translating foreign works into Chinese, as he also had to fulfil the "voluntary physical labor" at the same time; from 1968 to 1972, the couple suffered a hard time of being imprisoned due to the political unrest brought by the Cultural Revolution. During the two years of translating *Honglouloumeng* for the Foreign

Languages Press, they lost their beloved son. According to Xianyi Yang's autobiography (2012), they were never paid for the extra work on translation except *Honglouloumeng*, which was commissioned by the magazine *Chinese Literature*. Our findings corroborated with Li, Zhang and Liu (2011) that Xianyi Yang and Gladys Yang translated under censorship, grief, and tight schedule yet with little remuneration. This probably explains why a more literal approach was employed by the Yangs in rendering the fictional dialogues.

13.5.3 Translation Purposes

Finally, David Hawkes's translation purpose was to entertain readers and literary enthusiasts. To help reconstruct the dialogues, Hawkes has adopted a more liberal approach in his translation. For example, one of the most frequently occurring reporting verb phrase, 笑道, *xiao dao*, in the source text (literally: *said with a smile*) was translated in various ways (e.g., *childe*, *laugh*, *with a broad smile*, *with a meaningful smile*, *with a proud smile*) by Hawkes in relation to the context. Hawkes justified this approach as a measure to compensate for the absence of the tone of voice (Minford n.d., 32). In his preface to *The Story of the Stone Volume 1: The Golden Days*, Hawkes (1973, 46) stated his major concern in translating the novel: "If I can convey to the reader even a fraction of the pleasure this Chinese novel has given me, I shall not have lived in vain." When translating the dialogues, Hawkes preferred stance bundles, as they serve many communicative functions (e.g., expressing attitudes, desire, directives, intentions, predictions, abilities) which render the dialogues more engaging. On the other hand, in the Publisher's Note of *A Dream of Red Mansions Volume 1*, it was stated that *Honglouloumeng* is a book "about political struggle" (1978, iv), which "by presenting the prosperity and decline of the four typical noble families it truthfully lays bare the corruption and decadence of the feudal ruling class and points out its inevitable doom" (1978, vii). Though such a remark might result from self-censorship due to the political atmosphere of the time, such a depiction has clearly shown that ideological factors greatly outweighed aesthetic ones in the case of the Yangs. When translation becomes a task assigned by the officials, the translated work is to promote ideologies and hence leaves the translators little room for interpretation. Therefore, it is plausible that the Yangs opted for a more rigid approach to translate the novel.

13.6 Conclusion

This study sets out to compare different translators' use of lexical bundles in two *Honglouloumeng* translations. In line with Mastropierro's (2018) suggestion, we affirmed that lexical bundles can serve as a reliable indicator beyond other lexical devices for differentiating style in different translations. By examining the syntactic structures and functions of the key lexical bundles in Hawkes and the Yangs, we have found that the Yangs adopted a more literal and seemingly rigid approach to translating *Honglouloumeng*, as evidenced by the different use of key

lexical bundles from Hawkes. Our study has yielded some preliminary evidence that translators' styles may be influenced by the respective translator's language background, life experiences, and translation purposes. This study is, however, not without limitations. Only translation works by two groups of translators (i.e., Hawkes and the Yangs) were sampled in the current study. Future studies can compare more translation versions of *Honglouloumeng* to examine whether the use of lexical bundles differ among different translators as a result of their sociocultural background and translation purposes. Besides, as argued by Li and Zhang (2010, 250), "[a] corpus as well as a statistical presentation of translation or language facts is not the ultimate goal of our research, but rather the beginning and foundation for real research on whatever research questions the project is addressing." In this regard, more documentary evidence needs to be collected to verify the claims made based on corpus frequency data.

Notes

- 1 An earlier version first appeared in *Translation Quarterly* (2020), Issue 98, pp. 79–101. This present version is updated and modified based on the earlier version.
- 2 For a more detailed review of recent studies on HLM, readers may refer to Moratto et al. (2022).
- 3 Based on UCREL's (<https://ucrel.lancs.ac.uk/llwizard.html>) instruction on calculating log-likelihood and effect size, a critical log-likelihood value of 6.63 means that the null hypothesis is considered to be false (i.e., $p < 0.01$). Therefore, a log-likelihood value of 6.63 is set as threshold for Key-LBs in the current study.

References

- Axelsson, Karin. 2008. "Research on Fiction Dialogue: Problems and Possible Solutions." In *Corpora: Pragmatics and Discourse*, edited by Andreas H. Jucker, Daniel Schreier, and Marianne Hundt, 189–201. Leiden: Brill.
- Baker, Mona. 1993. "Corpus Linguistics and Translation Studies: Implications and Applications." In *Text and Technology: In Honor of John Sinclair*, edited by Mona Baker, G. Francis, and E. Tognini-Bonellis. Amsterdam: John Benjamins.
- Baker, Mona. 2000. "Towards a Methodology for Investigating the Style of a Literary Translator." *Target* 12, no. 2: 241–66.
- Biber, Douglas. 2009. "A Corpusdriven Approach to Formulaic Language in English: Multiword Patterns in Speech and Writing." *International Journal of Corpus Linguistics* 14, no. 3: 275–311.
- Biber, Douglas, and Federica Barbieri. 2007. "Lexical Bundles in University Spoken and Written Registers." *English for Specific Purposes* 26, no. 3: 263–86.
- Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. "If You Look At . . . : Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25, no. 3: 371–405.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Bonsall, Bramwell Seaton, Trans. 2004. *The Red Chamber (Honglouloumeng)*. Hong Kong: The University of Hong Kong. <https://lib.hku.hk/bonsall/honglouloumeng/title.pdf>.
- Bosseaux, Charlotte. 2007. *How Does It Feel? Point of View in Translation: The Case of Virginia Woolf into French*. Amsterdam: Rodopi.

- Cao Xueqin, and E. Gao. 1982. *Hong Lou Meng [in Chinese]*. Beijing: People's Literature Press.
- Chen, Yuhua, and Paul Baker. 2010. "Lexical Bundles in L1 and L2 Academic Writing." *Language Learning & Technology* 14, no. 2: 30–49.
- Conrad, Susan, M., and Douglas Biber. 2005. "The Frequency and Use of Lexical Bundles in Conversation and Academic Prose." *Lexicographica* 20: 56–71. <https://doi.org/10.1515/9783484604674.56>.
- Cortes, Viviana. 2004. "Lexical Bundles in Published and Student Disciplinary Writing: Examples from History and Biology." *English for Specific Purposes* 23, no. 4: 397–423.
- Crystal, David. 1999. *The Penguin Dictionary of Language*. London: Penguin Books.
- Fang, Yu, and Haitao Liu. 2015. "Comparison of Vocabulary Richness in Two Translated Hongloumeng." *Glottometrics* 31: 54–75.
- Hawkes, David. 1973. *The Story of the Stone Volume 1: The Golden Days*. London: Penguin.
- Hou, Yu. 2013. "A Corpus-Based Study of Nominalization as a Feature of Translator's Style (Based on the English Versions of Hong Lou Meng)." *Meta* 58, no. 3: 556–73.
- Hyland, Ken. 2008. "As Can be Seen: Lexical Bundles and Disciplinary Variation." *English for Specific Purposes* 27, no. 1: 4–21.
- Ji, Meng, and Michael P. Oakes. 2018. "A Corpus Study of Early English Translations of Cao Xueqin's Hongloumeng." In *Quantitative Methods in Corpus-based Translation Studies*, edited by Michael P. Oakes and Meng Ji, 177–208. Amsterdam: John Benjamins.
- Leech, Geoffrey N., and Mick Short. 1981. *Style in Fiction: A Linguistic Introduction to English Fictional Prose*. Longman: London.
- Li, Defeng, and Chunling Zhang. 2010. "Sense-Making in Corpus-Assisted Translation Research." In *Using Corpora in Contrastive and Translation Studies*, edited by Richard Xiao, 235–54. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Li, Defeng, Chunling Zhang, and Kanglong Liu. 2011. "Translation Style and Ideology: A Corpusassisted Analysis of Two English Translations of Hongloumeng." *Literary and Linguistic Computing* 26, no. 2: 153–66.
- Liu, Kanglong, and Muhammad Afzaal. 2021. "Translator's Style Through Lexical Bundles: A Corpusdriven Analysis of Two English Translations of Hongloumeng." *Frontiers in Psychology* 12. <https://doi.org/10.3389/fpsyg.2021.633422>.
- Liu, Zequan. 2008. "Translating Tenor: With Reference to the English Versions of Hongloumeng." *Meta* 53, no. 38: 528–48.
- Mahlberg, Michaela, and Michael Hoey. 2012. *Corpus Stylistics and Dicken's Fiction*. London: Routledge.
- Mahlberg, Michaela, Viola Wiegand, Peter Stockwell, and Anthony Hennessey. 2019. "Speech Bundles in the 19th Century English Novel." *Language and Literature* 28, no. 4: 326–53.
- Mastropierro, Lorenzo. 2018. "Key Clusters as Indicators of Translator Style." *Target* 30, no. 2: 240–59.
- Minford, John. 2012. "A Tribute to Brother Stone." In *Style, Wit and Word-Play: Essays in Translation Studies in Memory of David Hawkes*, edited by Tao Tao Liu, Laurence K. P. Wong, and Sin-wai Chan, 1–14. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Minford, John. n.d. "Hawkes' Approaches to Translating Fiction." In *A335 Culture and Translation [Course Materials]*. Hong Kong: OUHK.
- Moratto, Riccardo, Kanglong Liu, and Di-kai Chao, eds. 2022. *Dream of the Red Chamber: Literary and Translation Perspectives*. London and New York: Routledge.
- Mu, Yuanyuan. 2012. "Towards a Quantitative & Qualitative Stylistic Approach to Ideational Construal in the Translation of Narrative Discourse: Norms and Readers'

- Responses Revisited – A Corpus-based Study on Hong-lou Meng and Its Two English Translations.” PhD diss., City University of Hong Kong.
- Munday, Jeremy. 2012. *Evaluation in Translation: Critical Points of Translator Decision Making*. London and New York: Routledge.
- Pan, Fan, and Chen Liu. 2019. “Comparing L1L2 Differences in Lexical Bundles in Student and Expert Writing.” *Southern African Linguistics and Applied Language Studies* 37, no. 2: 142–57.
- Pan, Fan, Randi Reppen, and Douglas Biber. 2016. “Comparing Patterns of L1 Versus L2 English Academic Professionals: Lexical Bundles in Telecommunications Research Journals.” *Journal of English for Academic Purposes* 21: 60–71.
- Ran, Shiyang, and Ping Yang. 2013. “Breaking Through the Bottleneck: A Comparative Investigation into the ChineseEnglish Translation Studies of ‘Hong Lou Meng’ [in Chinese].” *China Publishing Journal* 12: 61–3.
- Saldanha, Gabriela. 2011. “Translator Style: Methodological Considerations.” *The Translator* 17, no. 1: 25–50.
- Scott, Mike. 2020. *WordSmith Tools Version 8*. Stroud: Lexical Analysis Software.
- Shrefler, Nathan. 2011. “Lexical Bundles and German Bibles.” *Literary and Linguistic Computing* 26, no. 1: 89–106.
- Su, Ke. 2021. “Translation of Metaphorical Idioms: A Case Study of Two English Versions of Honglouloumeng.” *Babel* 67, no. 3: 332–54.
- Tong, Jasmine Man, and David Morgan. 2021. “‘Twice Bitten’ Two Men and a Translation: The Making of the Stone.” *Babel* 67, no. 6: 791–818.
- Tsao, Lihqun. 2020. “On Translation of Social Terms in Hong Lou Meng.” *Arts Studies and Criticism* 1, no. 3: 68–73.
- Wang, Ning. 2016. “Chinese Literature as World Literature.” *Canadian Review of Comparative Literature* 43, no. 3: 380–92.
- Wei, Naixing. 2007. “Phraseological Characteristics of Chinese Learners’ Spoken English: Evidence of Lexical Chunks from COLSEC.” *Modern Foreign Languages* 30, no. 3: 281–91.
- Wei, Yaoyu, and Lei Lei. 2011. “Lexical Bundles in the Academic Writing of Advanced Chinese EFL Learners.” *RELIC Journal* 42, no. 2: 155–66.
- Wu, Chungming. 2021. “Towards Norms in Two Translations of Hong Lou Meng: A Corpus-based Study.” PhD diss., The Hong Kong Polytechnic University.
- Yan Minmin 閻敏敏. 2005. “Ershinianlai de Honglouloumeng yingyi yanjiu 二十年來的《紅樓夢》英譯研究 [Twenty Years’ Studies on the Translation of Honglouloumeng into English].” *Foreign Language Education 外語教學* 26, no. 4: 64–68.
- Yang, Hsien-Yi, and Gladys Yang, trans. 1978. *A Dream of Red Mansions Volume 1*. Beijing: Foreign Languages Press.
- Yang, Xianyi. 2002. *White Tiger: An Autobiography of Yang Xianyi*. Hong Kong: Chinese University Press.
- Yip, Virginia. 1995. *Interlanguage and Learnability: From Chinese to English*. Amsterdam: John Benjamins.
- Yu, Ke. 2020. “A Comparative Study of the Translation of Material Culture-loaded Words of Honglouloumeng in the Light of Skopostheorie.” *Journal of Language Teaching and Research* 11, no. 2: 318–23.

Appendix A

Yangs' 3-Word and 4-Word Key-LBs

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
A FEW CUPS	10	3.34	16.47	1,059.58	<0.001
ARE WE TO	11	4.99	18.11	1,059.71	<0.001
AS THE PROVERB	19	4.11	17.23	3.02	<0.001
AS THE PROVERB SAYS	18	2.76	15.88	2.94	<0.001
BOUND TO BE	20	19.81	32.94	1,060.58	<0.001
BUT MIND YOU	10	3.34	16.47	1,059.58	<0.001
CARRY ON LIKE	12	6.64	19.76	1,059.84	<0.001
COULD IT BE	10	3.34	16.47	1,059.58	<0.001
COUPLE OF DAYS	30	6.41	19.54	2.26	<0.001
DO SUCH A	10	3.34	16.47	1,059.58	<0.001
DO YOU EXPECT	11	4.99	18.11	1,059.71	<0.001
DOES IT MATTER	12	6.64	19.76	1,059.84	<0.001
DON'T YOU KNOW	11	4.99	18.11	1,059.71	<0.001
EVEN IF HE	10	3.34	16.47	1,059.58	<0.001
FOR A COUPLE	18	2.76	15.88	2.94	<0.001
FOR A COUPLE OF	18	2.76	15.88	2.94	<0.001
FOR A STROLL	12	6.64	19.76	1,059.84	<0.001
FOR A WHILE	29	7.26	20.38	2.40	<0.001
FOR NO REASON	19	18.17	31.29	1,060.50	<0.001
HAVE SUCH A	10	3.34	16.47	1,059.58	<0.001
HAVE THE SAME	12	6.64	19.76	1,059.84	<0.001
HIGH AND LOW	11	4.99	18.11	1,059.71	<0.001
HOW CAN I	36	13.68	26.81	2.52	<0.001
HOW CAN WE	25	28.05	41.17	1,060.90	<0.001
HOW CAN YOU	61	29.16	42.29	2.38	<0.001
HOW COULD I	20	5.47	18.59	3.09	<0.001
HOW IT IS	14	9.93	23.05	1,060.06	<0.001
HURRY UP AND	37	2.67	15.79	1.66	<0.001
I MEANT TO	17	14.87	27.99	1,060.34	<0.001
I'D NO IDEA	10	3.34	16.47	1,059.58	<0.001
IF NOT FOR	18	16.52	29.64	1,060.43	<0.001
IT'S NO USE	24	26.4	39.52	1,060.84	<0.001
IT'S NOT THAT	10	3.34	16.47	1,059.58	<0.001
JUST WHAT I	11	4.99	18.11	1,059.71	<0.001
MUCH THE BETTER	12	6.64	19.76	1,059.84	<0.001
MY ADVICE AND	12	6.64	19.76	1,059.84	<0.001
NOTHING BUT A	11	4.99	18.11	1,059.71	<0.001
ON LIKE THIS	20	19.81	32.94	1,060.58	<0.001
ON THE SLY	14	9.93	23.05	1,060.06	<0.001
SAY ONE WORD	10	3.34	16.47	1,059.58	<0.001
SO AS TO	30	19.64	32.76	3.68	<0.001
SO HOW CAN	13	8.28	21.41	1,059.96	<0.001
SO LONG AS	15	11.58	24.70	1,060.16	<0.001

(Continued)

Appendix A (Continued)

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
SO MUCH THE BETTER	12	6.64	19.76	1,059.84	<0.001
TAKE MY ADVICE	13	8.28	21.41	1,059.96	<0.001
TAKE MY ADVICE AND	11	4.99	18.11	1,059.71	<0.001
THE BLAME ON	11	4.99	18.11	1,059.71	<0.001
THIS CHANCE TO	11	4.99	18.11	1,059.71	<0.001
THIS IS JUST	15	11.58	24.70	1,060.16	<0.001
TO ASK FOR	26	3.68	16.80	2.25	<0.001
TO SEE TO	13	8.28	21.41	1,059.96	<0.001
TO SHOW MY	10	3.34	16.47	1,059.58	<0.001
WHAT DOES IT MATTER	12	6.64	19.76	1,059.84	<0.001
WHY NOT GO	11	4.99	18.11	1,059.71	<0.001
WHY SHOULD WE	15	11.58	24.70	1,060.16	<0.001
WOULDN'T THAT BE	20	19.81	32.94	1,060.58	<0.001
YOU DON'T UNDERSTAND	18	16.52	29.64	1,060.43	<0.001

Source: * Only Key-LBs loglikelihood > 6.63 (for pvalue < 0.01) are listed here.

Appendix B

Hawkes's 3-Word and 4-Word Key-LBs

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
A BIT AND	16	5.37	18.49	1,059.90	<0.001
A BIT BETTER	17	6.53	19.65	1,059.99	<0.001
A BIT OF	67	36.24	49.36	3.39	<0.001
A BIT TOO	15	4.22	17.34	1,059.81	<0.001
A FEW MINUTES	15	4.22	17.34	1,059.81	<0.001
A GOOD JOB	17	6.53	19.65	1,059.99	<0.001
A LOT OF	91	16.88	30.01	1.69	<0.001
A MATTER OF	33	3.66	16.78	2.37	<0.001
A QUESTION OF	16	5.37	18.49	1,059.90	<0.001
A THING LIKE THIS	14	3.06	16.18	1,059.71	<0.001
A WORD WITH	36	6.32	19.44	2.49	<0.001
ABLE TO SEE	14	3.06	16.18	1,059.71	<0.001
AND AFTER THAT	14	3.06	16.18	1,059.71	<0.001
AND GET IT	21	11.15	24.27	1,060.29	<0.001
AND I DON'T	18	7.68	20.81	1,060.07	<0.001
AND IN ANY	14	3.06	16.18	1,059.71	<0.001
AND IN ANY CASE	14	3.06	16.18	1,059.71	<0.001
ARE GOING TO	33	9.31	22.43	3.11	<0.001
ARE IN THE	14	3.06	16.18	1,059.71	<0.001
AS A MATTER	29	2.61	15.73	2.50	<0.001
AS A MATTER OF	29	2.61	15.73	2.50	<0.001
AWAY WITH IT	15	4.22	17.34	1,059.81	<0.001

Appendix B (Continued)

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
BE A BIT	20	9.99	23.12	1,060.22	<0.001
EXACTLY THE SAME	14	3.06	16.18	1,059.71	<0.001
FOR A BIT	35	8.13	21.25	2.78	<0.001
GET ON WITH	43	10.03	23.15	2.49	<0.001
GOING TO BE	47	19.75	32.87	3.20	<0.001
GOING TO DO	20	9.99	23.12	1,060.22	<0.001
GOT TO HEAR	14	3.06	16.18	1,059.71	<0.001
HAVEN'T GOT ANY	15	4.22	17.34	1,059.81	<0.001
HEAR ABOUT IT	14	3.06	16.18	1,059.71	<0.001
I AM AFRAID	31	22.71	35.83	1,060.86	<0.001
I AM NOT	18	7.68	20.81	1,060.07	<0.001
I AM SURE	28	19.24	32.36	1,060.71	<0.001
I DON'T KNOW WHY	19	8.84	21.96	1,060.15	<0.001
I DON'T THINK	53	16.44	29.56	2.57	<0.001
I HAVE BEEN	20	9.99	23.120	1,060.22	<0.001
I HOPE YOU	23	13.46	26.59	1,060.42	<0.001
I SHOULD HAVE	52	6.74	19.87	1.89	<0.001
I SHOULD LIKE	28	19.24	32.36	1,060.71	<0.001
I SHOULD LIKE TO	21	11.15	24.27	1,060.29	<0.001
I SHOULD THINK	19	8.84	21.96	1,060.15	<0.001
I THINK I	29	20.40	33.52	1,060.76	<0.001
I THINK IT	25	15.77	28.90	1,060.55	<0.001
I THINK IT'S	19	8.84	21.96	1,060.15	<0.001
I THINK WE	27	18.09	31.21	1,060.66	<0.001
I THINK WE OUGHT	14	3.06	16.18	1,059.71	<0.001
I THINK YOU	43	36.58	49.70	1,061.33	<0.001
I THOUGHT I'D	14	3.06	16.18	1,059.71	<0.001
I THOUGHT YOU	14	3.06	16.18	1,059.71	<0.001
I WONDER IF	15	4.22	17.34	1,059.81	<0.001
IF YOU ARE	32	5.34	18.46	2.65	<0.001
IF YOU ASK	23	13.46	26.59	1,060.42	<0.001
IF YOU ASK ME	20	9.99	23.12	1,060.22	<0.001
IF YOU WILL	25	15.77	28.90	1,060.55	<0.001
I'M AFRAID I	25	15.77	28.90	1,060.55	<0.001
I'M NOT SURPRISED	16	5.37	18.49	1,059.90	<0.001
I'M SURE YOU	19	8.84	21.96	1,060.15	<0.001
IN ANY CASE	69	23.09	36.21	2.43	<0.001
IS A VERY	20	9.99	23.12	1,060.22	<0.001
IS GOING TO BE	14	3.06	16.18	1,059.71	<0.001
IS SUCH A	16	5.37	18.49	1,059.90	<0.001
IS THE ONE	14	3.06	16.18	1,059.71	<0.001
IT MUST HAVE	14	3.06	16.18	1,059.71	<0.001
IT SEEMS THAT	23	13.46	26.59	1,060.42	<0.001
I'VE JUST BEEN	15	4.22	17.34	1,059.81	<0.001
KNOW WHAT THEY	16	5.37	18.49	1,059.90	<0.001
LOOK AT YOU	16	5.37	18.49	1,059.90	<0.001
ME ABOUT IT	19	8.84	21.96	1,060.15	<0.001
NOT GOING TO	38	10.97	24.09	2.89	<0.001

(Continued)

Appendix B (Continued)

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
OF THESE DAYS	17	6.53	19.65	1,059.99	<0.001
ONE OF THESE DAYS	17	6.53	19.65	1,059.99	<0.001
OUGHT NOT TO	15	4.22	17.34	1,059.81	<0.001
OUGHT TO BE	60	22.89	36.02	2.75	<0.001
OUT OF HERE	15	4.22	17.34	1,059.81	<0.001
SAY THAT I	14	3.06	16.18	1,059.71	<0.001
SHALL BE ABLE	16	5.37	18.49	1,059.90	<0.001
SHALL BE ABLE TO	16	5.37	18.49	1,059.90	<0.001
SHE HAS BEEN	16	5.37	18.49	1,059.90	<0.001
SHOULD LIKE TO	22	12.31	25.43	1,060.36	<0.001
SORT OF PERSON	18	7.68	20.81	1,060.07	<0.001
SORT OF THING	45	14.62	27.74	2.82	<0.001
SUPPOSED TO BE	20	9.99	23.12	1,060.22	<0.001
SURE TO BE	15	4.22	17.34	1,059.81	<0.001
TALK TO YOU	19	8.84	21.96	1,060.15	<0.001
TELL HER THAT	16	5.37	18.49	1,059.90	<0.001
TELL THEM THAT	18	7.68	20.81	1,060.07	<0.001
THAT I AM	27	18.09	31.21	1,060.66	<0.001
THAT I SHALL	15	4.22	17.34	1,059.81	<0.001
THAT I WAS	15	4.22	17.34	1,059.81	<0.001
THAT IF I	17	6.53	19.65	1,059.99	<0.001
THAT IF YOU	17	6.53	19.65	1,059.99	<0.001
THAT IT IS	14	3.06	16.18	1,059.71	<0.001
THAT SORT OF	34	26.18	39.30	1,060.99	<0.001
THAT SORT OF THING	18	7.68	20.81	1,060.07	<0.001
THAT THEY ARE	18	7.68	20.81	1,060.07	<0.001
THAT WE SHOULD	15	4.22	17.34	1,059.81	<0.001
THAT YOU ARE	30	6.39	19.51	2.97	<0.001
THAT YOU HAVE	37	7.22	20.34	2.53	<0.001
THE WAY I	17	6.53	19.65	1,059.99	<0.001
THERE IS A	27	3.52	16.64	2.82	<0.001
THERE WOULD BE	14	3.060	16.18	1,059.71	<0.001
THING LIKE THAT	17	6.53	19.65	1,059.99	<0.001
THINGS LIKE THAT	19	8.84	21.96	1,060.15	<0.001
THINK OF IT	15	4.22	17.34	1,059.81	<0.001
THINK WE OUGHT	19	8.84	21.96	1,060.15	<0.001
THINK WE OUGHT TO	19	8.84	21.96	1,060.15	<0.001
THINK YOU OUGHT	18	7.68	20.81	1,060.07	<0.001
THINK YOU OUGHT TO	17	6.53	19.65	1,059.99	<0.001
TO DO IS	22	12.31	25.43	1,060.36	<0.001
TO DO SOMETHING	14	3.06	16.18	1,059.71	<0.001
TO HAVE BEEN	17	6.53	19.65	1,059.99	<0.001
TO HEAR ABOUT	16	5.37	18.49	1,059.90	<0.001
TO TALK TO YOU	14	3.06	16.18	1,059.71	<0.001
TO TELL ME	35	5.42	18.55	2.45	<0.001
TO THINK THAT	16	5.37	18.49	1,059.90	<0.001
TO YOU ABOUT	14	3.06	16.18	1,059.71	<0.001
WANT TO GO	30	21.55	34.68	1,060.81	<0.001

Appendix B (Continued)

<i>Key-LBs</i>	<i>Freq.</i>	<i>BIC</i>	<i>Log-Likelihood</i>	<i>Log-Ratio</i>	<i>P-Value</i>
WE OUGHT TO	49	10.45	23.57	2.26	<0.001
WHAT IT IS	25	15.77	28.90	1,060.55	<0.001
WHAT YOU ARE	16	5.37	18.49	1,059.90	<0.001
WHAT YOU HAVE	20	9.99	23.12	1,060.22	<0.001
WHEN YOU ARE	27	18.09	31.21	1,060.66	<0.001
WHILE YOU ARE	14	3.06	16.18	1,059.71	<0.001
YOU ARE GOING	17	6.53	19.65	1,059.99	<0.001
YOU ARE GOING TO	15	4.22	17.34	1,059.81	<0.001
YOU ARE NOT	20	9.99	23.12	1,060.22	<0.001
YOU ARE TOO	17	6.53	19.65	1,059.99	<0.001
YOU DON'T NEED TO	14	3.06	16.18	1,059.71	<0.001
YOU KNOW WHAT	35	3.08	16.20	2.19	<0.001
YOU OUGHT TO	100	15.51	28.64	1.53	<0.001
YOU OUGHT TO BE	22	12.31	25.43	1,060.36	<0.001
YOU THINK THAT	16	5.37	18.49	1,059.90	<0.001
YOU WILL BE	28	4.47	17.59	2.87	<0.001
YOU WOULD BE	17	6.53	19.65	1,059.99	<0.001
YOU WOULD HAVE	14	3.06	16.18	1,059.71	<0.001
YOU'LL BE ABLE	15	4.22	17.34	1,059.81	<0.001
YOU'LL BE ABLE TO	15	4.22	17.34	1,059.81	<0.001
YOU'VE GOT TO	16	5.37	18.49	1,059.90	<0.001

Source: * Only Key-LBs loglikelihood > 6.63 (for pvalue < 0.01) are listed here.