

基于语料库的《觉醒》汉译本风格对比研究

蒋雨琦, 刘 锐

大连海事大学外国语学院, 辽宁 大连

收稿日期: 2023年9月27日; 录用日期: 2023年10月27日; 发布日期: 2023年11月8日

摘 要

本论文选取凯特·肖邦的著作《觉醒》为原本文语料, 选取齐彦婧译本和王骁双、张爽、戴婧译本作为语料素材, 自建《觉醒》中英文语料库。运用语料库研究方法, 使用几种语料库分析工具, 从语言特点入手, 对词汇和句法两个层面的基本特征进行数据统计和对比分析, 探讨两个汉译本中译者风格。研究发现译本在忠实原文的基础上, 不同译者的翻译风格略有不同。齐彦婧译本采用偏异化的翻译策略, 而王骁双、张爽、戴婧译本采用偏归化的翻译策略。

关键词

语料库, 《觉醒》, 译本风格, 异化, 归化

A Corpus-Based Study on the Comparison of the Styles of Chinese Versions of *The Awakening*

Yuqi Jiang, Rui Liu

School of Foreign Language, Dalian Maritime University, Dalian Liaoning

Received: Sep. 27th, 2023; accepted: Oct. 27th, 2023; published: Nov. 8th, 2023

Abstract

In this study, Kate Chopin's book *The Awakening* is selected as the original text corpus, and the translations of Yanjing Qi and Xiaoshuang Wang, Shuang Zhang, and Jing Dai are selected to build an English-Chinese translation corpus of *The Awakening*. Based on the application of corpus, several corpus analysis tools are used to explore the translator's style in the two Chinese versions, starting from linguistic features, and data statistics and comparative analyses are conducted on the basic features at both lexical and syntactic levels. It is found that the translations, on the basis

of being faithful to the original text, have slightly different styles of translation by different translators. The version of Qi Yanjing focuses foreignizing translation, while the version of Wang Qushuang, Zhang Shuang and Dai Jing focuses domesticating translation.

Keywords

Corpus, *The Awakening*, Translation Style, Foreignization, Domestication

Copyright © 2023 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

《觉醒》是美国女作家凯特·肖邦的经典作品,该书自问世以来争议不断,作为当时敢于挑战男性话语权的作品,曾一度被认为是污秽之作。凯特·肖邦所创造出的女主角,不再属于男性凝视下的产物,反而逐步产生自我意识,勇于打破婚姻的枷锁。然而就创作艺术而言,本书从一开始就被认为是上乘之作。凯特·肖邦赋予了《觉醒》深刻的精神内涵——鼓励女性敢于打破精神桎梏,寻求解放,重建自我。因此,从上世纪50年代《觉醒》便开始受到重视,甚至改编成歌剧演唱。

国内学者注意到了外国名著的文学价值和研究价值,大力发展学术翻译并开启了“汉译名著之路”。外国文学作品的译入不仅拓宽了国内的文学视野,也为国内研究提供了丰富的文本资源。然而,通过阅读文献发现,国内学者对《觉醒》的研究局限于女性主义批评,且往往仅通过学者自筛译例后的分析,缺少对其汉译本的对比研究。

Hermans (1996: p. 27)曾指出译本中隐藏着“译者的声音”[1],译者风格作为隐性特征存在于作品中。Baker (2000: p. 245)主张将语料库语言学与翻译学的理论和方法相结合来研究翻译中的译者风格[2],此论调为语料库和翻译的研究均开拓了新视野。近年来,国内外学者在此基础上探索了语料库翻译学中的译者风格,研究也取得了较快发展。杨柳和朱安博(2013)认为“自建语料库具有建库周期短、易于操作、针对性强的特点,不仅可以用于探讨‘翻译总特征’,还适用于量化分析译者个体的翻译风格和译本特点,为通过传统译本研究得出的观点提供直观的数据参考。”[3]。

因此,本论文将对《觉醒》的两个汉译本(即浙江教育出版社出版的齐彦婧译本和中信出版社出版的王骁双、张爽、戴婧译本)进行对比,基于自建的语料库和几种语料库分析工具,从词汇与句法两个层面来考察不同译本的差异,旨在探究不同译者所采用的翻译策略,浅析译者风格的成因。

2. 研究设计

2.1. 研究语料

本研究选取美国女作家凯特·肖邦的《觉醒》英文文本、浙江教育出版社出版的齐彦婧译本和中信出版社出版的王骁双、张爽、戴婧译本为研究语料。语料库的收集和处理在完成必要程序的基础上,根据研究的需要增加了一些额外的程序。第一,原始译文输入:电子版六篇英文译文均从网上采集,并分别以译者姓名作为文件名保存;第二,语料处理:将语料整理为文本文档格式,并清理文本。

《觉醒》两译本语料库进行对照的参考语料库为 ToRCH2014 现代汉语平衡语料库,该语料库 2015 年开始创建,2017 年夏正式建成发布。ToRCH2014 语料库的取样方案与布朗语料库相同。即包含 15 个

小的文类, 可合并为新闻(Press)、通用(General)、学术(Learned)、小说(Fiction)四大体裁。较为客观地反映了现代汉语普通文体书面语的特征, 且语料收集年份与两译本发布年份相近, 避免了语言历时发展特征造成的误差, 因此本文选择该库作为对照库。

本研究将研究语言特征中的句法和词汇两个层面, 旨在回答以下两个问题: 1) 两译本在词汇特征上是否存在差异? 2) 两译本在句法层面特征上是否存在差异?

2.2. 研究工具和方法

随着互联网的发展和技术的进步, 许多语料库工具被设计并应用到翻译研究中, 使得对文件量大的文本进行定量分析成为可能。一方面, 这些科学应用可以为研究人员节省大量的时间和精力, 另一方面, 研究结果具有更高的可行性和有效性。基于上述有利条件, 本研究采用了多种经济型语料库工具, 包括用于英文文本注释的 Tree Tagger 和中文文本注释的 ICTCLAS、用于收集指标相关数据的 Wordsmith 4.0 和 Antconc 4.2.0。

利用语料库检索工具 Wordsmith 4.0 从词汇和句法层面考察了两译文的定量特征, 主要参数包括标准化标记类型比、平均句长和词频表(前 10 位高频词)。需要强调的是, 本研究的语料库只保留了正文, 而删除了前言、注释、附录和后记等副文本信息。

要进行语料分析, 得到 TTR 等数据, 必须对汉语语料进行分词处理。本文使用的分词软件为中国科学院计算机技术研究所研发的 ICTCLAS 5.0 版。

Wordsmith Tools 由多个语言分析应用程序组成, 由利物浦大学教授 Mike Scott 设计。Wordsmith 可以应用于翻译研究的时间可以追溯到 1996 年, 目前已广泛应用于语言学和翻译研究。该应用程序有三个功能: 1) Concordance: 可以帮助研究人员快速找到文本中的任何单词或短语; 2) Keywords: 可以让用户发现文本中的关键词, 但前提是需要加载另一个语料库的参考词表; 3) Wordlist: 在该功能中, 研究人员可以获得精确的数据, 包括词类数量、标记、标准化词类 - 标记比、平均句长等。

3. 研究结果和讨论

3.1. 词汇层面

3.1.1. 词汇丰富度 TTR 和 STTR

类符是指文本中不同词汇的数量, 形符是指文本中单词的总数。类符/形符比可以用来评价大小相同的文本的词汇量和多样性[4], 但如果文本大小不同, 类符形符比可能会出现偏差, 无法准确判断。因此, 选择采用 Scott 提出的标准化类符形符比(STTR), 即每 1000 个单词计算一次文本的类符形符比, 根据这一原则, 数值越高, 文本的用词多样性越高, 反之亦然。

Table 1. Basic information of vocabulary in the English-Chinese corpus of the two self-constructed translations of *The Awakening*

表 1. 自建《觉醒》两译本英汉语料库词汇基本信息

	类符	形符	类符形符比	标准类符形符比
<i>The Awakening</i> 英文文本	6173	50,848	12.140	44.661
觉醒 王骁双、张爽、戴婧译本	7060	50,940	13.863	47.528
觉醒 齐彦婧译本	6945	47,852	14.515	46.367

利用 Wordsmith 4 计算出英文文本和两个中文译本的类符形符比和标准类符形符比。表 1 的结果如下: 王骁双、张爽、戴婧译本的译文词汇量更多, 丰富度更高, 齐彦婧译本的译文较低。ToRCH2014 语

料库的标准类符形符比为 46.49%。以上数据说明两译本的词汇丰富度比英文原文高, 而王骁双、张爽、戴婧译本比汉语原创文字的词汇丰富度高。

3.1.2. 高频词

高频词指的是出现频次高的词, 高频词反映出译者的词汇选择, 利用 Antconc 分别导出两译本的高频词, 在 ToRCH2014 语料库网站导出高频词。表 2 展示的是两译本以及参照语料库中出现的前十位高频词。

Table 2. High-frequency words in the corpus

表 2. 语料库高频词

	王译本	频次		齐译本	频次		ToRCH2014	频次	
1	的	2900	5.69%	的	2220	4.64%	的	61,816	5.06%
2	她	1525	2.99%	她	1510	3.16%	了	14,553	1.19%
3	了	1165	2.29%	了	1023	2.14%	是	13,333	1.09%
4	他	843	1.65%	他	899	1.88%	在	13,016	1.06%
5	在	836	1.64%	在	751	1.57%	和	9137	0.74%
6	我	698	1.37%	不	678	1.42%	一	8435	0.69%
7	不	675	1.33%	我	619	1.29%	我	8182	0.67%
8	是	594	1.17%	一	609	1.27%	不	6463	0.53%
9	着	554	1.09%	是	566	1.18%	他	6357	0.52%
10	一	551	1.08%	你	473	0.99%	有	5412	0.44%

如表 2 所示, 两译本和参照语料库中排名前十的词条都是单个汉字, 如“的、我、他、一、了、你、不、是”等。首先, 在二者的译文中, “的”是使用频率最高的一个。这主要是因为“的”是汉语中最为常见的虚词之一, “的”本身没有实际意义, 主要用来构成形容词性的偏正结构或者用作语气助词表示个人情感[5]。

其次, 两个译本中的前五个词条相同, 频率略有不同。第三, 在两个中文译本中, 前十个高频词中出现了较多人称代词, 包括“她、他、我、你”, 且频次都高于参照语料库。以“他”这个人称代词为例, 所占比例约为 1.65%或 1.88%, 高于参照语料库的 0.75%。小说中具有多于一般文体的心理剖析和人物对话[3]。由于中英文语言的差异性, 汉译者往往出于忠实原著的目的, 相比于汉语作家会更多地使用代词。同时就翻译技巧而言, 人称代词作为衔接手段也可以保证翻译作品的完整性和丰富性。因此, 小说的中文译本代词使用频率高, 有助于读者更好地理解原文, 这也显示了译者的异化翻译策略。

3.2. 句法层面

句法特征可以通过平均句子长度反映译者的风格, 不同的句子长度反映了不同的难易程度。平均句长是指文本中所有句子的平均单词数。句子长度主要受句子结构的影响, 译者在翻译过程中的操作规范会影响句子长度的多样性。在本研究中, 句子的数量可以通过统计特定的标点符号来计算, 这些标点符号具有结束句子的功能, 如句号、问号和感叹号等, 标点符号数量与句子数量的比值即为平均句子长度。

Table 3. Average sentence length of the two translated texts
表 3. 两译本平均句长

	王译本	齐译本	英文文本
句子总数	3288	3331	3701
平均句长	15.488	14.365	13.534

一般来说, 译文的句子数和平均句长都超过了原文, 这说明译文的阐释能力较强。在上面的数据中, 两译本的平均句长比原文本长。“就整个语料而言, 语料的平均句长可以在一定程度上反映句子的复杂程度。” [4]。

Butler (1985: p. 121)认为, 句子按长度可分为三类: 短句(1~9 个字)、中句(10~25 个字)和长句(25 个字以上) [6]。一般来说, 句子越长, 就越难理解, 语言就越正式。根据表 3, 两个译本的句子属于中句长度。且两个版本的平均句子长度长于源文本, 展现出句法显化的特征。这可能有很多原因, 但只需说明在翻译过程中存在大量本土化和改编的证据, 即调整文本以满足读者的期望和需求。在运用语料库方法分析翻译文本共性的时候, 通常认为, 译文文字总量和平均句长大于原文的主要原因在于显化。譬如说, Nida 就曾指出好的译文通常会比原文更长, 这是因为译者通过明示原文隐含的信息以帮助读者理解, 在一定程度上增加了译文的冗余程度 [5]。

此外, 王译本的平均句子长度多于齐译本。但齐译本的句子数比王译本多(见表 3)。一方面, 王译本的词汇丰富度较高, 在译文中采用了更多阐释和描述性的语言。另一方面, 齐译本可能采用了更多的短句, 在翻译时为了语言简洁, 使用了更多标点符号断句。

例:

“Edna Pontellier could not have told why, wishing to go to the beach with Robert, she should in the first place have declined, and in the second place have followed in obedience to one of the two contradictory impulses which impelled her.

A certain light was beginning to dawn dimly within her,—the light which, showing the way, forbids it.

At that early period it served but to bewilder her. It moved her to dreams, to thoughtfulness, to the shadowy anguish which had overcome her the midnight when she had abandoned herself to tears.” [7]

王译: “埃德娜·蓬特利尔不清楚自己为什么想和罗伯特一起去海滩, 两种互相矛盾的冲动驱使着她, 她先是觉得应该拒绝, 接着又想要去。

她的心里出现了一束朦胧的光——这束光为她指明了前路, 却又禁止她往前迈步。

在这样的初始阶段, 这束光让她迷茫, 引她入梦, 令她沉思, 迫使她再次体会那曾令她长夜痛哭的隐隐痛苦。”

[8]

齐译: “埃德娜·庞德烈其实是想跟罗伯特一起去海滩的。但她不明白自己为什么会受到两种矛盾的情绪驱使, 会先拒绝再答应。

她心里隐隐亮起一道模糊的光——它为她指明方向, 却禁止她迈出脚步。

那时一切才刚刚开始, 面对这光亮, 她只觉得不知所措。这让她浮想联翩, 思绪万千, 让她陷入午夜独自垂泪时那种隐痛。” [9]

以上例子中可以看出, 齐译本则运用了更多的短句来翻译, 多用标点符号来进行断句, 使译本简练、

易读性强。而王译本的句子数量更接近英文原文,王译本句子长度也相对比齐译本长。因此,在翻译过程中王译本更多地使用增译手法对原著进行更加细致的表达,更倾向于向译作受众传递原作者目的与小说精神内涵。

4. 结论

基于自建的语料库和几种语料库分析工具,本文研究了《觉醒》的两译本在语言上的特征。从类符形符比和高频词上分析了词汇层面的特征,从平均句子长度上分析了句法层面的特征。

总体而言,在词汇层面上,就标准类符形符比而言,王骁双、张爽、戴婧译本的译文词汇量及丰富度比齐彦婧译本的译文高,用词更加丰富;高频词中,汉译本小说的人称代词使用频率比汉语原创文本使用频率高。在句法层面,汉译本的平均句子长度总体比英文原文长,与以往研究结果相符;其次,王译本的平均句子长度比齐译本长,采用了较多阐释和描述性的话语,而齐译本采用了较多短句。

两个汉译本均遵循了忠实原文的准则,在此基础上,王骁双、张爽、戴婧译本使用了更加丰富的词汇表达,平均句子长度更长,说明译者对语篇的掌控力略强,考虑到了英文小说及其文化背景可能对中文读者带来的影响,更注重信息传达,采用了归化的翻译策略;而齐彦婧译本词密度较低,句式简短,注重对原文信息的保留,译文流畅,使得读者容易理解,采用了异化的翻译策略。因此,两个汉译本风格略有不同,读者可以根据自己的语言需求选择译本阅读。

上述的研究通过语料库和语料库分析工具得出了小说以及译本中的各项参数。语料库的建立可以帮助研究者从定量上研究译本的语言和翻译特征,为翻译研究提供了新角度。此外,本文研究的数据较少,没有全面研究各项数据,后面的研究可以加入更多数据和研究角度来对比分析小说不同译本。

参考文献

- [1] Hermans, T. (1996) The Translator's Voice in Translated Narrative. *Target*, 8, 23-48.
<https://doi.org/10.1075/target.8.1.03her>
- [2] 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>
- [3] 杨柳, 朱安博. 基于语料库的《温莎的风流娘儿们/妇人》三译本对比研究[J]. 外国语(上海外国语大学学报), 2013, 36(3): 77-85.
- [4] 杨惠中. 语料库语言学导论[M]. 上海: 上海外语教育出版社, 2002.
- [5] 牛桂玲, 田晓芳. 《麦田里的守望者》两译本对比研究——基于语料库视角[J]. 郑州航空工业管理学院学报(社会科学版), 2016, 35(2): 131-135.
- [6] Butler, C. (1985) *Statistics in Linguistics*. Basil Blackwell, Oxford.
- [7] Chopin, K. (2004) *The Awakening and Selected Short Stories of Kate Chopin*. Simon & Schuster, Inc., New York.
- [8] 肖邦. 觉醒[M]. 王骁双, 张爽, 戴婧, 译. 北京: 中信出版社出版社, 2016.
- [9] 肖邦. 觉醒[M]. 齐彦婧, 译. 杭州: 浙江教育出版社, 2020.