

基于语料库计量语言特征的译者风格对比研究——以《归园田居》三译本为例

雷成真

贵州大学外国语学院, 贵州 贵阳

收稿日期: 2024年4月22日; 录用日期: 2024年6月6日; 发布日期: 2024年6月17日

摘要

本文以陶渊明《归园田居》五首诗的汪榕培、方重、许渊冲英译本自建单语类比语料库, 采用熵值、关键词共现网络、连词使用等手段对三译本的词汇、句子、语篇三个方面进行量化分析, 在此基础上对不同译者风格产生的成因进行解读。研究发现, 汪译本倡导“传神达意”的翻译方法, 倾向还原原作的意象和主题, 凸显了“顺应自然、纯朴”的生态主题, 句式结构较为复杂, 但其音译的异化策略会造成理解困难; 方译本基于研究型翻译理念, 使用类比方式的显化策略来体现“追求自然, 厌恶世俗”的主题, 重视环境描写来勾勒陶渊明的生态追求; 许译本则遵循“三美”原则, 用词与句式表达更为灵活, 显化的翻译策略呈现了“享受自然, 任性存真”的主题。本文认为三者翻译风格迥异的主要原因在于译者的翻译理念。

关键词

语料库翻译研究, 语言计量特征, 译者风格, 《归园田居》

A Corpus-Based Comparative Study of Quantitative Language Features on Translator's Style of Three English Versions of *Back to Country Life*

Chengzhen Lei

College of Foreign Languages, Guizhou University, Guiyang Guizhou

Received: Apr. 22nd, 2024; accepted: Jun. 6th, 2024; published: Jun. 17th, 2024

文章引用: 雷成真. 基于语料库计量语言特征的译者风格对比研究——以《归园田居》三译本为例[J]. 现代语言学, 2024, 12(6): 65-74. DOI: 10.12677/ml.2024.126435

Abstract

This study builds a monolingual analogical corpus of the English translations of Tao Yuanming's *Back to Country Life* by Wang Rongpei, Fang Chong, and Xu Yuanchong, then quantitatively analyzes their lexical, syntactic and discourse aspects by means of entropy, co-occurrence network of keywords and conjunctions, etc. On this basis, this study tries to explain the causes of different translators' styles. It is found that: Wang advocates the translation method of "conveying the spirit and expressing the meaning", tends to restore the imagery and themes of the original text, and builds up the ecological theme of "conforming to nature and simplicity", but the strategy of foreignization by phonetic translation may cause difficulties in comprehension; Fang is based on the concept of research-based translation, and uses the analogical method of translation to reflect the ecological value of "pursuing nature and detesting the world", and emphasizes the environmental description; Xu follows the "three beauties theory", and prefers to use more abundant words and flexible sentence structure to constructs the ecological theme of "enjoying nature and being true to oneself". This study contends that the differing translation styles of the three versions are primarily ascribed to the translator's translation concept.

Keywords

Corpus Translation Studies, Quantitative Language Features, Translator's Style, *Back to Country Life*

Copyright © 2024 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

党的二十大指出：“中国式现代化是人与自然和谐共生的现代化”。该论述不仅创新性发展了“天人合一”的传统生态思想，更表明人类现代化进程中生态治理的中国智慧和方案，这正是我国对外传播正面国家生态形象的重要机遇。东晋诗人陶渊明是中国田园诗派的开拓者，其代表作《归园田居》五首诗呈现出了远离浊世、躬耕田园的乐趣，其中既可以探查出儒家、道家等“仁民爱物”、“天人合一”的传统生态伦理观，又能够溯源“人与自然和谐共生”的现代生态主张。因此，陶诗外译及其翻译研究成为促进中华文化国际传播，对外建构正面的国家生态形象的重要手段。《归园田居》的英译受到国内翻译家的宠爱，具有代表性的译者中属汪榕培先生于2000年翻译的《英译陶诗》时间最早[1]，其次是2016年方重先生翻译的《陶渊明诗文选》[2]，最后是许渊冲于2021年出版的《许渊冲译陶渊明诗选》[3]。因此本文以陶渊明《归园田居》五首诗的汪榕培、方重、许渊冲英译本自建单语类比语料库，试图从语言计量特征“远观”译者不同翻译风格下所凸显的陶诗生态理念，以期对自然诗歌翻译策略提出参考建议，促进中华优秀传统文化的国际传播。

2. 研究现状

语料库翻译学是运用语料库语言学数据统计的研究方法对翻译本质、翻译过程和各类翻译现象翻译进行分析的翻译学分支学科。自Baker (2000)首次运用形符/类符比、平均句长、动词say及其变体的指标考察了译者风格的特点，译者风格研究已成为语料库翻译学的一大研究课题[4]。胡开宝和谢丽欣(2017)

认为译者风格是指“译者在翻译过程中形成的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”[5]。此处的翻译选择包含宏观和微观层面，前者指翻译策略使用等非语言特征，而后者指的是译者在词汇、句法、语篇或叙事层面的语言表达特征。随着数字人文应用的进一步成熟，涉及到词汇、句子乃至语篇的语言计量特征分析已经成为探究译者风格的常用研究方法。国内学者将形符/类符比、平均句长、高频词等描写手段应用于文学作品、散文诗词、政治文献等领域的翻译风格研究中。如韩丹、张志军(2023)使用词汇丰富度、词汇密度与句子长度等计量特征对《阿Q正传》俄译本语料库进行了翻译风格的探究，最后将重点放在翻译风格差异成因的解读上[6]。可见，语言计量指标的使用以及对翻译风格成因的定性探索是近年来语料库翻译风格研究的趋势。然而相关研究在计量特征的选择上往往依托相对单一的参数统计，并且相关研究更注重译者个人风格的描述与对比，缺乏对导致翻译风格差异原因的实证性探究，容易导致人文学科变为技术的“殖民地”的境遇[7]。

鉴于此，本文借助 KH Coder 以及 WordSmith 等语料库工具，从词汇、句子和语篇层面考察《归园田居》汪榕培、方重和许渊冲英译本中熵值、关键词共现网络、平均句长以及连词使用等语言计量特征，在借助语料库定量统计数据的同时，本文兼顾源文本，定性探讨译者风格异同及成因，以期为田园诗歌翻译提供借鉴和启发，促进田园诗英译的研究与发展。

3. 研究设计

3.1. 参数设置

基于上文发现的语料库辅助翻译研究存在计量特征指标单一、定性层面研究深化的不足等局限性，本文将在几个传统描写参数指标的基础上，结合计量语言学设置了以下参数指标：

(1) 词汇层面：标准形符/类符比(STTR)、平均词长、相对熵值、关键词共现网络。在计量语言学中，熵(Entropy)是“基于信息论而实现的一个概念，以文本中词的出现概率为出发点，可反映文本中词的丰富程度。”[8]以下是熵值的计算方法：

$$H_{(x)} = -\sum p(x) \cdot \log_2(x) \quad (1)$$

其中， $p(x)$ 是一个语言实体在样本中的相对频率(频率为 f 的语言实体所占的比例)， $-\log_2 p(x)$ 表示每个实体所携带的自我信息，而 $-p(x) \log_2 p(x)$ 则是每个语言实体的数学期望值[9]。因此，熵值越高，由一个语言实体所携带的信息量越大。

为了观察整体译本用词的丰富程度，本文将会使用相对熵值，算法如下，该指标试图描述语言实体在文本中所附带的相对信息量，其中 V 为文本的形符。

$$H_{rel} = H / \log_2 V \quad (2)$$

(2) 句子层面：平均句长。

(3) 语篇层面：连词的使用。

3.2. 研究步骤

本研究将《归园田居》的三个英译本自建语料库，相关统计结果如表 1 所示。具体步骤如下。

首先，对词汇层面进行语言特征项的描写及对比。基于语料分析工具 WordSmith 6.0 对三个译本的词汇变量中的标准形符/类符比(STTR)、平均词长指标等进行检索并进行统计；再利用语料库检索工具 KH Coder 3 对译本进行关键词共现网络分析以呈现译本词汇层面的主题网络；最后通过 AntConc 3.4.3 对译本的所有词频进行统计，基于图 2 给出的熵值计算方法，呈现出三译本的词汇丰富的程度。其次，对句子层面的语言特征进行描写与统计，即句长特征，数据将通过 WordSmith 6.0 获得。最后，借助 WordSmith 6.0 的词表功能得到三译本连词使用数据结果。

Table 1. Information of three translation corpus
表 1. 三译本语料库信息

译本	类符数	形符数	出版时间
汪译本	305	538	2000
方译本	294	533	2016
许译本	307	500	2021

4. 结果与讨论

4.1. 词汇层面

4.1.1. 常规词汇参数

通过对词汇指标的考量，本文能够明晰译者对词汇的丰富度及搭配关系的设置进而看出译文对读者的可读性和易读性。从表 2 可知，三译本在常规的词汇变量上存在差异：

Table 2. Statistical comparison of the lexical level
表 2. 词汇层面的统计比较

译本	标准形符/类符比/ 百分比	平均词长 /个	词长 标准差	1~8 个字母的单词 占比/百分比	相对熵
汪译本	56.69	4.10	1.83	90.12	1.43
方译本	55.16	4.08	1.91	88.30	1.38
许译本	61.40	4.11	1.87	86.26	1.44

首先，标准形符和类符比(STTR)是文本中每千词的类符和形符之间的比例。李上荣(2023)指出 STTR 比值越低，一定程度上反映了语料中的语言表达相对单一，文本阅读难度较小，文本呈现出语言简化的特征[10]。汪译本和方译本的 STTR 接近，不存在明显差异，说明两译本的词汇量接近，阅读难度趋同。而许译本的 STTR 大于汪和方译本，说明的许译本词汇表达上灵活性更大，这从侧面佐证了许译本的编译策略，词汇选取上充分发挥了译者主体性。

其次，三译本的平均词长变量差异并不显著，均为 4 个字母，说明三译本基本都遵循简单词汇的使用。从词长标准差来看，相比于另两个译本，方译本的词长分布较不均衡，波动较大，而汪和许译本在形符数较大的情况下，词长分布整体还趋于稳定，这为译本句式结构层面的工整奠定了基础。三个译本 1~8 个字母的单词占比均达到总数的 85%以上，表明三译本以易读词汇为主，可读性较高。由此可见，三位译者试图从词汇的通俗易懂下手，促使更多的国外读者能感受到田园诗中的生态智慧。

最后，从相对熵的大小来看，许译本的编译策略导致其译本词汇选择较为灵活，丰富的词汇致使其文本相对熵多于另外两个译本。

通过词汇层面的变量对比我们发现，汪译本和方译本在词汇变量上具有相似性，而许译本更多地体现了译者主体性，用词较丰富与灵活，且译作在篇幅上做了较大调整。下文将就具体词汇特征做进一步探讨。

4.1.2. 关键词共现网络

本研究利用 KH Coder 3 选取了出现频次大于 2 且排序在前 60 的关键词，并通过 Co-Occurrence Network 功能形成译本关键词共现网络可视化图谱，这些词与词之间的联系代表了三位译者在翻译时最

侧重凸显的田园主题，节点的颜色越深，则越能体现译本的词汇主题。

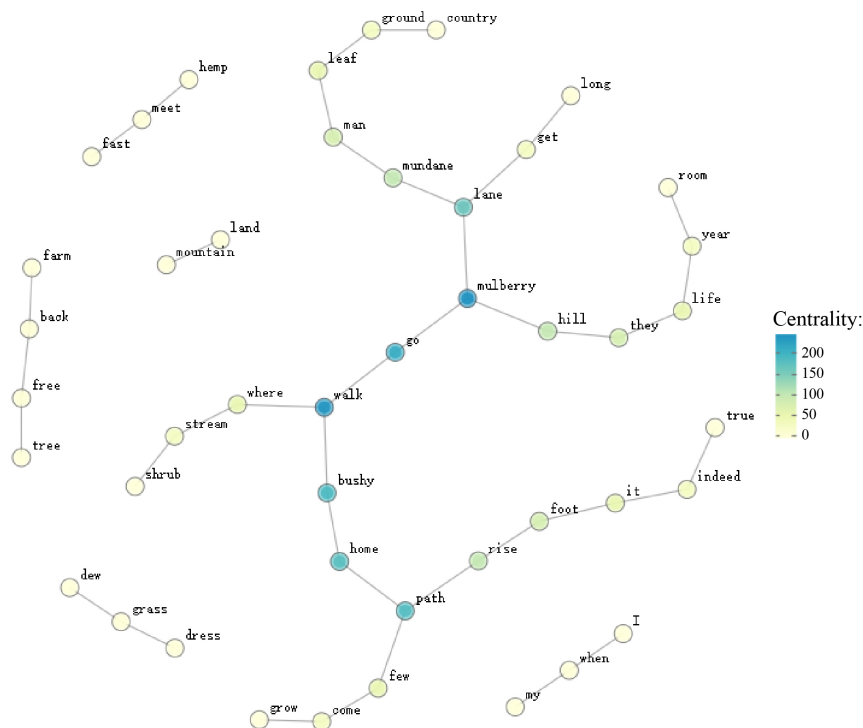


Figure 1. keywords co-occurrence network of Wang's version
图 1. 汪译本关键词共现网络图谱

首先,从共现集中度为 200,相互联系的节点个数来看,汪译本的主题呈现集中,节点词之间主次清晰(见图 1)。中心词 go、walk、mulberry 与 path、home、hill 等次一级搭配词主要凸显了诗人顺应自然,纯朴的生态取向语义(如例 1),结合再次一级的搭配如 come、life 等词则展现了对事物有生则有死自然走向的尊重,最外层的搭配如 grow、farm、mountain、tree、land 等则与作物与耕地、自然风光等概念关联。因此,作为田园诗歌译本,汪译本着重建构田园自然的各类核心概念,词汇表达准确分明,符合原作风格。但同时也应注意到,例 2 在翻译田园核心文化词“亩”时,汪译本采取了音译,此种异化策略一定程度上增加了目标语读者的理解难度。

例 1:

原文：薪者向我言：

“死没无复余。”

• • • • •

终当归空无。

汪译: “They’re dead and gone, with not a single heir.”

• • • • •

To nothingness it will eventually go.

例 1 中的“go”将人生有盛则有衰、有生则有死这样一个不可避免的自然法则生动地阐释出来，体现出汪译本对人与自然和谐的关注。

例 2:

原文: 方宅十余亩

汪译: My farm contains a dozen mu of ground.

其次, 方译本的共现集中度仅有 16, 且节点词最多的共现个数仅有 9 个, 节点之间没有交叉, 呈直线趋势, 其主题集中度较模糊(见图 2)。但仍可从中心词及其搭配看出汪译本主要基于追求自然, 厌恶世俗这两方面展开, 如将 *deep*、*few* 中心词结合深巷(*lane*)、寡轮鞅(*wheel*)以及最外层搭配词虚室(*vacant*)等, 呈现了诗人对闲情逸致的自然生活的向往, 而如尘网(*net*)、俗韵(*world of man*)等, 译者通过将 *dusty*、*world* 结合以上搭配描述了诗人对官场的唾弃。

例 3:

原文: 少无适俗韵, 性本爱秋山。

方译: From my youth I have loved the hills and mountains, Never was my nature suited for the world of men.

例 3 将 *hills* 和 *mountains* 作为所爱(*loved*)之物, 而 *world of man* 则是与自然不匹配(*not suited*)的“俗韵”, 汪译本在这里将中华文化独有的表达用人类所共同体验的人类世界进行类比处理, 能够激活受众有关人类社会中种种不公的现象, 这种语言表达便于目标读者更好地理解“俗韵”的内涵, 展现了译者在此主题网下采用的显化策略。

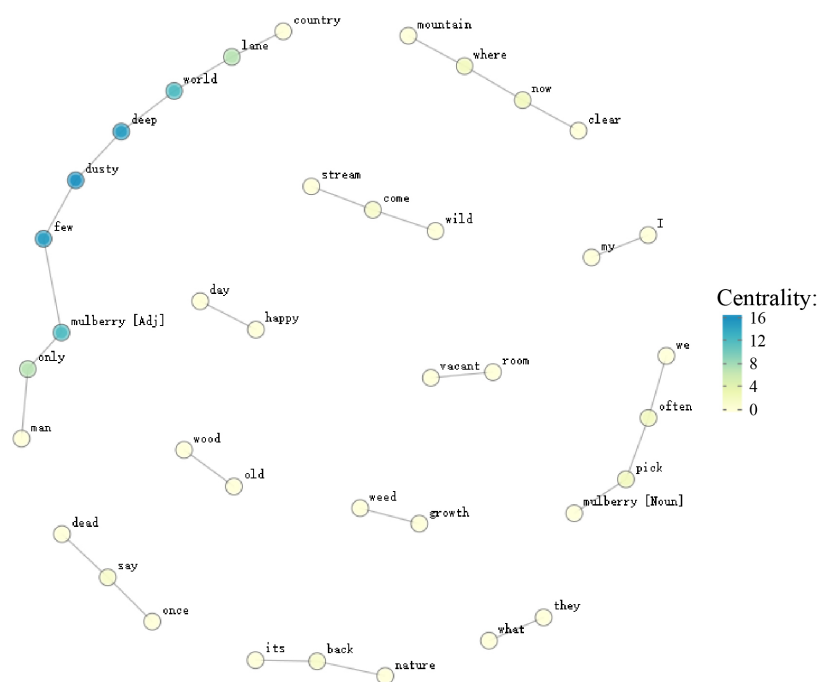


Figure 2. Keywords co-occurrence network of Fang's version
图 2. 方译本关键词共现网络图谱

最后, 许译本的词汇网络共现中心度达 200, 且节点之间存在紧密的交叉趋势, 其搭配网络密集, 主题连贯(见图 3)。同样, 许译本中 *go*、*room*、*find* 结合其核心搭配词 *back*、*field*、*short* 来看, 词网络强调了享受自然, 任性存真的表达, 并从发现自然之美(*find*, *see*, *hear*)、回归自我(*I*, *my*, *room*, *foot*)等主题展开。

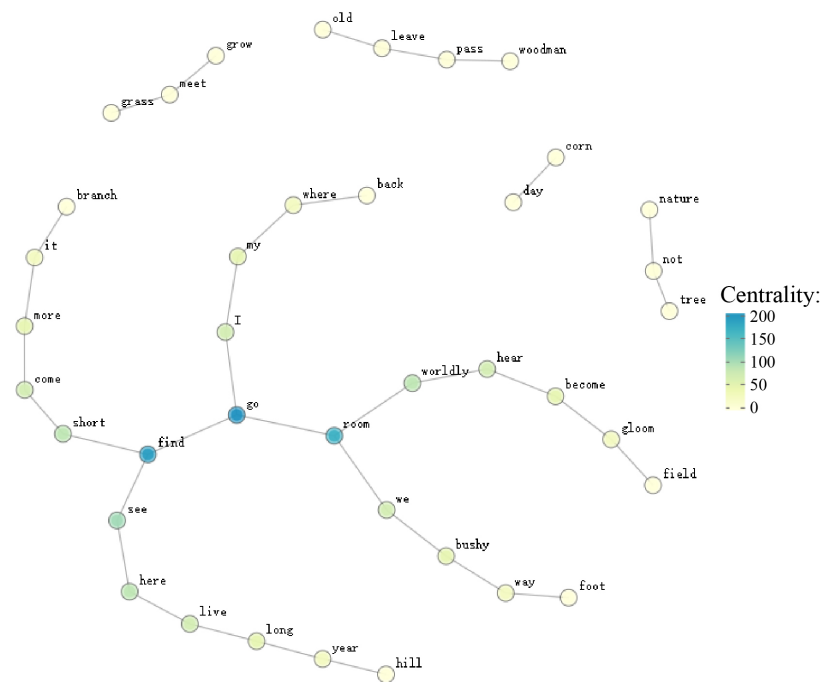


Figure 3. Keywords co-occurrence network of Xu’s version
图 3. 许译本关键词共现网络图谱

例 4 中，许译本将原文的“复得返自然”译为具有欣然发现美好的姿态-find homely pleasure，这种翻译策略的使用轻而易举就彰显出诗人归田之心的喜悦神态，许译本同样体现了在翻译上显化策略的使用。

例 4:
原文：久在樊笼里，复得返自然。
许译：After long, long official servitude, Again in nature find homely pleasure.

值得注意的是，在汪译本和方译本中都出现的作物桑麻(mulberry)在许译本中却未得到凸显，相反许则使用 corn 对其进行显化处理，mulberry 虽更具田园特色，但汪与方译本其对术语的直接使用，此种异化策略一定程度上增加了目标语读者的理解难度。

4.2. 句子层面

平均句长是指文本形符数和句子总数之间的比值，一般来说，“平均句长与文本的复杂程度成正比。”
[11]本研究使用 WordSmith 统计工具，结果如表 3 所示。

Table 3. Statistical comparison of the sentence level
表 3. 句子层面的数据统计

译本	平均句长/个	句子标准差
汪译本	8.43	1.84
方译本	8.20	2.23
许译本	7.97	2.34

从表 3 得知三译本的平均句长均低于 9,说明三者共同体现了诗歌简短精炼的文体特色。具体而言,汪译本的平均句长最高,而句子标准差又是最低,说明汪译本更多使用复杂长句,但句子变化波动小,句子分布均衡。而许译本平均句长最短,体现了其在形式上遵循形美理念,忠实于原文。

以上分析体现了三译本的词汇和句子风格,许译本的翻译最为灵活,以丰富的词汇将田园生活相关的概念明晰化。汪译本虽在主题集中和句子层面胜于许译本,但许译本的显化特征更加明显。下文将从语篇角度做进一步考察。

4.3. 语篇层面

译者在对整体语篇的翻译效果进行考量时,常借助连词来提高译文的连贯与逻辑性。三译本连词的使用情况如表 4。

Table 4. Statistical comparison of the textual level
表 4. 语篇层面的数据统计

	汪译本	方译本	许译本
and	28	30	23
but	1	2	5
or	2	1	1
so	1	2	1
when	8	1	1
if	1	2	1
总计	41	38	32
百分比	13.44%	12.91%	10.42%

通过表 4 发现,汪和方译本的连词使用在各自文本中的总形符数中占比相差不大,而许译本对于连词的使用则较少。结合前文得出的结论,汪和方译本篇幅较短,但句式较长的特点,两译本要通过连词将较长的句子链接在一起,在较短的篇幅中将原文的各种语义关系向读者传达,因此,汪和方译本语篇的连贯性高于许译本。

5. 翻译风格差异成因解读

以上研究表明,三译本翻译风格在词汇、句子及语篇层面均存在不同,但整体上再现了《归园田居》生态思想。三译本部分计量特征折射出翻译风格不同的成因,本文主要分析译者不同的翻译理念。

首先,方重所提倡的研究型翻译理念,确定了“方重的译本既是文学性强、与原著了无隔阂的文学译本,也是学术研究译本。”[12]因此,方重陶诗英译在词汇的共现搭配中大量选择与作物、耕地与自然风光等概念关联的搭配词,聚焦于使用具体的环境描写来证实反映陶渊明的生态追求,从而趋向展现出顺应自然,纯朴的生态取向语义,而在平均句句长与连词使用的计量特征介于方和许之间,呈现与原作“了无隔阂”的趋势。

其次,汪榕培倡导的“传神达意”翻译思想富有创见,旨在实现诗歌翻译的“神韵”。其中,汪榕培(2007)认为,诗歌的翻译方面的“传神达意”要做到传达原作的神情(形式、语气、意象等)表达原作的深层意义和表层意义[13]。基于此,汪译本的词汇网络共现密度集中,节点交互性强,对译本生态主题的相关表达丰富,说明了译本试图从词汇角度实现与原作之间在精神实质上的对应,还原原作的意象和主

题,以达到“传神达意”的效果。但此种“传神”追求也带来局限性,即采用异化策略对文化词进行音译处理,会对读者的理解带来困难。

最后,许渊冲主张中国古典诗词的翻译,应达到意美、音美、形美的“三美”境界。意美与音美的要求促使许译本趋向使用较为丰富的词汇表达以及主题性强的词汇搭配,但有小部分译文因其强调诗歌的形体美,限制了译文句长,具体为较少的连词使用,因此常采用显化的翻译策略,多选择目标读者容易理解的词汇,以至于未能很好地表现原诗的意境。

6. 结语

富有深厚生态思想的田园诗在中华优秀传统文化国际传播中发挥着重要的作用。本文利用语料库工具,采用“远观”的量化方法,从词汇、句法以及语篇层面考察了《归园田居》汪榕培、方重和许渊冲英译本翻译的风格特征,并结合“细读”定性分析了三译本翻译风格的差异及其成因。研究发现,三译本的翻译风格存在一定的差异,但都各具特色。汪译本与许译本的词汇与句子层面的变量相差不大,均体现了译文在词汇选择上的灵活与丰富性,而汪译本更注重传神达意,因此其呈现出的主题网络共现集中度更高,但其音译的异化策略会造成理解困难,而许译本遵循的“三美”原则促使了语义建构上显化策略使用。方译本的词汇计量特征最低,因其重视研究基础上的翻译实践,因此其客观性较强,更忠实于原诗结构,但其也采用了显化的策略,促进了陶诗生态意蕴的传达。总之,在数字人文发展以及着力建设生态文明话语体系的背景下,对陶诗翻译风格的研究,能为田园诗歌的英译带来新的启发。

基金项目

贵州省 2021 年度哲学社会科学规划课题《贵州濒危仡佬语语言生态深度调查及数据库建设研究》,项目编号: 21GZYB38;

教育部中外语言交流合作中心课题《<国际中文教育中文水平等级标准>与<欧洲语言等级框架>的对比研究》,项目编号: YHJXCX23-096;

教育部国际司中国教育交流协会课题《“一带一路”视域下中国—东盟国际教育交流与合作路径研究》,项目编号: BREC2023-006。

参考文献

- [1] 汪榕培. 英译陶诗[M]. 北京: 外语教学与研究出版社, 2000.
- [2] 方重. 陶渊明诗文选译[M]. 北京: 外文出版社, 2016.
- [3] 许渊冲. 许渊冲译陶渊明诗选[M]. 北京: 中国对外翻译出版社, 2021.
- [4] Baker, M. (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>
- [5] 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.
- [6] 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.
- [7] 耿强, 周知非. 数字人文视域下《人民日报》(1949-1966)生产的中国翻译话语研究[J]. 外语电化教学, 2023(1): 18-24+107.
- [8] 刘海涛. 计量语言学导论[M]. 北京: 商务印书馆, 2017.
- [9] Chen, R.N., Liu, H.T. and Altmann, G. (2016) Entropy in Different Text Type. *Digital Scholarship in the Humanities*, 32, 2-6. <https://doi.org/10.1093/llc/fqw008>
- [10] 李上荣. 文化“走出去”视域下《诗经》两译本翻译共性研究——基于语料库的统计和分析[J]. 今古文创, 2023(32): 108-111.
- [11] 杨惠中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2002.

- [12] 郑清斌. 虽不能至, 心向往之: “上外建校 60 周年: 纪念方重先生座谈会”纪要[J]. 中国比较文学, 2010(1): 149-151.
- [13] 汪榕培. 《诗经》的英译——写在“大中华文库”版《诗经》即将出版之际[J]. 中国翻译, 2007(6): 33-35.