Are Translated Chinese *Wuxia* Fiction and Western Heroic Literature Similar? A Stylometric Analysis Based on Stylistic Panoramas

Kan Wu and Dechao Li

1 Introduction

Wuxia, or Chinese martial arts fiction, is a traditional genre of Eastern heroic literature that originated from unique historical and cultural contexts during China's Warring States Period (475–221 BC) (Huang 2018, 152). Previous research on *Wuxia* (Flannery 2012; Vander Elst 2017; Keulemans 2020) has attempted to compare this type of Chinese heroic literature with Western chivalric stories and heroic fantasies – two subgenres of heroic literature deriving from a medieval background (Honegger 2010, 61). Those studies have demonstrated that *Wuxia* could be very different from the two Western subgenres in terms of cultural val-ues, religious belief, and above all, worldviews, even though they share the heroic theme. Wu and Li, however, discovered that when readers read a *Wuxia* translation, they sometimes experience a déjà vu–like reminder of chivalric stories or heroic fantasies (Wu and Li 2018, 102–3). This raises the question as to whether there are any possible stylistic connections between heroic literature in the East and that in the West. An examination of such stylistic connections may give us clues about the current reception of translated *Wuxia* and is hence our first research objective. To conduct the investigation, we turn to Stylometry – the statistical analysis of literary styles (Holmes 1998, 111) – for methodological support.

Existing stylometric research on (translated) texts of varied genres has employed a number of stylistic indices at such linguistic levels as characters (Daelemans 2013; Eder et al. 2016), words/lexes (Jones and Nulty 2019; Melka and Místecký 2020), *n*-grams/clusters (Mastropierro 2018; Valencia et al. 2019), sentences and paragraphs (Rong et al. 2006), tones and rimes (Hou and Huang 2020), and their combinations (Brocardo et al. 2014; Liu and Xiao 2020). These studies have unveiled the features of (translated) texts from multiple stylometric perspectives and reveal their most noticeable stylistic features. Nevertheless, they are not without methodological limitations. For one thing, they lack a panoramic view of different stylistic features in a text; for another, they do not always explain the selection criteria for the stylistic indices to be investigated. We believe that both the adoption of a panoramic view and the justification for selection criteria are vital in stylometric analyses because the former makes results more comparable

across research and the latter reveals intended functions associated with the chosen indices. Therefore, the second objective of the present study is to introduce the stylistic panorama, a novel concept proposed to describe the stylistic profile of a (translated) text in a relatively holistic and functional way.

To achieve the two research objectives, we formulate the following research questions:

RQ1: To what extent are translated *Wuxia* fiction similar to and/or different from Western chivalric stories and heroic fantasies, from a stylometric perspective?

RQ2: In what ways could such similarities and/or differences reveal current reception of English translations of *Wuxia*?

RQ3: How do the findings shed light on the use of stylistic panoramas in stylometric analyses?

Whereas an investigation of RQ1 is expected to contribute to the hypothesized stylistic connection(s) between translated *Wuxia* and the two subgenres of Western heroic literature, RQ2 contributes to practical outcomes for target readers of translated *Wuxia* and RQ3 explores theoretical implications for stylometric research into literary translation, specifically Chinese to English translation on *Wuxia* fiction.

2 Stylistic Panorama as a Stylistic Profile of a (Translated) Text

A stylistic panorama is defined in this research as a relatively complete stylistic profile that is based on a set of functionally related stylistic indices at multiple linguistic levels and is intended to satisfy certain research purpose(s). Theoretical significances of the concept can be observed in both stylometric analyses and translation studies. In stylometric analyses, the stylistic panorama emphasizes the combined efforts of varied stylistic features when exploring a (translated) text. For translation studies, the panorama values the stylometric approach in its methodological design. In other words, the concept can be a bridge connecting empirical translation studies with stylometric analyses: it lends viable methodological support to empirical translation studies, enriching the research scope of stylometric analyses.

To generate the stylistic panorama of a text, the first step is to choose proper stylistic indices. Such a choice is a varied process that depends largely on research needs. Considering that the research aim here is to conduct an overall stylistic comparison between the translated *Wuxia* and the two Western subgenres of heroic literature, we tend to select stylistic indices that could cover wide linguistic levels (words, word sequences, sentences, paragraphs, etc.) to capture basic features of this type of literature. Hence, the indices we choose for this research are the average word length (AWL), the dispersion of word lengths (DWL), the moving-average type-token ratio (MATTR), the verb-adjective ratio (VAR), the average sentence length (ASL), the dispersion of sentence lengths (DSL), the average paragraph length (APL), the most frequent words (MFWs), and the most frequent word

sequences (MFWSs). Next, we perform multivariate analyses on those indices to produce interpretable stylistic panoramas. This is because multivariate analyses are powerful enough to account for holistic stylistic similarities between the translated *Wuxia* and the two Western subgenres, as well as being amenable to varied sample sizes. Finally, for easier investigation, we categorize the stylistic panoramas according to the functions of the selected indices. Hence, two types of stylistic panoramas emerge in this work: one based on formal indices, and the other based on word and word sequence frequencies (MFWs and MFWSs, respectively).

The stylistic panorama based on formal indices describes formal features of a text. It is founded on such indices as AWL, DWL, MATTR, VAR, ASL, DSL, and APL, which are connected to comparatively small sample sizes (ca. 126 in total) in this research. At the word level, the AWL and DWL are selected to show the orthographical complexity of the heroic texts, and the MATTR and VAR are chosen to demonstrate the lexical richness of these texts. This ensures a breadth of features types under classification. It is worth stressing that, out of many possible indices depicting lexical richness, we prefer the MATTR and VAR for practical reasons. The MATTR is favored because "it takes into account all possible segmentation of the text" (Březina 2018, 58) and is thus believed to capture features of vocabulary richness of a heroic text. The VAR is adopted because it reflects the lexical richness of the heroic text – a genre that is likely to contain a multitude of verbs and adjectives depicting kung fu fighting scenes (Wu and Li 2018, 102). For sentences and paragraphs, the ASL and APL are chosen to partly reveal the typological complexity, and above all, the readability, of the heroic texts, which we believe have certain connections to the reception of translated *Wuxia*. In addition, we select the DSL because computations of sentence dispersion can unveil the rhythm and likewise the readability of a heroic text: a lower dispersion value suggests a higher level of repetitiveness in a text, and vice versa.

The stylistic panorama built on MFWs/MFWSs associates with lexes and their sequences in a text. In this study, the two stylistic indices are related to larger sample sizes (ca. 3,000 in total). The MFWs index is adopted because it has a long tradition of being an efficient classifier to distinguish stylistic features of one text from those of another (Burrows 2002). Meanwhile, word sequences (or *n*-grams, the contiguous sequences of *n* items of words within a given text or speech) have been widely applied in computational linguistics and information sciences for predictive or attributive purposes (Broder et al. 1997, 1157). In some quantitative linguistic/translation research (Rybicki 2012; Mastropierro 2018), frequency patterns of the MFWs and/or the MFWSs have been compared in an effort to assess their overall similarities and/or differences between texts. The MFWs and the MFWSs are often determined by the proportion of each word or word sequence in a text and are presented in parallel lists. The present study confines the scope of MFWSs to 2-grams and 3-grams, because they are the most common word sequences in (translated) literary texts (see Burrows 2002; Rybicki 2012; Mastropierro 2018). Furthermore, the choice of 2- and/or 3-grams as the MFWSs is further justified by the possibility that heroic literature is more likely to contain short phrases (of two to three words) that depict the quick action and short dialogue sequences of fighting scenes in the stories (Wu and Li 2018, 97).

Overall, the stylistic panorama is a concept that attempts to bind selected stylistic indices in texts together in a relatively holistic and functional way. In other words, when measuring the stylistic features of a text, the indices are no longer examined in isolation but are instead measured in a more comprehensive and interrelated way in line with specific research needs. For this work, because one objective is to explore potential stylistic connection(s) between the heroic literature of the East and that of the West, the selected stylistic indices and their resulting panoramas are expected to meet this goal.

3 Data and Methodology

3.1 Data and Corpora

The *Wuxia* novels[2] used in this study are the English translations of six different works by Louis Cha, a renowned Hong Kong *Wuxia* novelist. We choose those six works because they are the only English *Wuxia* translations published at the time of writing. The novels are translated by five translators who are experienced in rendering Chinese *Wuxia* fiction into English: Minford, Earnshaw, and Mok are sinologists dedicated to Chinese literary translation, whereas Holmwood and Chang are new generation translators who are interested in the dissemination of Chinese *Wuxia* overseas (Wu and Li 2018, 95).

For the Western heroic literature, the selected works[3] are chivalric stories translated into modern English and heroic fantasies written in modern English. We choose these works for a balance between comparability and representativeness. First, the choice of works in modern English is meant to ensure linguistic comparability across the subgenres and, as a consequence, facilitate where the style of the *Wuxia* translations could be comparatively located. Second, we expect the great popularity (based on Amazon/Goodreads ratings) and diverse source languages (i.e., English, German, Spanish) of the selected translated works would increase the representativeness in the two subgenres of Western heroic literature.

Details of all the works used in this research are summarized in Tables 1 through 3, including year of publication details regarding the versions used in the study. Comparability is additionally enhanced through similar token sizes (ca. 3 million) in texts across the subgenres. Representativeness is further captured

Table 1 Details of the Translated *Wuxia* Stories

Translated Wuxia *Stories*	*Year*	*Translator*	*Token Size*
Fox Volant of the Snowy Mountain	1993	Olivia Mok	120,613
The Book and Sword	2004	Graham Earnshaw	192,439
The Deer and the Cauldron	1997	John Minford	617,949
A Hero Born	2018	Anna Holmwood	127,123
A Bond Undone	2019	Gigi Chang	160,851
A Snake Lies Waiting	2020	Gigi Chang	140,539
Total Size			1,359,514

Table 2 Details of the Chivalric Stories

Chivalric Stories	Year	Author/Editor	Token Size
Don Quixote (translated from Spanish)	2003	M. De Cervantes	402,546
Ivanhoe	2005	W. Scott	182,732
Parzival (translated from German)	1980	W. Von Eschenbach	154,902
In the Days of Chivalry	2004	E. Everett-Green	153,079
Heroes and Heroines of Chivalry	2004	W. Patten	135,866
Castles, Knights, and Chivalry	2015	Kaufman et al.	351,311
Total Size			1,380,436

Table 3 Details of the Heroic Fantasies

Heroic Fantasies	Year	Author	Token Size
The Lord of the Rings 1–2	1954–1955	J. R. R. Tolkien	188,361
A Song of Ice and Fire 1	1996	G. R. R. Martin	293,856
Conan the Barbarian	1954	R. E. Howard	104,972
Wheel of Time 1	1990	R. Jordan	319,124
The Chronicles of Amber 1–4	1970–1976	R. Zelazny	239,940
The Chronicles of Narnia 1–5	1950–1954	C. S. Lewis	233,369
Total Size			1,379,622

through the inclusion of both earlier works (before the 1850s) and modern collections (the 1850s or later) for the chivalric stories.

3.2 Calculations and Algorithms

Whereas the average word, sentence, and paragraph length (AWL/ASL/APL) values are retrieved from the outputs of *Wordsmith 6.0* (Scott 2012), the most frequent word and word sequences (MFWs/MFWSs) and their proportions in the texts are obtained from *Intelligent Archive 3.0* (Craig 2018) as parallel lists. The dispersion of word and sentence lengths (DWL/DSL), moving-average type-token ratio (MATTR), and verb-adjective ratio (VAR) values, on the other hand, are calculated according to formulas 1 through 3, as follows.

$$SD = \sqrt{\frac{1}{n}\sum (X_i - X_0)^2} \qquad (1)$$

$$VAR = \frac{verbs}{verbs + adjectives} \qquad (2)$$

$$MATTR = \frac{\sum_{i=1}^{N-L} V_i}{L(N-L+1)} \qquad (3)$$

In formula 1, standard deviation (*SD*) gives the statistical expression of the DWL/DSL (Liu and Xiao 2020, 35), where *n* is the number of words/sentences in the text, X_i is the length of a single word/sentence, and X_0 is the average word/sentence length.

In formula 2, *verbs* and *adjectives* represent the total numbers of verbs and adjectives in a text, respectively. *Stanford Tagger 4.2.0* (Stanford NLP Group 2021) is used to obtain the numbers of verbs/adjectives in the texts through POS annotation.

In formula 3, N is the total length of a text, L is the selected length of a text chunk, and V_i is the number of types in the text chunk. To operationalize, the chunk size in the *MATTR* calculation is set to 500, a setting that has previously produced reliable results (see Covington and McFall 2010; Kettunen 2014).

Data related to the formal indices are normalized according to formula 4, while the Euclidean distance between two stylistic panoramas built on the MFWs and MFWSs are calculated based on formula 5, as follows:

$$Y = \frac{X}{\sqrt{\sum_{k=1}^{n} X^2}} \tag{4}$$

$$AB = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{5}$$

Formula 4 is the L2 regularization. Y represents the normalized value, X is the original value, and $\sum_{k=1}^{n} X^2$ is the sum of the squares of all original values in a dataset. A benefit of L2 regularization is that it handles the problem of overfitting well when the dataset is relatively small.

Formula 5 is used to calculate the Euclidean distance between two stylistic panoramas built on the MFWs/MFWSs, where x and y are the coordinates of each panorama and AB is the distance between the panoramas.

For the algorithms used to analyze the stylistic indices, the study employs hierarchical cluster analysis (HCA) and principal component analysis (PCA) for the reasons elaborated in Section 2. HCA is an unsupervised machine learning procedure that groups similar objects into a category (Christopher et al. 2008, 321) and is amenable to small sample sizes with the $O = 2^k$ principle[4] (Formann 1984). By contrast, PCA is based on the idea of reducing a substantial number of variables into a smaller number of transformed variables (Manly 2016, 103) and can thus help measure the overall similarities and/or differences between (translated) texts.

3.3 Analytic Steps

The stylometric analyses of the selected works require several steps. First, the data are cleaned by removing all the paratexts (prefaces, appendices, footnotes, etc.) as a preparatory step to minimize any possible influence of these on the results. Next, we retrieve the stylistic indices of the raw data from each text in the research, using software computation and manual calculations. Third, we build stylistic panoramas with the retrieved data, by using *R 4.03* to perform HCA and PCA separately on the normalized data. This step reveals the potential stylistic connections between the translated *Wuxia* and the two subgenres of Western heroic literature. Finally, the results are interpreted in light of the current reception of some *Wuxia* translations, with reflection on theoretical implications on the use of stylistic panoramas in stylometric analyses.

4 Results

4.1 Stylistic Panoramas Based on Formal Indices

Table 4 summarizes the raw values of the seven indices (AWL, DWL, MATTR, VAR, ASL, DSL, and APL). These raw values suggest that the translated *Wuxia* works and the two subgenres of Western heroic literature are stylistically similar

Table 4 Statistics of the Stylistic Indices across the Genres

Subgenre	Fiction	Word				Sentence		Para.
		AWL	DWL	MATTR	VAR	ASL	DSL	APL
Chivalric Stories	Don Quixote	4.27	2.21	0.50	0.77	48.17	26.56	71.83
	Ivanhoe	4.46	2.35	0.55	0.71	33.07	18.29	56.57
	Parzival	4.22	2.14	0.54	0.76	19.62	9.68	75.78
	In the Days of Chivalry	4.25	2.09	0.51	0.71	25.10	18.79	82.88
	Heroes and Heroines of Chivalry	4.07	1.93	0.46	0.80	23.12	14.53	67.60
	Castles, Knights, and Chivalry	4.32	2.14	0.53	0.81	11.75	6.72	29.38
Heroic Fantasies	The Lord of the Rings, 1–2	4.09	1.92	0.50	0.75	14.63	10.14	45.70
	A Song of Ice and Fire, 1	4.16	1.93	0.52	0.77	11.82	8.49	33.60
	Conan the Barbarian	4.44	2.23	0.55	0.72	26.43	10.25	49.68
	Wheel of Time, 1	4.26	2.08	0.53	0.79	18.75	8.71	48.26
	The Chronicles of Amber, 1–4	4.09	2.15	0.50	0.80	11.12	8.46	38.89
	The Chronicles of Narnia, 1–5	4.14	1.99	0.50	0.76	17.47	11.51	31.28
Translated Wuxia Fiction	Fox Volant of the Snowy Mountain	4.56	2.36	0.53	0.75	19.73	10.76	46.84
	The Book and Sword	4.29	2.10	0.50	0.80	13.07	8.20	33.31
	The Deer and the Cauldron	4.38	2.26	0.52	0.76	39.18	14.84	26.00
	A Hero Born	4.39	2.20	0.55	0.78	22.82	7.63	33.70
	A Bond Undone	4.47	2.27	0.56	0.77	21.67	7.40	29.06
	A Snake Lies Waiting	4.34	2.13	0.54	0.79	17.41	7.73	31.27

Chivalric Stories

AWL 1.41
DWL 1.42
MATTR 1.40
VAR 1.40
ASL 1.58
DSL 1.76
APL 1.83

Heroic Fantasies

AWL 1.38
DWL 1.35
MATTR 1.40
VAR 1.41
ASL 0.99
DSL 1.07
APL 1.18

***Wuxia* Translations**

AWL 1.45
DWL 1.47
MATTR 1.44
VAR 1.43
ASL 1.32
DSL 1.05
APL 0.96

AWL = Average Word Length
DWL = Dispersion of Word Lengths
MATTR = Moving-Average Type-token Ratio
VAR = Verb–adjective Ratio
ASL = Average Sentence Length
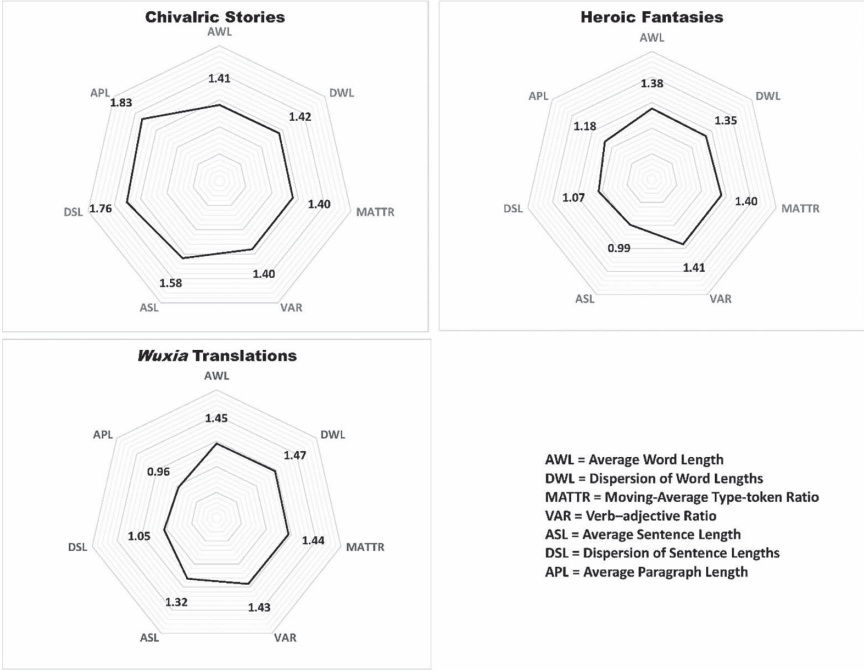DSL = Dispersion of Sentence Lengths
APL = Average Paragraph Length

Figure 1 Stylistic panoramas of the three subgenres, from a global view.

at the word level but divergent in terms of sentences and paragraphs. This trend could reflect multiple factors, from different literary norms to translatorial/authorial idiosyncrasies, thus giving target readers of the three subgenres varied reading experiences. To further probe the stylistic connections, we examine the stylistic panoramas formed by these formal indices from both global and local perspectives. While the examination at the global level attempts to capture the panorama of each subgenre, the investigation of local perspectives compares the panoramas of single works in the three subgenres.

4.1.1 Stylistic Panoramas from a Global Perspective

A global comparison of the stylistic panoramas between the *Wuxia* translations and the chivalric stories/heroic fantasies is meant to locate the stylistic features of the *Wuxia* translations in relation to those of the Western heroic literature. Before that comparison, however, the formation of a stylistic panorama requires that we normalize the data under each column in Table 4 for better data comparabil-ity across the indices, using formula 4 (explained in Section 3.2). Then, we tally the normalized data for each stylistic index to obtain a total value, which is the statistical ingredient of the stylistic panorama at this global level. The

stylistic panoramas that are formed in each of the three subgenres are presented in Figure 1 as radar charts. They reveal several stylistic patterns.

At the word level, a conspicuous pattern is that the translated *Wuxia* has higher normalized values than the two subgenres of Western heroic literature. This is easily detected in the moving-average type-token ratio (MATTR) values, in which the normalized value for the translated *Wuxia* novels is 1.44 and the values for the chivalric stories and heroic fantasies are both 1.40. That result may point to the use of a comparatively richer vocabulary in the translated *Wuxia*. In addition, the highest dispersion of word lengths (DWL) value is 1.47, demonstrating that the word length in the translated *Wuxia* novels is generally more variable than that in the two Western subgenres. Likewise, a normalized average word length (AWL) of the translated *Wuxia* stories at 1.45 indicates the use of longer and more complex words. One potential reason for such a tendency would be a shared preference by the *Wuxia* translators to use longer and more complex words for explanative renditions in English, because the original Chinese versions by Cha contain many culturally loaded *Wuxia* concepts. Also, the highest verb-adjective ratio (VAR) value is found in the *Wuxia* translations, which reveals that they use more verbs than their Western counterparts do. This has numerous stylistic effects, including a more vivid reading experience to the target readers.

At the sentence and paragraph levels, there are two noticeable patterns. First, the chivalric stories have higher values overall. For instance, the normalized ASL (average sentence length) and DSL (dispersal sentence lengths) values for the chivalric stories, being 1.58 and 1.76, respectively, are much higher than those of the other two subgenres. This could mean that most sentences in the chivalric stories are more complex and more varied than those in the other two subgenres. Similarly, the normalized APL (average paragraph length) for the chivalric stories is 1.83, a value far greater than those for the translated *Wuxia* novels and the heroic fantasies. A possible explanation would be the use of different literary norms between the subgenres of heroic literature. Close reading of the selected chivalric stories reveals that they tend to pack more sentences into a single paragraph, thus often pushing their APLs to higher values and presenting readers with longer paragraphs. All these may suggest that, as an old form of heroic literature, chivalric stories are stylistically more complex and more varied in terms of sentences and paragraphs than the other two subgenres. Second, the ASL and APL values in the *Wuxia* translations and the heroic fantasies demonstrate a less consistent but meaningful trend. Whereas the *Wuxia* translations have a higher ASL value of 1.32 yet a lower APL value of 0.96, the heroic fantasies bear a greater APL value of 1.18 but a lower ASL value of 0.99. This indicates that the *Wuxia* translations may have many short paragraphs built on relatively longer sentences – an important stylistic feature which distinguishes the *Wuxia* translations from the heroic fantasies.

4.1.2 Stylistic Panoramas from a Local Perspective

A local comparison is meant to investigate whether the stylistic panorama pattern deriving from the global comparison would vary when we make comparisons

Cluster Dendrogram

Figure 2 Cluster dendrogram of the HCA-based stylistic panoramas.

across the three subgenres based on single heroic works. The study holds that such a local comparison is necessary because it reveals how such extrastylometric factors as translatorial motivations and publication years might affect the stylistic patterns. To make that comparison, we use the normalized data and resort to HCA to produce stylistic panoramas. The HCA-based stylistic panoramas are produced according to the Euclidean distance between the texts and using the maximum distance method in computation. The output from *R 4.03* is presented in Figure 2 as a cluster dendrogram: the horizontal axis shows the titles of the 18 selected works, and the vertical axis records the divergence of clusters. The results show that the *Wuxia* translations differ in important ways from their Western counterparts and from each other.

The *Wuxia* translations published in more recent years (i.e., 2018, 2019, 2020) appear to form a distinct category that is not only stylistically different from the selected chivalric stories and heroic fantasies but also differs from the rest of *Wuxia* translations in the dendrogram. This trend is clearly shown by the stylistic panoramas of *A Snake Lies Waiting*, *A Bond Undone*, and *A Hero Born*, which indicate that the three recent *Wuxia* translations may share certain similarities in terms of language use. In that regard, the study demonstrates that a short publication span and close translatorial cooperation could be two factors that shape this stylistic similarity among the three *Wuxia* translations. Notably, the three translations were consecutively published in 2018, 2019, and 2020, and the two translators, Holmwood and Chang, had worked with a kindred spirit

in their translations in an effort to "see the foreign interest in China and its culture" (Mei 2019).

By contrast, the *Wuxia* translations from earlier periods (i.e., the 1990s and the 2000s) are stylistically diverse from each other but close to some works of the heroic fantasies. It is noteworthy in Figure 2 that the stylistic panoramas of *The Book and Sword*, *Fox Volant of the Snowy Mountain*, and *The Deer and the Cauldron* are comparatively divergent from each other, despite that the three translations are produced within a relatively short span of 11 years. Instead, the stylistic panoramas of *Fox Volant of the Snowy Mountain* and *The Book and Sword* are respectively similar to those of *Wheel of Time 1* and *A Song of Ice and Fire 1*, two works of heroic fantasies published likewise in the 1990s. The stylistic panorama of *The Deer and the Cauldron* is another story: it differs from that of the other works in the three subgenres and forms an independent category, which places the *Wuxia* translation somewhere between the chivalric stories and the heroic fantasies in terms of style. That uniqueness may indicate variations in the language use and thus the readability of the three *Wuxia* translations – a scenario that we tend to associate with different translatorial motivations. For example, the motivation behind the translation of *Fox Volant of the Snowy Mountain* may have been "promoting Chinese martial arts cultures overseas" (Wu and Li 2018, 100), while that behind the production of *The Deer and the Cauldron* may have been "winning overseas readership," and by juxtaposition, the motivation underlying the translation of *The Book and Sword* may have been "learning the Chinese language/culture" (Wu and Li 2018, 101). As a result, the translators of the three earlier *Wuxia* translations are likely to have used different vocabularies (more verbs, different culturally loaded words, etc.) and to have varied the sentence/paragraph lengths in their translations.

4.2 Stylistic Panoramas Based on MFWs/MFWSs

The stylistic panoramas based on the formal indices show that the *Wuxia* translations are not only largely different from the works of the two Western subgenres but also divergent from each other. Therefore, the study seeks to determine whether such stylistic patterns would remain or change when we compare the panoramas on the basis of the most frequent words (MFWs) and most frequent word sequences (MFWSs) through the PCA.

4.2.1 Parallel Lists of MFWs/MFWSs

Parallel lists of the MFWs/MFWSs in the texts (shown partially in Tables 5–7[5]) include information about the MFWs/MFWSs and the titles of each work. The remaining rows list the words/word sequences and their proportions in each work. To facilitate comparability, proportions are analyzed rather than raw frequencies, given the different lengths of the works. Finally, contractions such as "I've," "he'll," and "you'd" in all works are analyzed as single words rather than as separate words. Such a decision contrasts with previous studies (see Tognini-Bonelli 2001; Laviosa 2002; Mastropierro 2018), which treated those forms as separate

Fiction	the	and	to	of	a	he	his	in	was	i	[...]
A Bond Undone	4.994	2.149	2.473	1.937	1.854	1.77	1.771	1.125	1.178	1.038	[...]
A Hero Born	5.483	2.747	2.592	1.935	1.759	1.732	1.715	1.213	1.347	0.546	[...]
A Snake Lies Waiting	5.492	2.491	2.788	1.999	1.831	1.816	1.612	1.175	1.532	0.681	[...]
A Song of Ice and Fire 1	6.038	3.015	2.234	1.994	2.185	1.756	1.732	1.102	1.326	0.95	[...]
Amber Chronicles 1-4	4.623	2.917	2.663	2.25	2.154	1.587	0.683	1.128	1.547	4.531	[...]
Castles Knights and Chivalry	4.799	2.367	2.861	1.647	1.912	2.397	2.286	1.169	1.058	1.001	[...]
Chronicles of Narnia 1-5	5.4	4.076	2.253	2.188	2.269	1.253	0.717	1.273	1.414	1.066	[...]
Conan the Barbarian	7.316	3.257	1.723	2.869	2.659	1.551	1.657	1.635	1.175	0.7	[...]
Don Quixote	5.166	4.651	3.09	2.475	1.633	1.514	1.087	1.551	0.886	1.489	[...]
Fox Volant of the Snowy Mountain	6.93	2.421	2.975	2.243	2.058	1.39	1.871	1.549	1.318	1.008	[...]
Heroes and Heroines of Chivalry	5.232	4.841	2.871	2.302	1.65	2.176	1.683	1.387	1.199	1.449	[...]
In the Days of Chivalry	6.049	3.774	2.883	3.407	1.605	1.518	1.55	1.664	1.28	1.007	[...]
Ivanhoe	7.023	3.247	2.819	4.076	1.902	0.969	1.452	1.425	0.793	1.151	[...]
Lord of The Rings 1-2	6.24	4.016	2.107	2.701	1.984	1.597	0.839	1.577	1.31	1.446	[...]
Parzival	3.872	2.567	2.956	2.236	1.754	1.483	1.566	1.396	1.405	1.446	[...]
The Book and Sword	6.17	3.547	2.813	2.102	2.025	1.893	1.123	1.201	1.108	0.809	[...]
The Deer and the Cauldron	5.773	2.606	2.907	2.63	2.142	1.8	1.317	1.384	1.398	0.688	[...]
Wheel of Time 1	6.282	2.613	2.365	2.03	2.18	2.128	1.483	1.319	1.232	0.822	[...]

Figure 3 Sample parallel list for the MFWs in the selected works.

Fiction	of.the	in.the	to.the	and.the	on.the	it.was	he.was	he.had	at.the	from.the	[...]
A Bond Undone	0.521	0.319	0.24	0.107	0.174	0.133	0.184	0.223	0.164	0.136	[...]
A Hero Born	0.606	0.347	0.308	0.14	0.207	0.125	0.192	0.151	0.173	0.133	[...]
A Snake Lies Waiting	0.608	0.296	0.291	0.133	0.197	0.187	0.207	0.185	0.168	0.134	[...]
A Song of Ice and Fire 1	0.519	0.378	0.243	0.245	0.191	0.195	0.182	0.141	0.139	0.161	[...]
Amber Chronicles 1-4	0.453	0.283	0.276	0.181	0.168	0.304	0.153	0.133	0.119	0.095	[...]
Castles Knights and Chivalry	0.372	0.294	0.284	0.092	0.184	0.134	0.146	0.177	0.162	0.14	[...]
Chronicles of Narnia 1-5	0.576	0.376	0.268	0.321	0.245	0.332	0.161	0.143	0.175	0.104	[...]
Conan the Barbarian	0.834	0.565	0.296	0.273	0.298	0.178	0.155	0.121	0.202	0.251	[...]
Don Quixote	0.538	0.392	0.268	0.28	0.138	0.11	0.148	0.155	0.111	0.09	[...]
Fox Volant of the Snowy Mountain	0.839	0.363	0.377	0.127	0.348	0.113	0.168	0.117	0.202	0.165	[...]
Heroes and Heroines of Chivalry	0.558	0.327	0.373	0.29	0.116	0.124	0.197	0.171	0.133	0.092	[...]
In the Days of Chivalry	0.938	0.427	0.336	0.257	0.056	0.198	0.142	0.218	0.107	0.151	[...]
Ivanhoe	0.999	0.376	0.429	0.203	0.164	0.104	0.072	0.093	0.148	0.172	[...]
Lord of The Rings 1-2	0.807	0.598	0.29	0.313	0.241	0.22	0.147	0.113	0.139	0.168	[...]
Parzival	0.268	0.198	0.253	0.123	0.14	0.182	0.171	0.147	0.093	0.095	[...]
The Book and Sword	0.698	0.335	0.361	0.28	0.238	0.122	0.142	0.103	0.175	0.129	[...]
The Deer and the Cauldron	0.695	0.405	0.317	0.171	0.232	0.19	0.232	0.181	0.129	0.143	[...]
Wheel of Time 1	0.597	0.472	0.235	0.258	0.217	0.223	0.2	0.198	0.217	0.188	[...]

Figure 4 Sample parallel list for the MFWSs (2-grams) in the selected works.

Fiction	one.of.the	out.of.the	there.was.i	he.did.not	it.was.a	as.soon.as	there.was.i	that.he.wa	it.would.be	for.a.mome	[...]
A Bond Undone	0.022	0.037	0.007	0.03	0.012	0.002	0.017	0.01	0.028	0.011	[...]
A Hero Born	0.058	0.029	0.007	0.016	0.015	0.01	0.017	0.007	0.011	0.004	[...]
A Snake Lies Waiting	0.036	0.033	0.015	0.031	0.019	0.022	0.035	0.019	0.023	0.01	[...]
A Song of Ice and Fire 1	0.017	0.017	0.015	0.024	0.017	0.007	0.021	0.009	0.012	0.022	[...]
Amber Chronicles 1-4	0.023	0.023	0.036	0.009	0.037	0.02	0.021	0.021	0.025	0.021	[...]
Castles Knights and Chivalry	0.024	0.03	0.009	0.013	0.015	0.011	0.013	0.005	0.009	0.028	[...]
Chronicles of Narnia 1-5	0.033	0.06	0.046	0.012	0.051	0.033	0.032	0.013	0.02	0.021	[...]
Conan the Barbarian	0.022	0.041	0.03	0.041	0.028	0.001	0.02	0.022	0.006	0.003	[...]
Don Quixote	0.043	0.012	0.006	0.033	0.009	0.044	0.01	0.02	0.024	0.002	[...]
Fox Volant of the Snowy Mountain	0.02	0.027	0.004	0.007	0.014	0.008	0.01	0.014	0.008	0.004	[...]
Heroes and Heroines of Chivalry	0.038	0.02	0.011	0.012	0.007	0.018	0.007	0.034	0.01	0.003	[...]
In the Days of Chivalry	0.025	0.012	0.014	0.013	0.02	0.001	0.018	0.016	0.014	0.02	[...]
Ivanhoe	0.023	0.015	0.01	0.003	0.007	0.008	0.004	0.006	0.004	0.001	[...]
Lord of The Rings 1-2	0.021	0.063	0.073	0.017	0.02	0.021	0.021	0.022	0.014	0.024	[...]
Parzival	0.007	0.001	0.019	0.016	0.012	0.005	0.016	0.019	0.014	0	[...]
The Book and Sword	0.077	0.081	0.024	0.01	0.015	0.026	0.013	0.011	0.026	0.029	[...]
The Deer and the Cauldron	0.08	0.031	0.036	0.006	0.019	0.017	0.015	0.028	0.017	0.018	[...]
Wheel of Time 1	0.04	0.053	0.019	0.052	0.02	0.017	0.025	0.013	0.009	0.023	[...]

Figure 5 Sample parallel list for the MFWSs (3-grams) in the selected works.

words for certain morphological and/or phonetic reasons. However, these linguistic options have stylistic impact in the present study, so it is reasonable to account for them as instances of stylistic choice.

Before proceeding to the actual analyses, it is important to decide how many word/word sequences in each text to consider from the tops of the parallel lists in order to gain the MFWs/MFWSs-based panoramas. Because previous studies

(see Burrows 2002; Rybicki 2012; Grabowski 2013; Mastropierro 2018) share no agreement on this number, we determine it through repeated pilot studies according to this principle: the stability of analytic results improve with the increase of the MFWs/MFWSs numbers but remain relatively unchanged once the numbers reach a certain point, at which a stylistic panorama is formed. Therefore, in the present analysis, we test with numbers from 100 to 2,000, using increments of 50, to arrive at the point at which stable results form a stylistic panorama. That point turns out to be the top 1,000 in the case of the MFWs-based panoramas, the top 950 in the case of 2-grams, and the top 900 in the case of 3-grams. For better consistency, we use the top 1,000 MFWs/MFWSs entries as the benchmark for conducting PCA analysis.

4.2.2 Overall Patterns of Stylistic Panoramas

The PCA results based on the MFWs, 2-grams, and 3-grams are depicted in Figures 6 through 7, respectively.

Figure 6 shows the overall extent to which the selected works from the three subgenres differ on the basis of the stylistic panoramas formed by the top 1,000
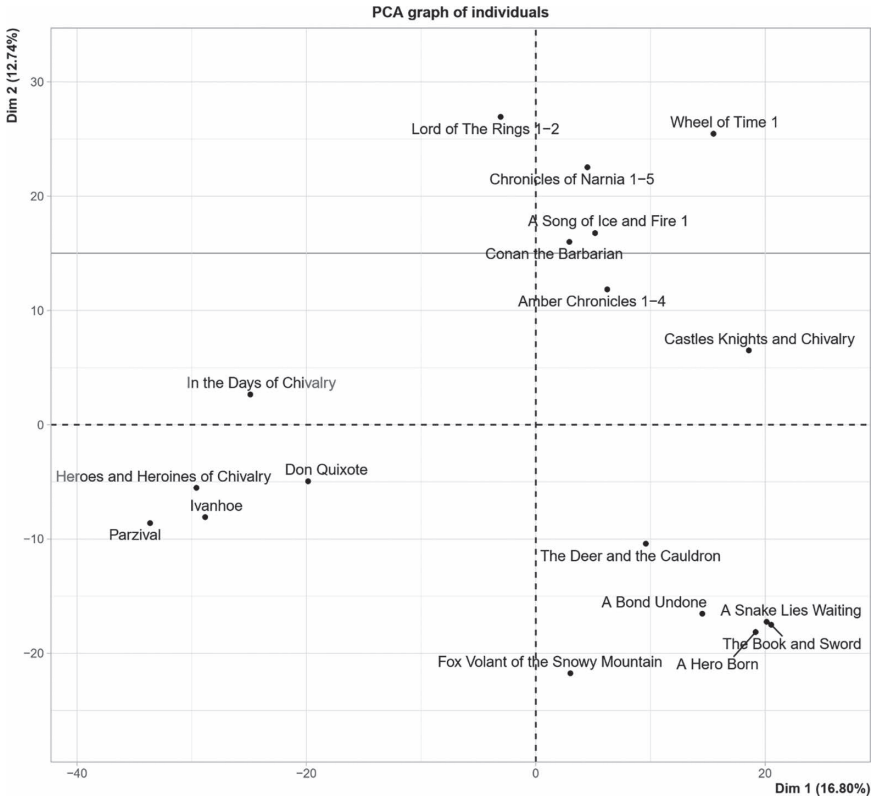

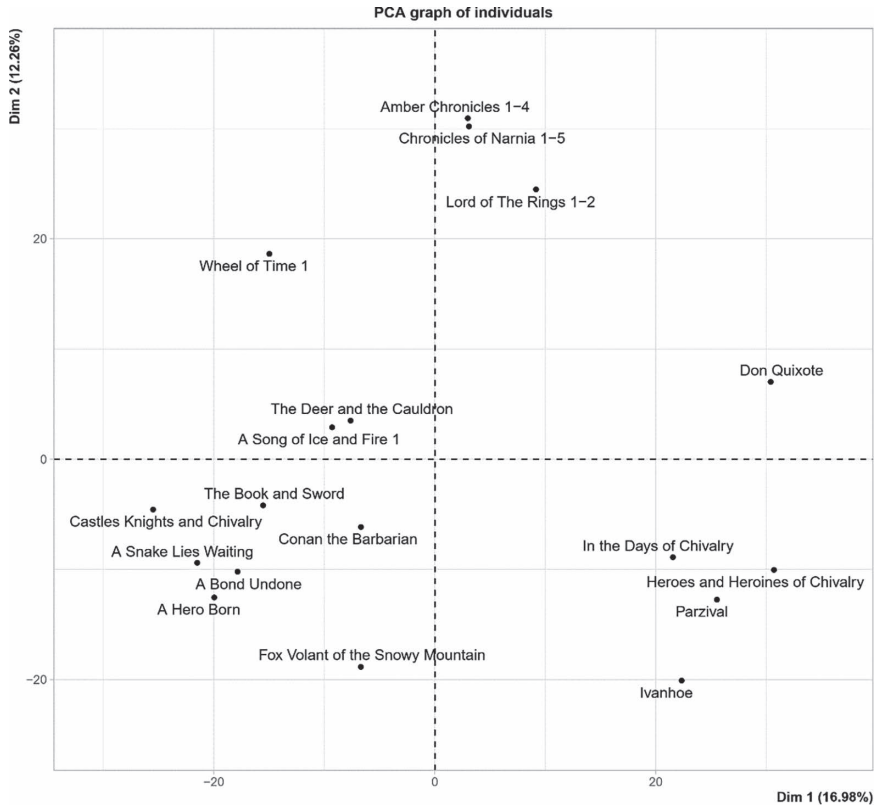*Figure 6* PCA graph of individuals, based on the top 1,000 MFWs.

Figure 7 PCA graph of individuals, based on the top 1,000 2-grams

most frequent words (MFWs). The dots in the figure are the stylistic panoramas of the heroic literary works represented by the MFWs, and the horizontal and vertical axes are the two principal components (dimensions) that represent the majority of data variance in the parallel word list. The metric distance between two dots signifies a possible diversity level between the stylistic panoramas of two works. The general principle is that the greater the metric distance between two dots, the higher the diversity level between the stylistic panoramas of the two works. The axes are unlabeled because they are the results of a dimensionality reduction[6] through the principal component analysis (PCA) – specifically, an unsupervised machine learning method, in which datasets are often unlabeled, unclassified, or uncategorized (Saslow 2018). The two percentiles in the brackets along the axes are the level of variance carried by the two components (dimensions): the first component (Dim 1) represents 16.80% of the variance across the data, whereas the second component (Dim 2) represents 12.74% of that variance. Those results
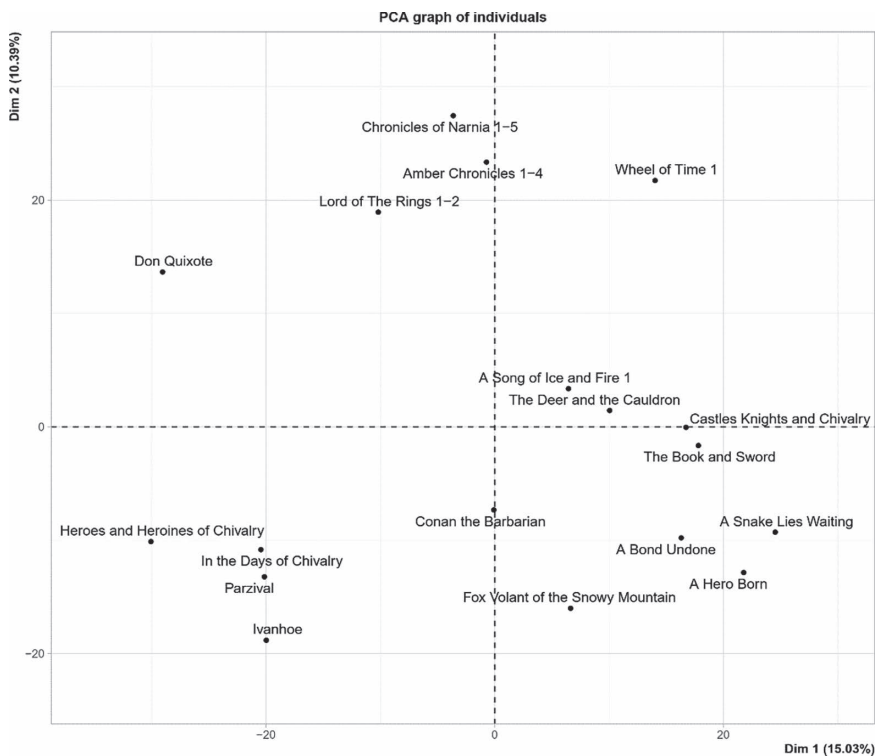
Figure 8 PCA graph of individuals, based on the top 1,000 3-grams.

show that Dim 1 has more data variance than Dim 2 does, thus implying that the distances between the data points along the horizontal axis bear greater variance than those along the vertical axis do.

In that light, the message conveyed by Figure 6 is clear: the relatively short distances between the dots representing the stylistic panoramas of the six *Wuxia* translations suggest that these translations share similarities in their MFWs. Meanwhile, longer distances between these dots and dots symbolizing the panoramas of the chivalric stories and heroic fantasies suggest that the translated *Wuxia* works are very different from the two Western subgenres in terms of their MFWs. Similarly, the dots that represent the chivalric stories (except for *Castles, Knights, and Chivalry*) and the heroic fantasies are mainly packed within their own subgenres and are distant from the ones representing works of other subgenres. That orientation indicates that most heroic works belonging to the same subgenre are prone to sharing their MFWs in texts.

When it comes to the panoramas formed by the top 1,000 most frequent word sequences (MFWSs), the stylistic pictures are largely similar to those stemming

from the top 1,000 MFWs, despite there being some slight differences. In Figure 7, the first principal component (Dim 1) shows 16.98% of the variance across the data, whereas the second principal component (Dim 2) carries 12.26% of that variance – a pattern that resembles the MFWs-based PCA results. Likewise, five of the six dots representing the *Wuxia* translations are close to each other but distant from the dots representing the works of other subgenres, except for the one for *The Deer and the Cauldron*, which is closest to the dot representing *A Song of Ice and Fire 1*, a work of heroic fantasies. This suggests that Minford's *Wuxia* translation may bear strong similarities to the novel by Martin in terms of 2-grams. In addition to the dots for the *Wuxia* translations, the dots symbolizing works of the two Westerns subgenres largely stay within their own subgenres, with the exception of the dot for *Castles, Knights, and Chivalry*, which lies closer to dots representing the *Wuxia* translations. In Figure 5, the stylistic panoramas formed by 3-grams demonstrate a mirrored but otherwise almost identical pattern with that in Figure 7, even though Dim 1 on the horizontal axis shows 15.03% of the data variance and Dim 2 on the vertical axis has 10.39% of such variance.

4.2.3 Metric Distances between Stylistic Panoramas

The above PCA results reveal how the MFWs/MFWSs-based stylistic panoramas of the *Wuxia* translations are similar to and/or different from those of the chivalric stories and heroic fantasies. However, they have only illustrated a general side of the stylistic picture wherein the exact level of similarities between a *Wuxia* translation and other works in the graphs is still unknown. To determine that level, we need to calculate the metric distances between each of the two dots through their coordinates, which are simultaneously generated in the PCA. With the coordinates of each dot, we use formula 5 to compute their metric distances, and then we focus on the average distances between the dots representing the *Wuxia* translations and the dots of the chivalric stories/heroic fantasies. In that way, the exact metric distances between the *Wuxia* translations and the chivalric stories/heroic fantasies are measured. Those average distances are reported in Table 5.

As the table shows, the MFWs-based stylistic panorama of *The Deer and the Cauldron* by Minford has the shortest average distance (32.80) to the panoramas

Table 5 Average Distances between Each *Wuxia* Translation and the Western Counterparts

Fiction	MFWs	2-grams	3-grams
Fox Volant of the Snowy Mountain	38.32	34.91	31.34
The Book and Sword	43.12	31.86	31.44
The Deer and the Cauldron	32.80	26.55	25.91
A Hero Born	42.74	37.88	38.75
A Bond Undone	39.15	35.47	33.43
A Snake Lies Waiting	42.72	37.42	39.35

of the selected chivalric stories and heroic fantasies, whereas the MFWs-based panorama of *A Hero Born* by Holmwood has the longest average distance to the other subgenres (42.74). The average distances from the MFWs-based panoramas of the remaining *Wuxia* translations to those of the chivalric stories and heroic fantasies fall within the range of 38.32 to 42.74 and hence are significantly greater than that of Minford's translation. Of the MFWSs-based panoramas, the previously described stylistic scenario seems to repeat itself in the case of 2-grams, but it bears nuances in the case of 3-grams. The 3-grams-founded panoramas illustrate that even though Minford's translation still has a noticeably shorter average distance to the other subgenres at 25.91, *A Snake Lies Waiting* by Chang has the longest distance of 39.35, a value that is slightly higher than that of Holmwood's translation at 38.75.

All these numbers suggest that with regard to the top 1,000 MFWs- and MFWSs-based panoramas, most of the chosen *Wuxia* translations are stylistically different from the selected chivalric stories and heroic fantasies, with the exception of *The Deer and the Cauldron* by Minford. This result of metric distances is in line with our direct observations of the PCA individuals, as shown in Figures 3 through 5, which we would relate again to translatorial motivations: Minford's motivation to win readerships in the English-speaking world (Cf. Section 4.1) may partly explain why his *Wuxia* translation resembles the chivalric stories/heroic fantasies in terms of the MFWs/MFWSs.

5 Discussion

With the two types of stylistic panoramas, the present study has illustrated the extent to which translated *Wuxia* and the two Western subgenres are stylistically connected. The stylistic panoramas based on the formal indices show that there are few similarities between the *Wuxia* translations and the stories of two Western subgenres, despite that certain similarities do exist between those *Wuxia* translations and heroic fantasies published in the 1990s/2000s. The MFWs-/MFWSs-based stylistic panoramas reveal that the *Wuxia* translations are stylistically different from those of most of the chivalric stories and heroic fantasies, with the exception of *The Deer and the Cauldron*. These findings, together with the study, have practical and theoretical implications with respect to the research questions.

On the practical side, the findings indicate reasons for the reception of *Wuxia* translations: unique stylistic features (richer *Wuxia*-specific vocabularies, shorter paragraph lengths, etc.) which distinguish *Wuxia* from both chivalric stories and heroic fantasies could be a possible reason that these *Wuxia* translations are well received. Table 9 summarizes the five-scale ratings of the six *Wuxia* translations by readers from four well-known websites of book promotions and reviews. Because some ratings are not available in *Novelupdates* and/or *Audible*, we focus on the average rating of each translation for better comparability. The table shows that *A Hero Born*, *A Bond Undone*, and *A Snake Lies Waiting* have the top three average ratings. We attribute the favorable ratings of the three *Wuxia* translations to their stylistic uniqueness, which is partly shown in the following two aspects.

Table 6 Reception of the Six *Wuxia* Translations in English (Up to 02/2021)

Fiction	Amazon	Goodreads	Novelupdates	Audible	Average
Fox Volant of the Snowy Mountain	3.40 of 5	3.84 of 5	3.00 of 5	n/a	3.41 of 5
The Book and Sword	4.80 of 5	3.89 of 5	3.20 of 5	n/a	3.96 of 5
The Deer and the Cauldron	4.20 of 5	4.28 of 5	4.40 of 5	n/a	4.29 of 5
A Hero Born	4.60 of 5	4.02 of 5	4.30 of 5	4.70 of 5	4.41 of 5
A Bond Undone	4.70 of 5	4.39 of 5	n/a	5.00 of 5	4.70 of 5
A Snake Lies Waiting	4.80 of 5	4.39 of 5	n/a	4.70 of 5	4.63 of 5

First, regarding the stylistic panoramas founded on the formal indices, relatively higher MATTR but lower DSL and APL values could contribute in part to favorable ratings. A high MATTR value suggests a rich vocabulary, which in *Wuxia* translations could mean the readers may receive an extended cultural experience of martial arts with a greater use of *Wuxia*-specific words. For instance, when rendering the original names of martial heroes and kung fu fighters, Holmwood and Chang both use more creative words, such as "*Ryder Han*," "*Ironheart Yang*," "*Twice Foul Dark Wind*," "*Nine Yin Skeleton Claw*," and the like. By contrast, in the earlier (the 1990s/2000s) *Wuxia* translations, those elements are sometimes presented less interestingly because of transliteration and/or omission. In addition, a lower DSL value in these translations could indicate a more repetitive yet consistent translation of the *Wuxia*-specific terms across sentences. Such an explanation would help us form a coherent impression of the fictional *Wuxia* world created by these terms. Finally, the lower APL values may reduce the readers' reading efforts when they come across certain culturally alien and linguistically idiosyncratic *Wuxia* elements. For instance, when the readers read a paragraph which contains many *Wuxia*-specific words (Ryder Han, Ironheart Yang, etc.), a relatively short paragraph with a low APL value around 30 (see Table 4) is more likely to reduce their cognitive load as they process these *Wuxia* elements. All these unique stylistic features may motivate readers to rate the three translations favorably.

Second, in terms of the MFWs-/MFWSs-based stylistic panoramas, a greater use of words and word sequences related to body language, body parts, mood, or inner feelings could likewise lead to more favorable ratings. When we look through the parallel lists, body-language words, such as "sighed," "pointed," "nodded," "shouted," and the like, appear frequently in the three translations, and words related to mood, such as "worried," "angry," "surprised," "scared," and so on, are also widely used in the same translations. In addition, 2- and/or 3-grams about body parts, such as "his neck," "his chest," "head and arms," and "in his hand," as well as ones for inner feelings, such as "dared to," "refused to," "had no idea," and "he wondered about," create vivid characterizations. This is because a preservation of the original descriptions of body language in the *Wuxia* translations may shorten the psychological distance between target readers and

the reconstructed *Wuxia* heroes/heroines, who are "alive" with perceptible human kinetic and/or mental presentations.

On the theoretical side, the study casts light on the use of stylistic panoramas in stylometric analyses in the following ways. First, the study attaches importance to the intended function of the chosen stylistic indices when using them as building blocks of a stylistic panorama. The seven formal indices are selected to show the general stylistic features of the heroic works, while the purpose in analyzing the MFWs and MFWSs is to identify the lexical resources in the same works. This could be important to a stylometric study because it binds indices together through a shared function. Nonetheless, the selection criteria of stylistic indices in some previous studies (Hossain et al. 2017; Liu and Xiao 2020) are not always made clear to readers. As a result, possible functions associated with those indices are often underexplored, which could lead to a tenuous connection between the selected indices.

Second, the study values triangulation of different types of stylistic panoramas when exploring holistic stylistic pictures of (translated) texts under investigation. For example, when the study concludes that the *Wuxia* translations are stylistically different from the chivalric stories and heroic fantasies, it has done so by triangulating the results stemming from the stylistic panoramas based on the seven formal indices and ones built by the MFWs and MFWSs. In this way, the study not only takes multiple functionally related stylistic features into account but also locates the stylistic pictures of the same genre from different stylistic perspectives. By contrast, some existing stylometric analyses of (translated) texts have confined their stylistic explorations to formal indices (Hossain et al. 2017; Liu and Xiao 2020) or MFWs/MFWSs (Eder 2017; Haverals et al. 2022) without attempting to triangulate the results from both sides. Consequently, extra stylistic pictures stemming from such triangulation are sometimes ignored in those analyses. In this light, we hold that such triangulation of stylistic panoramas may benefit stylometric analyses, as it helps the analyses transcend a single stylistic perspective by bringing multiple stylistic perspectives into play.

Third, the study holds that when we use stylistic panoramas in stylometric studies that highlight linguistic characteristics at a single level, however, there might be some weakness. This is especially evident when we attempt to use both types of stylistic panoramas. Despite the edge offered through triangulation, it would be less appropriate to use them simultaneously in studies dedicated to such single-level characteristics as words, word sequences, or sentences, since their scopes of investigations would be too narrow. Nevertheless, as the concept of stylistic panoramas is now in its infancy, it still has room for further development to satisfy the theoretical and methodological needs of varied stylometric studies.

6 Conclusion

Returning to the original research interest, the study can now give a clear answer: stylistic connections between translated Chinese *Wuxia* and Western heroic literature

are weak because the stylistic panoramas founded on the formal indices and the MFWs/MFWSs have demonstrated important stylistic differences across the genres. Despite these divergences, the study has made the following contributions to *Wuxia* translation research and stylometric studies: first, it highlights possible stylistic connections between heroic literature in the East and that in the West, clues which may help understand the reception of Chinese *Wuxia* in the West. Second, it demonstrates the use of the stylistic panorama, a concept that seeks to describe the stylistic picture of a (translated) text in a relatively holistic way by binding different stylistic indices together, with respect to function.

Nonetheless, this study has several limitations, one of which is that the stylometric analyses are founded on a relatively small number of *Wuxia* translations. Even though the study has included all the English *Wuxia* translations published at the time of writing, it is assumed that when there are more *Wuxia* translations in the future, the results might be slightly different due to various translatorial (translators' motivations, preferences, etc.) and/or extratranslatorial (patronage intervention, sociocultural influences, etc.) reasons. Furthermore, the present selection of stylistic indices in the formation of panoramas has considered only general features, whereas an alternative selection that favors more idiosyncratic features (hapax legomena related to martial arts, chivalry, fantasies, etc.) pertaining to the heroic literature could be equally potent in uncovering stylistic connections between heroic literature in the East and that in the West.

For further research along this line, the publication of additional *Wuxia* translations could allow works by different authors to be incorporated into the corpus to produce more insightful results. In the meantime, stylistic indices that focus on various idiosyncratic features of heroic literature can be considered to widen the scope of meaningful research.

Funding

Notes

1 The article was originally published on April 23, 2022, in *Digital Scholarship in the Humanities* (DSH), DOI: 10.1093/llc/fqac019. It is reused under license 5304200620907 permitted by Oxford University Press. Credit goes to DSH, Oxford University Press, the European Association for Digital Humanities (EADH), and Alliance of Digital Humanities Organizations (ADHO).
2 All *Wuxia* translations used in the study were purchased from Amazon.com as e-books.
3 All the Western heroic works used in the study were available in the public domain and were downloaded freely from Gutenberg.org as "txt" files.
4 O is the minimum sample size, and k is the number of variables.
5 Full tables are available in Figshare (DOI: 10.6084/m9.figshare.19361468).
6 Because PCA is a multivariate statistical analysis that operates according to dimensionality reduction (Manly 2016, 102–3), multiple dimensions in the analysis were compressed into two dimensions – a more manageable scale for the present work.

References

Březina, Vaclav. 2018. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.

Brocardo, Marcelo Luiz, Issa Traore, and Isaac Woungang. 2014. "Toward a Framework for Continuous Authentication using Stylometry." 2014 IEEE 28th International Conference on Advanced Information Networking and Applications, Victoria, BC, Canada, 106–15.

Broder, Andrei Z., Steven C. Glassman, Mark S. Manasse, and Geoffrey Zweig. 1997. "Syntactic Clustering of the Web." *Computer Networks and ISDN Systems* 29, no. 8: 1157–66.

Burrows, John. 2002. "'Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing* 17, no. 3: 267–87.

Christopher, Manning, Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Covington, Michael, and Joe D. McFall. 2010. "Cutting the Gordian Knot: The Moving-Average Type – Token Ratio MATTR." *Journal of Quantitative Linguistics* 17, no. 2: 94–100.

Craig, Hugh. 2018. *Intelligent Archive 3.0*. Newcastle: University of Newcastle.

Daelemans, Walter. 2013. "Explanation in Computational Stylometry." In *Computational Linguistics and Intelligent Text Processing*, edited by Alexander Gelbukh, 451–62. Berlin: Springer.

Eder, Maciej. 2017. "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities* 32, no. 1: 50–64.

Eder, Maciej, Jan Rybicki, and Mike Kestemont. 2016. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8, no. 1: 107–21.

Flannery, Mary. 2012. "The Concept of Shame in Late-Medieval English Literature." *Literature Compass* 9, no. 2: 166–82.

Formann, Anton. K. 1984. *Die Latent-class-analyse: Einführung in Theorie und Anwendung*. Weinheim: Beltz.

Grabowski, Łukasz. 2013. "Interfacing Corpus Linguistics and Computational Stylistics: Translation Universals in Translational Literary." *International Journal of Corpus Linguistics* 18, no. 2: 254–80.

Haverals, Wouter, Lindsey Geybels, and Vanessa Joosen. 2022. "A Style for Every Age: A Stylometric Inquiry into Crosswriters for Children, Adolescents and Adults." *Language and Literature* 31, no. 1: 62–84.

Holmes, David I. 1998. "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing* 13, no. 3: 111–17.

Honegger, Thomas. 2010. "Heroic Fantasy and the Middle Ages – Strange Bedfellows or An Ideal Cast?" *Itinéraires: Littérature, Textes, Cultures*, no. 3: 61–71.

Hossain, M. Tahmid, Md. Moshiur Rahman, Sabir Ismail, and Md Saiful Islam. 2017. "A Stylometric Analysis on Bengali Literature for Authorship Attribution." In *ICCIT 2017: 20th International Conference on Computer and Information Technology*, 1–5. Dhaka, Bangladesh: IEEE.

Hou, Renkui, and Churen Huang. 2020. "Robust Stylometric Analysis and Author Attribution Based on Tones and Rimes." *Natural Language Engineering* 26, no. 1: 49–71.

Huang, Yonglin. 2018. *Narrative of Chinese and Western Popular Fiction*. Berlin: Springer.

Jones, Ewan, and Paul Nulty. 2019. "Quantitative Measures of Lexical Complexity in Modern Prose Fiction." *Digital Scholarship in the Humanities* 34, no. 4: 914–37.

Kettunen, Kimmo. 2014. "Can Type-Token Ratio Be Used to Show Morphological Complexity of Languages?" *Journal of Quantitative Linguistics* 21, no. 3: 223–45.

Keulemans, Paize. 2020. *Sound Rising from the Paper: Nineteenth-Century Martial Arts Fiction and the Chinese Acoustic Imagination*. Leiden: Brill.

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

Liu, Ying, and Tianjiu Xiao. 2020. "A Stylistic Analysis for Gu Long's Kung Fu Novels." *Journal of Quantitative Linguistics* 27, no. 2: 32–61.

Manly, Bryan F. J. 2016. *Multivariate Statistical Methods: A Primer* (4th ed.). London: Chapman and Hall.

Mastropierro, Lorenzo. 2018. *Corpus Stylistics in Heart of Darkness and Its Italian Translations*. London: Bloomsbury Publishing.

Mei, Jia. 2019. "Turning Action into Words." *China Daily*, April 19. http://global.chinadaily.com.cn/a/201904/19/WS5cb91987a3104842260b70d3_2.html. Accessed 28 February 2021.

Melka, Tomi S., and Michal Místecký. 2020. "On Stylometric Features of H. Beam Piper's Omnilingual." *Journal of Quantitative Linguistics* 2, no. 7: 204–43.

Rong, Zheng, Li Jiexun, and Chen Hsinchun. 2006. "A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques." *Journal of the American Society for Information Science and Technology* 57, no. 3: 378–93.

Rybicki, Jan. 2012. "The Great Mystery of the Almost Invisible Translator." In *Quantitative Methods in Corpus-Based Translation Studies*, edited by M. P. Oakes and M. Ji, 231–48. Amsterdam: John Benjamins.

Saslow, Elliott. 2018. "Unsupervised Machine Learning." *Towards Data Science*. https://towardsdatascience.com/unsupervised-machine-learning-9329c97d6d9f. Accessed 28 February 2021.

Scott, Mike. 2012. *WordSmith Tools version 6*, Stroud: Lexical Analysis Software. https://www.lexically.net/publications/citing_wordsmith.htm.

Stanford NLP Group. 2021. "Stanford Tagger 4.2.0." https://nlp.stanford.edu/software/tagger.html. Accessed 28 February 2021.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at Work*. Amsterdam: John Benjamins.

Valencia, Alex I., Helena Gómez-Adorno, Christophe Rhodes, and Gibran Fuentes Pineda. 2019. "Bots and Gender Identification Based on Stylometry of Tweet Minimal Structure and N-Grams Model." In *Working Notes of CLEF 2019 – Conference and Labs of the Evaluation Forum*, 1–8. Lugano: CLEF.

Vander Elst, Stefan. 2017. *The Knight, the Cross, and the Song: Crusade Propaganda and Chivalric Literature, 1100–1400*. Philadelphia, PA: University of Pennsylvania Press.

Wu, Kan, and Dechao Li. 2018. "Lexical Normalization in English Translations of Jin Yong's Martial Arts Fiction: A Corpus-Based Study." In *Asia Pacific Interdisciplinary Translation Studies*, edited by X. Luo, 93–106. Beijing: Tsinghua University Press.