

基于语料库的《长恨歌》四译本的译者风格研究

张源迅

贵州财经大学外语学院，贵州 贵阳

收稿日期：2026年7月13日；录用日期：2026年8月10日；发布日期：2026年8月24日

摘 要

本研究通过自建平行语料库，运用语料库统计方法，借助WordSmith9.0与AntConc4.3.1等语料库检索软件，从词汇、句子及语篇三个层面，对《长恨歌》的四个译本：弗莱彻译本、翟里斯译本、杨宪益、戴乃迭译本、许渊冲译本展开定量分析，尝试从语言数据视角管窥并解读各译本所呈现的译者风格，并以此作为基础，比较这四部作品的语言特点及其翻译手法，从而深度挖掘译者所处的年代和国籍如何影响其译著，以期为类似古诗词的翻译工作带来启示。

关键词

语料库，译者风格，《长恨歌》译本

A Corpus-Based Study on Translators' Styles of Four English Versions of *Song of Everlasting Sorrow*

Yuanxun Zhang

School of Foreign Languages, Guizhou University of Finance and Economics, Guiyang Guizhou

Received: July 13, 2026; accepted: August 10, 2026; published: August 24, 2026

Abstract

By constructing a self-built parallel corpus and adopting corpus statistical methods, this study employs retrieval software including WordSmith 9.0 and AntConc 4.3.1 to conduct quantitative analysis on four English translations of *Song of Everlasting Sorrow* respectively rendered by Fletcher,

Giles, Yang Xianyi & Gladys Tayler, and Xu Yuanchong. From lexical, syntactic and textual dimensions, it attempts to explore and interpret translators' styles reflected in each version based on linguistic data. On this basis, the paper compares linguistic features and translation strategies of the four translations, and further investigates how translators' historical eras and national backgrounds exert influences on their works. It is expected to provide references for future translation practices of *Song of Everlasting Sorrow* and other classical Chinese poems.

Keywords

Corpus, Translator's Style, Translations of *Song of Everlasting Sorrow*

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

唐代诗人白居易的长篇叙事诗《长恨歌》以唐玄宗与杨贵妃的爱情悲剧为主线，融合历史、政治、人性与神话想象，把现实主义的精神内涵和浪漫主义的幻想元素巧妙融为一体，对中唐时期诗歌的发展起到了积极的促进作用。

译者风格，又称译者文体或翻译文体，有广义和狭义之分。前者指译者在语言应用方面所表现出来的个性特征以及包括译本选择、翻译策略应用、序跋和译注等在内的非语言特征。后者则指译者语言应用或语言表达的偏好，或在译本中反复出现的语言表达方式[1]。1993年莫娜·贝克(Mona Baker) [2]发表“语料库语言学与翻译研究：启示和应用”(“Corpus Linguistics and Translation Studies: Implications and Applications”)一文，首次系统阐述语料库方法在翻译研究中的应用价值，标志着语料库翻译研究的诞生。其中，基于语料库对译者风格进行考察是语料库在翻译研究中的一项重要应用。莫娜·贝克(Mona Baker) [3]在2000年利用TEC语料库对Peter Bush和Peter Clark的译者风格进行了分析，为后续译本译者风格的研究奠定了基础研究框架。目前，国内外已有众多学者借助语料库技术开展文学作品译者风格研究，积累了丰富的研究成果。截至2026年7月13日，在中国知网以“语料库译者风格”为检索词进行检索，与外国语言文学相关的学术论文，包括254篇期刊论文、204篇学位论文、13篇会议论文、17篇辑刊论文以及19篇特色期刊。这些研究的对象多为文学名著以及古典诗歌为主，也有部分论文以具体译者为研究核心，结合其相关译作，重点分析知名译者的翻译风格。

现有《长恨歌》译本研究多偏向定性分析，少有探讨译者时代与国别文化对译文的影响。本文借助语料库搭建四个译本的平行语料库，对比译本语言特征与翻译策略，探究译者背景带来的翻译差异，以期为古典诗词翻译研究提供参考。但本研究也有其局限性，仅仅分析了通用书面语料库量化指标(平均句长、词汇密度、类符形符比等)，未针对古典叙事诗的韵律、分行、诗行结构、押韵格式设置专属统计维度，通用计量标准难以完全适配诗歌文体特殊性。

2. 研究过程

2.1. 研究工具

WordSmith Tools是一款专业语料库检索与分析软件，是一套集成了Concord(语境共现检索)、WordList(词表生成)和Keyword(关键词分析)三大核心功能的综合性分析套件。本研究在词汇和句子层面主要借助

WordSmith 9.0 从类符形符比、平均句长等角度进行对比分析。

AntConc4.3.1 是一款免费、轻量级、跨平台的语料库检索与分析软件。本研究在语篇层面则主要借助 AntConc4.3.1 从显性连接与隐性连接两个维度对不同译本的语篇进行分析。

2.2. 研究对象

《长恨歌》的英译版本众多，本文选取了最具代表性的译本。笔者综合考量了许多因素如译者的国别、译者的翻译风格、译本的影响力、译本的出版时间等，最终选择了这四位译者的译本作为本文研究对象，分别是英国译者弗莱彻(W. J. B. Fletcher) [4]、翟里斯(Herbert Allen Giles) [5]和国内译者杨宪益和戴乃迭[6]、许渊冲[7]的译本，见表 1：

Table 1. Basic information of four English translations of *Song of Everlasting Sorrow*
表 1. 《长恨歌》四种英译文基本信息

译者	译者身份	翻译出版时间	英译名	译本	出版社
弗莱彻(W.J.B. Fletcher)	英国外交家、著名汉学家	1923 年	THE BALLAD OF ENDLESS WOE	<i>More Gems of Chinese Poetry</i>	THE COMMERCIAL PRESS
翟里斯(Herbert Allen Giles)	英国前驻华外交官、著名汉学家	1973 年	The Everlasting Wrong	<i>A History of Chinese Literature</i>	Tuttle Publishing
杨宪益、戴乃迭	中国翻译家、外国文学研究专家/英籍汉学家、翻译家	2001 年	Song of Eternal Sorrow	《唐诗 汉英对照》	北京：外文出版社
许渊冲	中国翻译家、诗人、大学教授	2014 年	The Everlasting Regret	《许渊冲英译白居易诗选 汉英对照》	北京：中国对外翻译出版有限公司

2.3. 研究方法

本研究整理筛选《长恨歌》四个英译本并清洗语料，先提取文本、搭建平行语料库，再用 Treecatcher 标注词性，依托 WordSmith 9.0 做数据统计，从词汇、句子、语篇多维定量分析，结合实例对比归纳四位译者的风格异同。

3. 基于语料库的译者风格分析

3.1. 词汇层面

在词汇层面将从词汇变异、词汇密度两方面展开，对《长恨歌》四个英译本进行统计分析。

3.1.1. 词汇变异

胡开宝(2011)指出词汇变异主要通过类符/形符比(TTR)体现，其探索的是给定文本中的范畴范围。类符指语料库中出现的重复的词数，形符指语料库中词汇的总数[8]。类符/形符比(TTR)与词汇变异呈正相关关系[9]。此外，在某种程度上，类型/形符比还能反映译者受源语言影响的程度，即该比例越低，译者就越接近英语使用者的语言习惯。经任敬敏[10]等学者论证所得当语料文本较长时，语料文本中的虚词数目会增多，从而会使类符/形符比的可信度降低，类符/形符比不能全面反映词汇的丰富程度。因此，本研究使用了当前语料库语言学领域普遍认可的标准化类符/形符比(STTR)，来衡量文本的词汇变异度，以更为准确地体现语料的用词丰富度。由于《长恨歌》源文本的形符数少于 1000，故将本研究中的标准类符/形符比的固定值设为 100。

Table 2. Statistics of standardized type/token ratio of the four English translations (Generated by WordSmith 9.0)
表 2. 四个英译本标准化类符/形符比统计表(by WordSmith 9.0)

	源语文本	弗译本	翟译本	杨、戴译本	许译本
类符数 (Types)	447	612	652	547	549
形符数 (Tokens)	580	1340	1433	1187	1160
类符/形符比 (TTR)	77.07%	45.67%	45.50%	46.08%	47.33%
标准化类符/形符比 (STTR)	94.80%	76.23%	73.50%	74.91%	76.55%

由表 2 数据可知, 许译本标准化类符/形符比高于翟译本, 说明许渊冲选词范围更广、词汇变化丰富, 译文笔法灵活。许译本总形符数偏少却拥有高类符形符比, 可见译者依靠优化用词丰富表达, 而非单纯扩充篇幅, 这得益于其深厚的中文功底。翟译本形符数量不多, 偏向直译, 常对原文内容多加阐释。受汉英语言特征影响, 汉语重意合、英语重形合, 英译篇幅普遍长于原文, 两个译本的形符数据均契合这一语言规律。

Example 1

遂令天下父母心, 不重生男重生女。

许译本: Of fathers and mothers who would rather give birth. To a fair maiden than to any son on earth.

翟译本: For thus it came about that fathers and mothers through the length and breadth of the empire. Rejoiced no longer over the birth of sons, but over the birth of daughters.

杨、戴译本: All parents wished for daughters
Instead of sons!

弗译本: Until throughout the Empire
All parents hailed with joy
The birth of some fair maiden;
And wanted not a boy.

“天下”是东亚国家特有的概念, 其含义等同于“普天之下”, 在地理和空间范围上都没有限制。从这个意义上来说, “天下”所涵盖的区域并不局限于帝国的领土, 这一点可能让西方人难以理解。在这种情况下, 翟里斯和弗莱彻将其直译为“the length and breadth of the empire”(帝国的广袤疆域)和“throughout the empire”(整个帝国), 译法虽然不够地道, 但其生动的描述更容易被西方读者接受。相比之下, 许渊冲先生的“on earth”(在世间)最符合上述所讨论的范围。与其他人相比, 杨虽然省略了前半句, 直接翻译了后半句, 但他的表达清晰, 目标读者相对容易理解。

3.1.2. 词汇密度

词汇密度反映语料中实义词所占比重。胡壮麟将英语词汇划分为实义词与功能词。英语实义词通常指具有实在意义的词语, 包括名词、动词、形容词和副词[11]。1995 年莫娜·贝克(Mona Baker) [12]指出词汇密度一般指实词在一个文本总词数中所占的百分比, 是文本信息量大小的一个衡量标准。词汇密度是衡量文本信息量与语体特征的重要指标。密度越高, 意味着实词在总词数中的占比越大, 文本所传递的信息量也越丰富。同时, 词汇密度也与语体的正式程度呈正相关, 密度越高, 通常表明文本越正式。

由表 3 可知，许译本的词汇密度最高，翟译本的词汇密度最小，这说明许译本所承载的信息量相对更大，语体也相对更为正式，对源文本表达比较准确。而翟译本的词汇密度最低，说明翟里斯考虑到目标语读者的接受度，通过增加功能词以简化译文，使译文更加通俗易懂。但也在一定程度上反映出作为非母语者，翟里斯对源文本的理解不如母语为汉语的本土译者。此外，四个译本的实词数量由大到小分别是名词、实义动词、形容词、副词，这也符合英语的表达习惯。

Table 3. Statistics of word classes and lexical density of the four English translations (Generated by AntConc)
表 3. 四个英译本词性及词汇密度统计表(by AntConc)

译本	名词数量/个	实义动词数量/个	形容词数量/个	副词数量/个	实词数量/个	词汇密度/%
弗译本	378	221	96	83	778	58.06
翟译本	408	192	96	99	795	55.48
杨、戴译本	360	190	106	53	709	59.73
许译本	336	189	102	69	696	60.00

注：常用的词汇密度计算公式为：词汇密度 = 实词数/总词数 × 100%。

3.2. 句子层面

句子也是分析文本的重要维度之一，本文在句子层面将对《长恨歌》四个英译本的平均句长进行统计分析。

平均句长

一般而言，句子数量和平均句长可以作为译者风格研究的重要参数[13]。一般而言，平均句长越长，文本的句子结构就越复杂，句子的理解难度越高。

Table 4. Sentence statistics of the four English translations (Generated by WordSmith 9.0)
表 4. 四个英译本句子统计表(by WordSmith 9.0)

译本	句子个数/句	平均句长/词	句子长度标准差
弗译本	101	13.27	3.63
翟译本	53	27.04	1.15
杨、戴译本	84	14.13	3.63
许译本	53	21.89	4.38

从表 4 可知平均句长由大到小依次是翟译本、许译本、杨、戴译本和弗译本。

Example 2

行宫见月伤心色，夜雨闻铃肠断声。

许译本：The moon viewed from his tent shed a soul-searing light;
The bells heard in night rain made a heart-rending sound.

杨、戴译本：In his refuge when he saw the moon, Even it seemed sad and wan.
On rainy nights, the sound of bells Seemed broken-hearted.

翟译本：Travelling along, the very brightness of the moon saddens his heart,
And the sound of a bell through the evening rain severs his viscera in twain.

弗译本: When from his tent the Moon he sees,
His breast is charged with woe.
The rain of night,
the watches' bell.

从例 2 数据可见, 翟译本平均句长最长, 弗莱彻译本最短, 数据差异源于二者翻译理念不同。翟理斯身为英国汉学家, 素来推崇典雅韵体直译, 主张译诗严守诗体、保留韵律, 但翻译《长恨歌》时打破惯例: 译文行文工整、用词考究却不押韵。为还原诗作叙事内容、补充历史背景, 他舍弃诗歌审美与押韵要求, 因此译文句子偏长。弗莱彻深受翟理斯译论影响, 坚持译诗要押韵、守格律、忠于原文。翻译时采用诗体排布, 摒弃大段划分, 依照原诗联句拆分小节, 每节四句, 句式规整、押韵严谨, 句式更为凝练, 最终译本平均句长在四个译本中最低。

3.3. 语篇层面

在语言学领域, 英语和汉语的区别主要体现在意合和形合上。韩礼德(Halliday, M)认为, 汉语主要是意合的。也就是说, 汉语中一组句子的逻辑意义不是通过使用连接词来体现, 而是通过语境来传达。此外, 汉语中衔接手段较少, 非常注重隐性连贯和语境意义。英语主要是形合的[14]。

“英语的形合通常通过各种关系词、连接词、介词、小品词短语等体现[14]。”在《长恨歌》的源文本中, 虽然没有使用明确的连接词来表明逻辑关系, 但读者仍能理解诗句中的语法含义。然而, 将这首诗翻译成英语时, 译者需要添加一些显性连接词, 如 “and, but, since, if, or” 来阐明逻辑关系。

3.3.1. 显性连接

Table 5. Statistics of conjunctions in the four English translations (Generated by AntConc)
表 5. 四个英译本连词统计表(by AntConc)

译本	并列连词数量/个	从属连词数量/个	并列连词与从属连词在全文总词数中的占比/%
弗译本	61	168	17.09
翟译本	66	226	20.38
杨、戴译本	39	162	16.93
许译本	53	167	18.97

Table 6. Statistics of pronouns in the four English translations (Generated by AntConc)
表 6. 四个英译本代词统计表(by AntConc)

译本	人称代词数量/个	物主代词数量/个	人称代词与物主代词在全文总词数中的占比/%
弗译本	132	61	14.40
翟译本	104	44	10.33
杨、戴译本	112	50	13.75
许译本	192	39	19.91

从表 5、表 6 可知四位译者均在译文中不同程度地增加了显性连接词。其中许译本许渊冲的译文在显性连接成分的使用上明显多于另外三个译本。

Example 3

侍儿扶起娇无力，始是新承恩泽时。

杨、戴译本：When her maids helped her rise,
She looked so frail and lovely,
At once she won the emperor's favour.

许译本：Without her maids' support, she was too tired to move.
And this was when she first received the monarch's love.

翟译本：Then, when she came forth, helped by attendants, her delicate and graceful movements,
Finally gained for her gracious favour, captivating his Majesty's heart.

弗译本：When thence attendants bore her,
So helpless and so fair;
Then first beat in her Prince's breast
Desire and tender care.

例 3 数据显示，许译本频繁使用连词，依托英语形合特点实现以形传神。大量连接词使译文契合译入语表达与文化特点，与其偏重意译的翻译思想相符。许渊冲主张翻译贵在神似而非形似，古典诗词富含意境韵律，意译能够挣脱原文句式桎梏，留存原作情思与美感。他力求译作让西方读者读懂、赏美、体味韵味，而意译恰好可以消解中西语言文化隔阂，实现这一翻译追求。

3.3.2. 隐性连接

所谓的隐性联系是指那些并不需要借助连词或者连接语就能被理解的关系，它们主要是由语言中的自然逻辑顺序所决定的。然而，英文的特点在于其形式化，因此许多这样的关联通常会以标点符号的形式出现，例如句号“.”、“？”、“！”、“，”和分号等。

Table 7. Statistics of implicit cohesive devices in the four translations (Generated by AntConc)
表 7. 四个译本隐性连接成分统计表(by AntConc)

译本	弗译本	翟译本	杨、戴译本	许译本
“；”	21	10	11	17
%	22.58	21.82	18.97	31.58

由表 7 可知，许译本的隐性连接成分占比最大，弗译本、翟译本以及杨、译本数值比较接近，这是由于翟里斯、弗莱彻、戴乃迭都是英语母语者，语言习惯相近。且杨、戴译本的数值最小，这表明他的译文更注重意义的传达，译文结构相对简单，翻译风格偏向直译。在 20 世纪中西方文化交流匮乏背景下，杨宪益、戴乃迭翻译核心目标是促进中西方文化交流，而直译能减少译者主观解读带来的偏差，为西方读者提供最贴近原文的文本，助力文化传播的准确性。

4. 结论

本文依托自建《长恨歌》四译本语料库，从词、句、语篇开展定性与定量分析，探析四位译者的翻译风格。本研究结果表明，四位译者呈现出迥异的翻译风格，差异成因可归为外部时代译介环境与译者个人翻译理念两大层面。

其一，不同时代译诗传统与出版受众需求，直接影响译者翻译策略。弗莱彻、翟里斯活跃于维多利

亚晚期至 20 世纪初, 彼时英语读者偏爱符合本土诗歌审美的流畅文本, 汉诗英译普遍简化异域文化表达, 适配英文阅读节奏。二人曾任驻华外交官, 译诗面向普通西方读者普及中国古典诗歌, 因此多选用贴合英语表达的译法, 删减复杂文化表述, 降低阅读理解难度。 “20 世纪的大部分时间里, 归化翻译占据着主导地位; 但是到了 21 世纪, 局面将会有所改变, 随着国际间文化交流的日益频繁, 随着各国人民之间的不断沟通, 向作者接近的异化译法将越来越广泛地被采用, 与归化译法取得平衡, 甚至可能占据上风 [15]。” 杨宪益、戴乃迭 2001 年推出译本, 恰逢翻译策略多元发展阶段, 中外文化互通渠道拓宽, 译文兼顾原诗意蕴与英文行文规范, 不会单一偏向归化或异化。许渊冲的译介工作扎根国内古典诗词英译研究环境, 学界注重完整还原诗歌意象与文化底蕴, 因此其译文着力保留唐诗独有的韵律与文化内涵, 不会为迎合外国读者随意改写原文。

其二, 译者的文学积累、审美取向与翻译准则, 塑造各自遣词、造句、谋篇的语言风格。弗莱彻与翟里斯熟习维多利亚叙事诗创作范式, 受当时文学风潮影响, 译文中侧重描摹人物心理活动, 句式排布贴合英语诗歌传统, 优先保障阅读流畅度。杨宪益自幼研习文史, 搭配戴乃迭的母语英语优势, 二人以忠实还原典籍原貌为翻译目标, 行文克制简练, 极少增加主观抒情修饰, 句式严格贴合原诗叙事脉络, 避免过度扩充。许渊冲长期对比中西诗歌, 没有直接套用西方对等翻译理论, 自主提出兼顾意美、音美、形美的“三美原则”。翻译《长恨歌》时, 他灵活调整句式、补充逻辑连接词, 以意译还原诗歌意境, 平衡原作审美韵味与英文文本可读性。

参考文献

- [1] 胡开宝, 田绪军. 基于多语语料库的翻译研究: 议题与意义[J]. 北京第二外国语学院学报, 2023, 45(2): 3-17, 101.
- [2] Baker, M. (1993) Corpus Linguistics and Translation Studies—Implications and Applications. In: Baker, M., Francis, G. and Tognini-Bonelli, E., Eds., *Text and Technology*, John Benjamins Publishing Company, 233-250. <https://doi.org/10.1075/z.64.15bak>
- [3] Baker, M. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. *Target. International Journal of Translation Studies*, 12, 241-266. <https://doi.org/10.1075/target.12.2.04bak>
- [4] Fletcher, W.J.B. (1923) *More Gems of Chinese Poetry*. The Commercial Press Limited.
- [5] Giles, H.A. (1973) *A History of Chinese Literature*. Tuttle Publishing.
- [6] Yang, X.Y. and Yang, G. (2001) *Poetry and Prose of the Tang and Song*. Foreign Languages Press, 105 p.
- [7] 许渊冲. 许渊冲英译白居易诗选[M]. 北京: 中国对外翻译出版有限公司, 2014: 8-9.
- [8] 胡开宝. 语料库翻译学概论[M]. 上海: 上海交通大学出版社, 2011.
- [9] 王克非. 语料库翻译学探索[M]. 上海: 上海交通大学出版社, 2012.
- [10] 任敬辉. 基于语料库的《长恨歌》译者文体研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2013.
- [11] 胡壮麟. 理论文体学[M]. 北京: 外语教学与研究出版社, 2002.
- [12] Baker, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target. International Journal of Translation Studies*, 7, 223-243. <https://doi.org/10.1075/target.7.2.03bak>
- [13] 胡开宝, 谢丽欣. 基于语料库的译者风格研究: 内涵与路径[J]. 中国翻译, 2017, 38(2): 12-18, 128.
- [14] Halliday, M. and Hasan, R. (1976) *Cohesion in English*. Longman.
- [15] 孙致礼. 翻译: 理论与实践探索[M]. 南京: 译林出版社, 1999.