

基于语料库的《老人与海》译者风格研究

汪菁雅

上海海事大学外国语学院, 上海

收稿日期: 2026年3月13日; 录用日期: 2026年4月10日; 发布日期: 2026年4月22日

摘要

本研究基于自建的《老人与海》英汉一对二平行语料库, 采用定量与定性相结合的方法, 对比分析了吴劳和余光中两位译者的风格。研究从词汇、句法、语篇三个层面展开, 考察了标准类符形符比、词汇密度、句子总数与平均句长、被动式使用、衔接手段以及语篇情感等多个语言特征。研究发现, 两位译者的风格存在显著差异: 吴劳的译本在显化程度、欧化程度上更高, 更忠实于原文的形合特点; 余光中的译本则词汇更丰富, 句式更灵活多变, 更贴近汉语“意合”的语言习惯。研究认为, 语料库驱动的量化分析方法能够使译者风格研究更为科学, 为定性分析提供实证支持。

关键词

《老人与海》, 平行语料库, 译者风格

A Corpus-Based Analysis of Translator Style in *The Old Man and the Sea*

Jingya Wang

College of Foreign Languages, Shanghai Maritime University, Shanghai

Received: March 13, 2026; accepted: April 10, 2026; published: April 22, 2026

Abstract

This study, based on a self-constructed English-Chinese one-to-two parallel corpus of *The Old Man and the Sea*, employs a combination of quantitative and qualitative methods to compare and analyze the translator styles of Wu Lao and Yu Guangzhong. The study is carried out from three levels—lexical, syntactic, and discourse—examining a range of linguistic features, including the standardized type-token ratio, lexical density, total number of sentences and average sentence length, usage of passive constructions and cohesive devices, and sentiment. The findings reveal significant differences between the two translators: Wu Lao's version demonstrates a higher degree of explicitation

and Europeanization and is more faithful to the hypotactic features of the source text; Yu Guangzhong's translation, by contrast, conforms more closely to the paratactic conventions of Chinese, featuring richer vocabulary and more flexible and varied sentence structures. The study suggests that corpus-driven quantitative analysis enables a more scientific study of translator style and provides empirical support for qualitative analyses.

Keywords

The Old Man and the Sea, Parallel Corpus, Translator Style

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

《老人与海》(*The Old Man and the Sea*)是美国作家欧内斯特·米勒尔·海明威(Ernest Miller Hemingway)最具代表性的作品之一。小说讲述了老年渔夫圣地亚哥在连续八十四天未能捕到鱼的情况下,仍坚持独自出海捕鱼,并最终钓到一条巨大的马林鱼的故事。作为海明威的经典之作,这部作品在世界范围内产生了深远影响,在中国的译介也十分活跃。其中张爱玲译本是第一个汉译本,海观、吴劳、余光中和孙致礼等知名作家和学者也陆续出版了译作。

这些经典译作产生的时代多有不同,也发生过译者重译的情况,在时代背景、文化背景以及翻译观的影响下,不同译本得以体现不同译者的风格。Baker认为,相对于其他译者,风格是译者对语言的个性化使用,是个人的语言习惯;从语言表达方面,是一种个性特色,同时也是非语言特征[1]。尽管译者风格的定义尚未统一化,但学界普遍认为译者风格是指译者在翻译过程中,形成了区别于其他译者且体现于译者本人不同译作中的翻译选择[2],这些翻译选择是译者的语言运用和非语言特征的综合体现。译者风格使某位译者作品区别于其他译者作品,并且表现于其不同译作,无论是目的语词汇和句式结构的选择或语篇结构的重构,还是翻译策略和翻译方法的应用,都必然会打下译者的烙印,从而形成译者风格[3]。

由此可见,译者的语言特征能够通过译作的词汇、句法和语篇等层面加以体现,通过对语言特征的系统考察,译者风格的识别更具客观性。本研究基于语料库辅助,选取《老人与海》吴劳和余光中两译本进行译者风格分析,吴劳译本流传广泛、颇具权威,余光中译本则经历了译者的重译,两个译本均具有一定的代表性,适合应用于对比分析。

2. 基于语料库译者风格研究文献综述

20世纪90年代后,语料库开始被应用于译者风格研究,Baker首次提出了“译者风格”这一概念,将语料库用于译者风格研究当中。近年来,国内基于语料库的译者风格研究逐渐成为语料库研究的重要课题之一,借助语料库研究方法,能够对译者风格进行更为系统的分析,关注译者的主体性及其无意识的风格特征。整体来看,译者风格研究正逐步走向更加客观的发展方向。在Baker的译者风格研究基础上,Gabriela Saldanha将研究方法分为源语取向视角和目的语取向视角[4]。

回顾过去二十年国内基于语料库的译者风格研究情况,源语取向视角的研究占多数,侧重原文语言特征在译文中的再现,一般采用双语平行语料库。张丹丹(2014)重点考察了《红楼梦》乔利译本前24回

与后 32 回在韵式、句子、语境等三个层面的真实风貌, 研究发现乔利译本前后风格迥异, 得出其前后译者可能并非同一人, 此外, 该研究还以《红楼梦》其他三个英译本作为参照进行对比分析[5]。刘泽全(2017)自建《老人与海》英汉平行语料库, 采用定量与定性结合的方法考察了原文及六个译本, 从对原文的语言风格的系统化分析出发进行不同译本的风格对比, 并与语境相结合深入考察文体风格[6]。韩丹(2023)自建《阿 Q 正传》两个俄译本语料库, 对比分析两位译者在词汇、句法层面的计量特征和文化负载词翻译策略的异同, 以数据证明了前人研究的部分结论存在疑问, 研究表明这些计量特征相差无几, 但在原文各章节中存在明显差异[7]。目的语取向视角侧重译者自身的风格, 研究通常基于同一个译者的多个作品进行考察分析, 且一般采用单语语料库, 并使用对照语料库作为参照。黄立波(2012)考察葛浩文的十部中国小说英译本并以戴乃迭的翻译作品为参照, 探究两位译者的翻译风格差异, 研究认为 S-型译者风格的关注对翻译研究的意义更为深远[8]。祝朝伟(2023)通过自建庞德翻译作品语料库和庞德原创作品语料库, 从词汇、句法和语篇三个层面比较庞德的翻译作品与原创作品, 对于身兼译者和作者双重身份的译者, 对比分析其翻译作品和非翻译作品是译者风格研究的有益尝试, 有助于拓展译者风格研究的视域[2]。

在既有研究中, 学界对《老人与海》的译者风格研究普遍集中于吴劳与张爱玲等译者之间的对照, 而对吴劳与余光中的对比研究相对较少。在研究方法上, 早期研究更多依赖主观判断的定性分析, 并未充分借助量化的手段。随着语料库语言学的发展, 近年来研究虽然开始引入定量方法, 但多集中于词汇和句法层面, 考察的指标不够全面, 特别是对语篇层面的考察相对不足。基于此, 本研究在综合运用定量与定性方法的基础上, 从词汇、句法和语篇三个层面对吴劳与余光中的译者风格进行对比分析, 以弥补既有研究在研究对象与分析层面上的不足。

3. 研究设计

3.1. 研究语料和研究问题

本研究选取的语料源于《老人与海》英文原文、2009 年上海译文出版社出版的吴劳译本和 2010 年译林出版社出版的余光中译本, 研究采用了三个版本的正文部分。本研究采用语料库分析方法, 基于源语取向视角, 探究两位译者翻译风格的差异, 提出具体研究问题, 即两版译文在词汇、句法和语篇层面体现出怎样的译者风格?

3.2. 研究方法

本研究通过自建英汉一对二平行语料库对《老人与海》吴劳译本(以下简称吴译本)与余光中译本(以下简称余译本)进行译者风格分析。首先将原文和译文材料转化为软件可识别的 txt 格式, 使用文本编辑器和 Emeditor 分别对英文文本和中文文本进行清洁降噪, 接着使用 ABBYY Aligner 对英汉文本进行双语语料对齐, 对齐后导出为 docx 格式, 通过转化得到英汉共 3 个 txt 格式的文本文件。使用 TreeTagger 和 CorpusWordParser 分别对英文文本和中文文本进行分词赋码, 最后使用 WordSmith 和 AntConc 进行检索统计, 统计内容包括类符形符数、词汇密度、句子总数、平均句长、特定词语等, 并利用 UCREL Significance Test System 平台进行显著性检验。结合具体数据, 从词汇、句法和语篇三个层面进行两个译本的译者风格分析。

3.3. 研究工具

本研究采用的研究工具包括软件 Word Office 16、文本编辑器 3.3.6.1、Emeditor14.4.4、TreeTagger2.0、CorpusWordParser3.0、ABBYY Aligner1.0.6.59、WordSmith8.0、AntConc3.3.0, 以及在线平台 UCREL Significance Test System。

4. 研究结果与分析

4.1. 词汇层面

4.1.1. 标准类符形符比

类符形符比(TTR)是文本中类符数与形符数的比值,是衡量文本词汇丰富度的重要指标。然而,类符形符比容易受到语篇长度的影响,语篇越长该比值就越小。因此,当两个文本长度不等时,类符形符比无法准确反映出二者词汇丰富度的差异,因而学界通常采用计算每个文本的标准化类符形符比(STTR)的做法[7]。STTR 值越高,词汇变化就越丰富。研究使用 WordSmith8.0 统计《老人与海》原文和译文的标准类符形符比,结果如表 1 所示。

Table 1. Statistical analysis of types and tokens in three texts

表 1. 原文和两译本类符形符统计数据

类别	英文原文	吴译本	余译本
类符	2528	3950	3909
形符	26,690	26,759	23,501
标准类符形符比	34.52%	45.39%	47.15%

原文类符数为 2528,形符数为 26,690,标准类符形符比为 34.52%;吴译本类符数为 3950,形符数为 26,759,标准类符形符比为 45.39%;余译本类符数为 3909,形符数为 23,501,标准类符形符比为 47.15%。通过利用 UCREL Significance Test System 对数据进行显著性检验,能够计算出对应的 P 值。一般来说,P 值越接近 0,说明结果的差异性越大。当 P 值<0.01 时,能够佐证样本之间存在显著差异。此处对三组类符数据进行两两卡方检验,得出吴译本与原文类符数相比的卡方值为 274.87, $P < 0.01$,具有显著差异;余译本与原文类符数相比的卡方值为 444.07, $P < 0.01$,具有显著差异;两译本类符数相比的卡方值为 24.10, $P < 0.01$,具有显著差异。前贤的研究普遍认为,译文文本的长度会超过原文,这是语际显化的特征之一[6]。从形符数量上看,吴译本高于余译本,可见吴译本的显化程度也更高。如例 1:

例 1:

原文: It was getting into the afternoon and the boat still moved slowly and steadily. But there was an added drag now from the easterly breeze and the old man rode gently with the small sea and the hurt of the cord across his back came to him easily and smoothly.

吴译: 时间已是下午,船依旧缓慢而稳定地移动着。不过这时东风给船增加了一份阻力,老人随着不大的海浪缓缓漂流,钓索勒在他背上的感觉变得容易忍受而平和了些。

余译: 渐渐到了下午,小船依旧缓缓不断地前进。这时东风加重了拉力,老人在微波的海上乘浪航行,背上的绳伤也感觉比较温和、好受。

例 1 中,吴译本忠实地译出了原文的全部信息,在处理“added drag now from the easterly breeze”时,对原文没有的受力对象即船体进行了增译。此外,后半句中对“the cord across his back”的翻译,增加了动词“勒”进行补充说明,对原文信息的阐释更加细致,进一步表示吴译本的显化程度更高。

两个译本的标准类符形符比均远高于原文,说明译文较原文词汇更丰富,用词变化性更大。其中余译本的标准类符形符比略高于吴译本,可知余译本的词汇丰富度也高于吴译本。如例 2:

例 2:

原文: "I told the boy I was a strange old man," he said. "Now is when I must prove it."

吴译: "我跟那男孩说过来着, 我是个不同寻常的老头儿," 他说。"现在是证实这话的时候了。"

余译: "我对那孩子说过, 自己是个老精灵," 他说, "现在正好证明。"

贯穿原文的人物及称谓 "old man" 在不同译本中的处理方式虽然存在差异, 但在绝大多数译本中仍是以 "老人"、"老头子" 等为主。在本例中, 对于原文 "strange old man" 的表达, 余译本将其处理为 "老精灵", 这种称谓比之吴译的 "老头儿" 更加鲜活诙谐, 丰富了同一词汇翻译方式的选择。

4.1.2. 词汇密度

词汇密度是文本中实词类符数与文本总形符数的比值。文本信息量大小与词汇密度成正比, 词汇密度越高, 文本信息量越大, 其阅读难度也越大, 反之则越小[6]。在词类频率统计中, 通过计算实词形符在总形符中所占比例, 可得出文本的词汇密度, 以此反映文本信息量大小[9]。研究使用 Antconc3.3.0 统计《老人与海》原文和译文的实词数, 根据 Ure 提出的词汇密度计算公式 "词汇密度 = 实词数/总词数 × 100%" [9], 计算结果如表 2 所示。

Table 2. Statistical analysis of lexical density in three texts

表 2. 原文和两译本词汇密度统计数据

类别	英文原文	吴译本	余译本
实词	12,585	16,923	15,797
形符	26,690	26,759	23,501
词汇密度	47.15%	63.24%	67.21%

原文实词数为 12,585, 总形符数为 26,690, 词汇密度为 47.15%; 吴译本实词数为 16,923, 总形符数为 26,759, 词汇密度为 63.24%; 余译本实词数为 15,797, 总形符数为 23,501, 词汇密度为 67.21%。通过对三组词汇密度数据进行两两卡方检验, 得出吴译本与原文词汇密度相比的卡方值为 404.56, $P < 0.01$, 具有显著差异; 余译本与原文词汇密度相比的卡方值为 565.76, $P < 0.01$, 具有显著差异; 两译本词汇密度相比的卡方值为 18.32, $P < 0.01$, 同样具有显著差异。由此可知, 两译本的词汇密度均远高于原文, 说明译本的文本信息量大于原文, 阅读难度也随之加大, 进一步体现了翻译文本的语际显化特征。同时, 余译本的词汇密度高于吴译本, 说明余译本中实词占比更高, 文本负载的信息量更大。

4.2. 句法层面

4.2.1. 句子总数与平均句长

句法层面的语言特征通常通过平均句长体现。一般来说, 句子总数越少, 平均句长越大, 则句式越复杂, 长难句越多。平均句长是总形符数除以句点标记(包括句号、问号和感叹号等)的个数得出的数据[10]。研究使用 WordSmith8.0 统计《老人与海》原文和译文的句子总数和平均句长, 结果如表 3 所示。

原文共有 1781 句, 平均句长为 14.75; 吴译本共有 1832 句, 平均句长为 14.61; 余译本共有 1736 句, 平均句长 13.54。通过对三组平均句长数据进行两两卡方检验, 发现两译本的平均句长与原文相比的卡方值均接近 0, $P > 0.05$, 不能判断有显著差异。鉴于海明威的电报式写作风格以简洁干练为特色, 两位译者在翻译中都把控了原文的这一风格。二者的平均句长仍然存在差异, 如例 3:

Table 3. Statistical analysis of sentence length in three texts**表 3.** 原文和两译本句长统计数据

类别	英文原文	吴译本	余译本
句子数	1781	1832	1736
平均句长	14.75	14.61	13.54

例 3:

原文: He was an old man who fished alone in a skiff in the Gulf Stream and he had gone eighty-four days now without taking a fish.

吴译: 他是个独自在湾流中一条平底小帆船上钓鱼的老人, 这一回已去了八十四天, 没逮上一条鱼。

余译: 那老人独驾轻舟, 在墨西哥湾暖流里捕鱼, 如今出海已有八十四天, 仍是一鱼不获。

英汉语言在句子结构上的不同, 往往导致两种语言在静态与动态表达方式上存在差异。在本例中, 吴译保留了原文的结构与顺序, 将前半句从句“who fished alone in a skiff in the Gulf Stream”处理为大段定语修饰主语, 句式变化较小; 余译则灵活地拆分了较长的前半句, 强调了动作, 更符合中文的行文习惯, 句式变化较大。结合数据, 吴译本的平均句长略高于余译本, 进一步说明吴译本更习惯使用长句, 句式更复杂, 而余译本更多使用短句, 整体来看可读性更高。

4.2.2. 被动式

在现代汉语中, “把”字句与“被”字句应用广泛, 同时也在一定程度上反映译者的语言欧化程度。被动式在汉语中常有消极意义, 不是所有的陈述句都有被动式, 而欧化翻译则不再依照这一习惯, 常常保留原文中表示中性或积极意义的被动语态[11]。通过 Antconc3.3.0 检索发现, 吴译本中“被”字使用频次为 40, 余译本中“被”字使用频次为 4, 如表 4 所示。将这些被动句按积极、中性和消极意义分为三类, 尽管中性和消极意义占比较高, 但吴译本中“被”字使用频次仍远远高于余译本。由此可见, 吴译本的欧化程度高于余译本, 余译本则更趋近于汉语的传统表达方式。

Table 4. Statistical analysis of three semantic categories of the “bei” in two translations**表 4.** 两译本“被”字的三类语义统计数据

类别	吴译本	余译本
积极	2	1
中性	25	2
消极	13	1
共计	40	4

4.3. 语篇层面

4.3.1. 衔接与连贯

衔接是语篇的重要特性, 体现在结构上, 通过词汇和语法手段得以实现; 连贯是语篇的主要特性, 通过逻辑推理来实现语义连接[12], 因此衔接与连贯在构成语篇时紧密联系。翻译时, 译者要识别原语篇的各种衔接手段和连贯性, 同时还需要在译文语篇中, 合理运用衔接手段, 重构原文的连贯性[13]。汉语

是意合语言，句与句之间的过渡较为隐含，而英语语句之间主要采用形合手段实现连接。语篇衔接通常可以依靠使用代词和连接词等词汇或语法手段。通过 Antconc3.3.0 统计《老人与海》原文和译文中连词和代词的使用频次，结果如表 5 所示。

Table 5. Statistical analysis of conjunctions and pronouns in three texts

表 5. 原文和两译本连词和代词统计数据

类别	英文原文	吴译本	余译本
连词	1571	755	536
代词	4211	3654	3059
共计	5782	4409	3595
全文中占比	21.66%	16.48%	14.45%

原文中连词频数为 1571，代词频数为 4211，合计 5782；吴译本连词频数为 755，代词频数为 3654，合计 4409；余译本连词频数为 536，代词频数为 3059，合计 3595。通过对三组连词和代词总频数进行两两卡方检验，得出吴译本与原文连词和代词总频数相比的卡方值为 90.77， $P < 0.01$ ，具有显著差异；余译本与原文连词和代词总频数相比的卡方值为 123.73， $P < 0.01$ ，具有显著差异；两译本连词和代词总频数相比的卡方值为 5.85， $P > 0.05$ ，不能判断有显著差异。

结合在全文中占比的数据，发现两译本的连词和代词使用频数占比均远低于原文，对应了英语和汉语语言在衔接与连贯上的差异，即英语重形合，需要依靠连词和代词实现衔接，汉语重意合，衔接主要体现在意义上。两译本之间比较，余译本的连词和代词使用频数占比低于吴译本，利用词汇和语法手段来实现衔接与连贯的情况少于吴译本，因此余译本更符合汉语意合的语言习惯。

4.3.2. 语篇情感

翻译能体现作者蕴藏的情感寄托，译本中一些不表示具体“意念”和“概念”的指征(感叹词、语气词以及感叹号和问号)的使用在一定程度上体现了译者对原文情感的转换，能够反映译者的风格特征[14]。研究使用 Antconc3.3.0 检索《老人与海》原文和译文中感叹词等的使用频次，如表 6 所示。

Table 6. Statistical analysis of interjections and other marks in three texts

表 6. 原文和两译本感叹词等统计数据

类别	英文原文	吴译本	余译本
感叹词	14	76	19
感叹号和问号	94	92	99
共计	108	168	118
全文中占比	0.40%	0.63%	0.50%

经过统计，发现原本文中感叹词等总使用频次为 108，占全文比例 0.40%；吴译本中感叹词等总使用频次为 168，占全文比例 0.63%；余译本中感叹词等总使用频次为 118，占全文比例 0.50%，三文本的统计结果在全文中占比均较低，其中两译本的总占比均略高于原文，说明译本的情感传达更浓厚。从标点

符号角度来看,三文本使用感叹号和问号的频次差异不大,说明两位译者都没有过多改变原文中的非陈述语气,较为完好地保持了原文整体较为冷静沉稳的陈述方式。从感叹词角度来看,吴译本的感叹词使用频次远高于余译本,多是在句末使用感叹词为主,说明吴译本的口语化程度更高,情感表达更明显。

5. 结语

本研究通过自建《老人与海》英汉一对二语料库,分别统计标准类符形符比、词汇密度、句子总数、平均句长、连词和代词以及感叹词等使用情况,从词汇、句法和语篇三个层面分析吴译本和余光中译本的译者风格。在词汇层面,吴译本的显化程度更高,而余译本用词更丰富,且实词占比更高,文本的信息量更大。在句法层面,吴译本句子较长较为复杂,而余译本的翻译方式更加灵活,此外,吴译本更多使用被动态,其欧化程度更高。在语篇层面,余译本更符合汉语意合的语言规范,而吴译本的口语化程度更高,情感表达更明显。总体而言,研究证实两位译者的译者风格存在较为显著的差异,同时也受到原文风格的影响。吴译本采用更忠实原文的翻译,余光中译本则更贴近译入语的语言习惯。

《老人与海》在译介之初就显示出有别于同时代欧美作品的独特优势,几十年来不断重译的道路上获得重生,显示出旺盛的生命力,其涉及出版社数量之多、覆盖范围之广、参与译者之众,在整个外国文学翻译史的长河中都蔚为壮观[15]。由此可见,海明威的这部杰作对于译者风格研究的价值极高,其深远影响跨越时代,至今仍值得学界深入探索。基于语料库的译者风格研究可以从量化分析的视角对译者的独特风格进行统计分析,为定性的结论提供有力的数据支持。

参考文献

- [1] Baker, M.J. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. *Target*, 12, 241-266. <https://doi.org/10.1075/target.12.2.04bak>
- [2] 祝朝伟, 李润丰. 基于语料库的庞德中国典籍英译译者风格探析[J]. 外语教学, 2023, 44(4): 75-82.
- [3] 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.
- [4] Saldanha, G.J. (2011) Translator Style: Methodological Considerations. *Translator*, 1, 25-50. <https://doi.org/10.1080/13556509.2011.10799478>
- [5] 张丹丹, 刘泽权. 《红楼梦》乔利译本是一人所为否?——基于语料库的译者风格考察[J]. 中国外语, 2014, 11(1): 85-93.
- [6] 刘泽权, 王梦瑶. 量性结合的《老人与海》及其六译本风格对比考察[J]. 外语教学, 2017, 38(5): 73-79.
- [7] 韩丹, 张志军. 基于语料库的《阿Q正传》俄译本翻译风格对比研究[J]. 外语与外语教学, 2023(1): 105-116.
- [8] 黄立波, 朱志瑜. 译者风格的语料库考察——以葛浩文英译现当代中国小说为例[J]. 外语研究, 2012(5): 64-71.
- [9] Ure, J.J. (1971) Lexical Density and Register Differentiation. In: *Applications of Linguistics*, Cambridge University Press, 443-452.
- [10] 胡开宝, 朱一凡, 李晓倩. 语料库翻译学[M]. 上海: 上海交通大学出版社, 2018: 163-167.
- [11] 张威, 孙泽瑞. 基于语料库的《李尔王》三译本欧化现象比较研究[J]. 渤海大学学报(哲学社会科学版), 2020, 42(2): 106-110.
- [12] 杨黎霞. 语篇语言学与翻译[J]. 中国科技翻译, 2003, 16(3): 1-4+55.
- [13] 杨惠英. 语篇翻译中的衔接与连贯[J]. 现代英语, 2022(20): 53-56.
- [14] 赵宇霞. 基于语料库的傅雷翻译风格新探: 语言与情感的融合[J]. 外语电化教学, 2022(2): 96-103.
- [15] 高存. 《老人与海》在中国的译介[J]. 北京第二外国语学院学报, 2016, 38(2): 68-79+136.