Global Language and Cultural Studies, 2026, Volume 1, Issue 1
https://bbwpublisher.com/journal/GLCS
Online ISSN : 3083-3035
Print ISSN : 3083-3043

Translator's Style in Different English Versions of *Shiji* : A Corpus-based Comparative Study

Jiaxin Tian, Shiying Liu*

School of Foreign Languages, Chongqing University of Posts and Telecommunications (CQUPT), Chongqing 400065, China

Author to whom correspondence should be addressed.

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: This study employs a corpus-based methodology to explore the stylistic differences among three prominent English translations of *Shiji* by Yang Xianyi & Gladys Yang, Burton Watson, and William H. Nienhauser, Jr. Focusing on lexical, syntactic, and textual dimensions, the research investigates core indicators such as lexical diversity, lexical density, sentence structure, cohesive devices, and overall readability to identify their distinct translator styles. The findings reveal that each translator's stylistic choices are closely aligned with their respective translation orientations and target audiences. Yang's version prioritizes fidelity to the source text; Watson's translation emphasizes narrative flow and accessibility for general Western readers; Nienhauser's rendition optimizes academic clarity and readability.

Keywords: Translator style; corpus-based translation studies

Online publication: February 10, 2026

1. Introduction

As China's first biographical history book in Chinese literary, *Shiji*, compiled by Sima Qian during the Western Han Dynasty, occupies an irreplaceable position in both Chinese and world historiography. Beyond its historical value as a comprehensive record of Chinese history from the legendary Yellow Emperor to the reign of Emperor Wu of Han, *Shiji* is also a carrier of traditional Chinese culture, serving as a crucial text for cross-cultural communication. Amid the accelerated globalization and cultural exchanges between China and the West, English translations of *Shiji* have emerged as vital bridges facilitating Western readers' understanding of ancient Chinese history and culture.

There have been to date several influential English translations of *Shiji*, among which the versions by Yang Xianyi & Gladys Yang, Burton Watson and William H. Nienhauser, Jr. are the most widely recognized and studied. Yang Xianyi and Gladys Yang, completed and published the *Selections from Records of The Historian* in 1979. Their translation was a government-led cultural communication[1]. Burton Watson, a prominent American translator, produced *Records of The Grand Historian*, an abridged English translation of *Shiji*, first published in 1961. William H. Nienhauser, Jr., an American sinologist specializing in Chinese classical literature, led a team of

scholars to complete *The Grand Scribe's Records* since 1994.

Despite the existence of these influential translations, comparative studies of their stylistic features still remain insufficient. Traditional translator style research often relies on subjective criticism. In contrast, the rise of corpus-based translation studies (CTS) since the 1990s has provided a rigorous quantitative research paradigm for translator style analysis. By constructing specialized corpora and utilizing corpus tools to extract and analyze linguistic features, CTS enables objective and systematic comparisons of translator styles, thereby overcoming the limitations of traditional qualitative research. This paradigm shift has laid a solid foundation for the in-depth exploration of stylistic differences among the three English translations of *Shiji*.

This study selects *Basic Annals of The First Emperor of the Qin* as the source text and compares its three renditions. This chapter focuses on Qin Shi Huang, the first emperor who unified China and established the first centralized feudal dynasty in Chinese history, recording pivotal historical events such as the unification of the six kingdoms, the establishment of imperial institutions, and the implementation of uniform codes, weights, and measures. Thus, the English translations of this chapter not only convey historical facts but also reflect how translators interpret and transmit the core elements of China's unification civilization to Western audiences.

This study adopts a mixed-methods research design that combines both quantitative analysis and qualitative interpretation. Employing WordSmith 9.0 for quantitative data extraction and analysis, the study compares a series of lexical, syntactic, and textual indicators from the three translated corpora. These indicators will be used to objectively identify the stylistic differences among the three translations. On the basis of the quantitative results, the study conduct in-depth interpretation of the observed stylistic differences.

This study is structured into five chapters. Chapter 1 serves as the introduction. Chapter 2 is the literature review and the research gaps addressed by this study are identified. Chapter 3 details the research methodology and corpus construction process. Chapter 4 conducts the core corpus-based comparative analysis of translator styles. Chapter 5 is the conclusion and implications.

2. Literature Review

Translator's style refers to translator's preference in language use, or the recurrent linguistic patterns in the target text[2]. Traditionally, translation was regarded as an act of imitation, and the translator's style was considered subordinate to the source text (ST) style, leading to its "invisibility"[3]. The cultural turn of translation studies indicated a shift of focus from the centeredness of the source text to the centeredness of the target text. The translator can interpret the source text in many different ways and demonstrate his/her interpretation in the target text, which constitutes his/her style[2]. Mona Baker, who pioneered the application of corpus methods to explore literary translator's style, conceptualized translator's style as a "thumbprint" manifested in a range of linguistic and non-linguistic features, emphasizing that it refers to the consistent linguistic choices made by translators across multiple translations, independent of the influence of source texts[4].

The term, corpus-based translation studies, was first introduced in 1996 by Mona Baker. It can be defined as the branch of the discipline that uses corpora of original and/or translated text for the empirical study of the product and process of translation, the elaboration of theoretical constructs and the training of translators" [5].

Corpus-based translation studies is essentially a partnership marriage between corpus linguistics and descriptive translation studies[2]. According to Hu, corpus-based translation studies has the following characteristics. Firstly, the research is based on the analysis of a sizable collection of naturally occurring texts in a corpus, with focus on the

actual patterns of the translated texts. Secondly, it combines quantitative analysis and qualitative research. Thirdly, it applies both linguistic and cultural studies approaches to translation studies. Corpus-based studies into translator's style can be categorized into two types : T-type (target text type) and S-type (source text type)[6]. T-type studies are based on comparable corpora and focus on the habitual linguistic behaviour in many translations by an individual translator. S-type studies are based on parallel corpora and take source text features into consideration.

A number of studies provide insights into the translation of *Shiji*. Gao finds that Yang uses distinctive translation strategies in the translation process: faithfulness to the original text, recreation of a classic, flexibility as necessary, and ensuring readability[7]. Wu holds the belief that Watson's specific translation strategy favors continuity over annotations and avoided covering up the work's literary characteristics with academic characteristics[8]. Watson paid meticulous attention to the readability of his English translation and made great achievements in language, literature, and art. Li and Wang discuss the influence of "patronage, ideology and poetics" on Watson's and Nienhauser's translation[9]. Wei (2023) reveals that Watson's version targets general readers and non-specialized students, emphasizing the literary value of *Shiji*.

While Nienhauser's translation is a collaborative academic full-translation project involving scholars from multiple countries, designed for experts and specialized students[10]. There exist few studies that delve into the comparison of the most renowned translated versions of *Shiji* from a corpus-based perspective.

3. Research Design

The source text is derived from chapter six of *Shiji*, Basic Annals of The First Emperor of the Qin. Yang Xianyi & Gladys Yang's translation is excerpted from *Selections from Records of The Historian* (Foreign Languages Press, 1979); Burton Watson's translation is extracted from *Records of the Grand Historian: Qin Dynasty* (Columbia University Press, 1993); and William H. Nienhauser, Jr.'s translation is taken from *The Grand Scribe's Records, Volume I: The Basic Annals of Pre-Han China* (Indiana University Press, 1994). Only the common part of the chapter that the three versions all cover is included in the corpora, so that they are comparable. Prior to corpora construction, non-translation components such as annotations and editorial notes are all removed to ensure that the corpus only contains the translated text.

Subsequently, the cleaned texts are formatted into plain text files and imported into WordSmith 9.0 for corpus building.

4. Results and Discussion

4.1. Lexical Level

4.1.1. Lexical Diversity

Table 1. Standardized type/token ratio (STTR) and mean word length

	Tokens	Types	TTR	STTR	STTR Std. dev.	STTR basis	Mean word length
Yang	12,614	2,462	19.52	44.73	50.73	1,000	4.52
Watson	15,696	2,765	17.62	43.09	53.05	1,000	4.52
Nienhauser	15,352	2,958	19.27	44.29	51.55	1,000	4.48

Type refers to different words in a corpus, or a word form that appears for the first time[11]. Token refers to all word forms that appear in a corpus[2] .A type/token ratio (TTR) is a measure of the range and diversity of vocabulary used by a writer or in a given corpus[12]. When the text length is long, the standardized type/token ratio (STTR) that refers to the type/token ratio every 1000 words is usually employed to compare the vocabulary richness among translated texts[13]. Yang's version exhibits the highest STTR (44.73%), indicating a notably richer lexical diversity despite having the smallest token count (12,614). In contrast, Watson's translation presents the lowest STTR (43.09%) among the three, even though it has the largest token count (15,696).

The lower lexical diversity suggests a preference for repetitive core vocabulary, a strategy that effectively reduces reading barriers for audiences unfamiliar with Chinese ancient history and culture. William H. Nienhauser, Jr.'s version falls midway between the two in terms of STTR (44.29%) , with a token count (15,352) close to Watson's but a higher type count (2,958). He retains sufficient lexical variation to convey the complexity of historical information, with a slightly higher mean word length (4.48), while avoiding overly obscure vocabulary that might hinder comprehension, as evidenced by his moderate lexical diversity metrics.

4.1.2. Lexical Density

Table 2. Ratio of Content Words

	Nouns	Verbs	Adjectives	Adverbs	Content words	Tokens	Lexical density
Yang	3,910	2,308	739	549	7,506	12,614	59.51%
Watson	4,772	2,879	904	808	9,363	15,696	59.7%
Nienhauser	4,773	2,653	842	735	9,003	15,352	58.64%

Another measurement at the lexical level is lexical density, which refers to the ratio of the number of content words (nouns, verbs, adjectives, adverbs) per number of running words (tokens) [14]. Generally speaking, the smaller the lexical density, the simpler the language being used[2]. And the higher the lexical density, the more information included in the language. All the three texts are tagged with Part of Speech (POS) through TreeTagger 3.0 before being imported into WordSmith 9.0 for the computation of different content words. The values are remarkably close: Yang's version at 59.51%, Watson's at 59.7% and Nienhauser's, at 58.64%. It indicates that Nienhauser's translation carries slightly less information and is relatively easier for readers. Watson's version, though with the lowest STTR, or lexical diversity, stands out to have the highest lexical density among the three.

4.2. Syntactic level

Table 3. Number of Sentences and Mean Sentence Length

	Sentences	Mean sentence length	std.dev.
Yang	636	18.99	11.02
Watson	803	19.31	10.17
Nienhauser	935	16.17	11.91

Mean sentence length is another metric that reveals the complexity and readability of a text. A longer mean sentence length corresponds to a rather complex text. The longer the sentence is, the more difficult the text is, which can be used as a marker of the general translator`s style[15]. Nienhauser's translation stands out with the highest number of sentences (935) and the shortest mean sentence length (16.17 words), which reflects his pursuit of clarity and readability with shorter sentences. Burton Watson's version has a moderate number of sentences (803) and the longest mean sentence length (19.31 words) among the three. While Yang's translation has the fewest sentences (636) and a moderate mean sentence length (18.99 words). The difference can be evidenced by one of the translation case:

Source: 秦始皇帝者，秦庄襄王子也。庄襄王为秦质子于赵，见吕不韦姬，悦而取之，生始皇。以秦昭王四十八年正月生于邯郸。及生，名为政，姓赵氏。

Yang: Shih Huang Ti was the son of King Chuang-hsiang of Chin, who while a hostage in Chao fancied one of Lu Pu-wei's concubines and took her. She gave birth to the First Emperor in Hantan in the first month of the forty-eighth year of King Chao of Chin. At his birth he received the name Cheng, with Chao as his surname.

Watson: The First Emperor of the Qin was a son of King Zhuangxiang of Qin. When King Zhuangxiang was a hostage for the state of Qin in Zhao, he happened to see a concubine belonging to Lü Buwei. Pleased with her, he took her for himself, and she bore the First Emperor. He was born in Handan in the first month of the forty-eighth year of King Zhaoxiang of Qin (259 BC). At his birth he was given the personal name Zheng; his family name was Zhao.

Nienhauser: The First Emperor of Ch'in 秦始皇帝 was the son of King Chuang-hsiang 庄 襄 (r.249-247 B. C.) of Ch'in. When King Chuang-hsiang was Ch'in's hostage prince in Chao, he saw Lü Pu-wei's concubine. Pleased with her, he married her and begot the First Emperor. The First Emperor was born in Han-tan 邯 郸 in the first month of the forty-eighth year of King Chao 昭 of Ch'in(259 B. C.). At birth he was given the praenomen Cheng 政 and the nomen Chao 趙 .

In this case, Yang's translation has a tighter structure, conveying all the information with just three concise sentences. While Watson and Nienhauser both use five relatively shorter sentences.

4.3. Textual Level

4.3.1. Cohesion

Cohesion occurs where the interpretation of some elements in the discourse is dependent on that of another. The one presupposes the other, in the sense that it cannot be effectively decoded except by recourse to it[16]. Halliday and Hasan identify two types of cohesion: "grammatical cohesion", that can be achieved by reference, substitution, conjunction and ellipsis, and "lexical cohesion", that can be achieved by reiteration or repetition and collocation. The study compares the use of two distinctive indicators, conjunctions and pronouns, in the three translation versions.

Table 4. Percentage of Conjunctions

	Coordinating conjunctions	Subordinating conjunctions	Percentage
Yang	606	1,287	15%
Watson	808	1,721	16.11%
Nienhauser	704	1,398	13.7%

Coordinating conjunctions (e.g., “and,” “but”) connect equal syntactic units, while subordinating conjunctions (e.g., “when,” “because”) establish hierarchical relationships (e.g., time, cause, condition), directly reflecting how translators interpret and re-present the sequence and causality of the source text for their target audiences.

The percentage of conjunctions relative to total tokens reveals the extent to which translators rely on explicit cohesive devices to enhance textual cohesion. Watson’s version exhibits the highest overall conjunction percentage (16.11%), with 808 coordinating conjunctions and 1,721 subordinating conjunctions. This emphasis on explicit conjunctions reflects his focus on accessibility for general Western readers.

He reduces the cognitive burden of interpreting the implicit logic inherent in classical Chinese, ensuring the narrative flows smoothly for audiences less familiar with Chinese historical contexts. Yang’s translation follows with a 15% conjunction percentage (606 coordinating, 1,287 subordinating). Nienhauser’s version has the lowest conjunction percentage (13.7%) despite 704 coordinating and 1,398 subordinating conjunctions. This aligns with his syntactic strategy of shorter, more fragmented sentences. The discrete structure of his text inherently reduces the need for excessive conjunctions.

Table 5. Percentage of Pronouns

	Personal pronouns	Possessive pronouns	Percentage
Yang	448	317	6.1%
Watson	642	318	6.12%
Nienhauser	673	361	6.74%

Personal pronouns (e.g., “he”, “they”) and possessive pronouns (e.g., “his”, “their”) are key referential devices in discourse. The percentage of these pronouns relative to total tokens directly impacting readability and show translator’s different styles, as pronoun use is a critical marker of how translators prioritize clarity for their intended readers. Nienhauser’s version stands out with the highest overall percentage of personal and possessive pronouns (6.74%), including 673 personal pronouns and 361 possessive pronouns. This aligns with his syntactic choice of shorter sentences. The fragmented structure increases the risk of referential ambiguity, so the higher pronoun frequency clarifies entities across discrete clauses. Watson’s and the Yang’s versions exhibit nearly identical percentages (6.12% and 6.1%).

4.3.2. Readability

Table 6. Readability

	Flesch reading ease	Text difficulty	Flesch-Kincaid grade level
Yang	66.07	33.93	9.11
Watson	62.65	37.35	9.73
Nienhauser	68.11	31.89	7.79

The Flesch Reading Ease Score is a readability test designed to indicate how difficult a passage in English is to understand. A higher score means a more difficult text fore readers. Text difficulty is an inverted score (100-Flesch Reading Ease score), so that a score of 0 represents the easiest text and 100 represents the most difficult.

Flesch-Kincaid Grade Level score rates text on a U.S. school grade level, which is also an indicator of readability. All the three texts are processed with BFSU Readability Analyzer 3 for these readability data. Nienhauser's version achieves the highest reading ease score (68.11), the lowest text difficulty (31.89), and the most accessible grade level (7.79), confirming his success in prioritizing clarity for academic and student audiences. This outcome is consistent with his syntactic choices—splitting the source text into the most sentences (935) with the shortest mean sentence length (16.17 words). Yang's translation follows with a moderate reading ease score (66.07), text difficulty (33.93), and grade level (9.11). In contrast, Watson's version exhibits the lowest reading ease (62.65), highest text difficulty (37.35), and most demanding grade level (9.73) .

5. Conclusion

This study adopts a corpus-based approach to explore the stylistic differences among three influential English translations of Sima Qian's *Shiji* by Yang Xianyi & Gladys Yang, Burton Watson, and William H. Nienhauser, Jr. Through quantitative analysis of lexical, syntactic, and textual level indicators, the research identifies the distinct stylistic characteristics of each translator.

At the lexical level, Yang's version demonstrates the highest lexical diversity, while Watson's translation exhibits the lowest. Lexical density across all three translations is remarkably consistent, confirming their shared commitment to retaining the information-rich nature of the source historical text, while subtle variations mirror their audience-specific strategies.

Syntactically, Nienhauser's style is defined by the highest number of sentences and the shortest mean sentence length, a choice that decomposes dense classical Chinese into structured, readable units.

At the textual level, cohesion and readability metrics further validate these stylistic orientations. Watson's translation relies most heavily on explicit conjunctions to clarify implicit logical relationships in the source text, reducing cognitive load for readers. Nienhauser's lowest percentage of conjunctions and the highest percentage of pronouns, coupled with his syntactic data, the largest number of sentences and shortest mean sentence length, show that he addresses referential ambiguity inherent in his fragmented sentence structure. Readability metrics reinforce these patterns: Nienhauser's version is the most accessible, Yang's is moderately accessible, and Watson's is the most demanding.

These results enrich corpus-based translator style research in classical Chinese historical text translation. However, This study is subject to limitations. The research corpus is confined only to a single chapter of *Shiji*. While this focus ensures depth of analysis, it limits the generalizability of findings to the translators' overall stylistic tendencies across the entire *Shiji*. Translators may adapt their styles to different text types, and a broader corpus may provide a more comprehensive understanding of their consistent stylistic traits.

Disclosure statement

The author declares no conflict of interest.

References

[1] Yan P, 2021, The translation of Basic Annals of the First Emperor of the Qin: An imagological approach. Foreign Language Research, 9(1): 74–81+109.

[2] Hu K, 2016, Corpus-Based Study of Translator's Style. In: Introducing Corpus-based Translation Studies. New Frontiers in Translation Studies. Springer, Berlin, Heidelberg.

[3] Venuti L, 1995, The Translator's Invisibility: A History of Translation. Routledge.

[4] Baker M, 2000, Towards a methodology for investigating the style of a literary translator. Target, 12(2): 241–246.

[5] Laviosa S, 2007, Learning creative writing by translating Witty Ads. The Interpreter and Translator Trainer (ITT), 1(2): 197–222.

[6] Huang L, 2021, New Trends in Corpus-Based Translator's Style Studies. In: Wang, V.X., Lim, L., Li, D. (eds) New Perspectives on Corpus Translation Studies. New Frontiers in Translation Studies. Springer, Singapore.

[7] Fengping G, 2024, Replication and recreation of a classic: Yang Hsien-yi and Gladys Yang's English translation of The Records of the Grand Historian (Shiji). Chinese Studies in History, 57(1): 58–66.

[8] Tao W, 2024, An analysis of the image positioning and literary characteristics of Burton Watson's English translation of Records of the Grand Historian (Shiji). Chinese Studies in History, 57(1): 47–57.

[9] Li X, Wang R, 2011, Writing Shi Ji into world literature on Burton Watson's translation. Across Languages and Cultures, 12(2): 95–134.

[10] Wei H, 2022, A comparative study of Watson's and Nienhauser's English translations and introductions of Shiji. Journal of Huaibei Normal University (Philosophy and Social Sciences), 43(5): 69–73.

[11] Yang HZ, 2002, An Introduction to Corpus Linguistics. Shanghai Foreign Language Education Press.

[12] Baker M, 1995, Corpora in translation studies: An overview and some suggestions for future research. Target, 7(2): 223–243.

[13] Gao Y, Zhou G, 2024, Revealing the translator's style: A corpus-based study of english translations of Mencius. PLOS ONE, 19(7): e0305894.

[14] Ure J, 1971, Lexical Density and Register Differentiation. Contemporary Educational Psychology, 5: 96-104.

[15] Olohan M, 2004, Introducing Corpora in Translation Studies. Rutledge.

[16] Halliday MAK, Hasan R, 1976, Cohesion in English. Routledge.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.