

基于语料库的译者风格研究

——以《秘密花园》为例

傅至轩

上海海事大学, 上海

收稿日期: 2022年8月4日; 录用日期: 2022年9月12日; 发布日期: 2022年9月19日

摘要

本研究以小说《秘密花园》的英文原著、李文俊与梅静两位不同译者的汉译本为研究对象, 通过自建语料库采用相关软件做出研究和探讨。本研究主要以词汇与句子两个层面为主, 采用定量与定性研究的方法对两个译本做出比较和分析, 通过不同语料项目之间的对比研究来分析两位译者可能拥有的不同翻译风格及其对读者可能产生的阅读方面之影响。经研究发现, 两译本的词汇丰富程度和阅读难度上接近。梅译本有着更高的类符/形符比和词密度, 平均句段长度更短, 叠词、拟声词等特殊词汇使用较多; 而李译本词汇变化度较小, 平均句段长度更长, 叠词使用较少但儿化音的使用更多。两份译本虽然在翻译风格上有所区别, 但都达到了《秘密花园》这部小说作为儿童文学作品应有的阅读层次。

关键词

语料库, 译者风格, 《秘密花园》, 儿童文学

A Corpus-Based Study of Translator Style

—Taking *The Secret Garden* as an Example

Zhixuan Fu

Shanghai Maritime University, Shanghai

Received: Aug. 4th, 2022; accepted: Sep. 12th, 2022; published: Sep. 19th, 2022

Abstract

This study takes the original English text of the novel *The Secret Garden* and two translated Chinese versions by two different translators, Li Wenjun and Mei Jing, as the material. This study is made through a self-built corpus, retrieved and analyzed with some corpus-related software. This study focuses mainly on lexical and syntactic aspects, and uses both quantitative and qualitative

research methods to compare and analyze the two translations. This study can thus draw conclusions on the different translator styles that Li Wenjun and Mei Jing may have had and the impact they may have had on their readers in terms of reading. It has been found that the two translations are close in terms of vocabulary richness and reading difficulty. Mei's version has higher Type/Token Ratio and word density, shorter Mean Sentence Length, and higher frequency of some special words. While Li's version has less variation in vocabulary and more rhotic words. Although the two translations differ in translator style, they both achieve what the novel *The Secret Garden* deserves as a work of children's literature.

Keywords

Corpus, *The Secret Garden*, Translator Style, Children's Literature

Copyright © 2022 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

《秘密花园》是 20 世纪美国女作家弗朗西斯·霍其森·伯内特的代表作，具有极高的美学价值。作为一部畅销百年的经典儿童文学著作，它在世界范围内先后被翻译成了 50 多种语言，衍生出许多电影、电视以及话剧等艺术作品，仅在中国大陆地区就有数十种不同的译本。只要说起《秘密花园》，人们就会想起一个成长的安乐窝、一个蜕变的发生地、一个灵魂的避难所。这本关于如何寻找幸福和快乐的书，不仅是对孩子心灵的滋养、愉悦和启迪，而且对不同家庭的儿童教育有着重要的借鉴作用[1]。

2. 研究现状

语料库翻译学最早在上世纪 90 年代由英国学者 Mona Baker 提出，之后她将这一研究方法与翻译学创新结合，应用于文学作品的译者风格研究之上[2]。Saldanha 将译者风格研究分为翻译文体(Translation Style)和译者风格(Translator Style)两大基本分类，前者侧重目标文本与源文本的对应方式，后者则指的是译者特有的表达方式[3]。作为一项重要内容，语料库翻译学在国内也被学者广泛讨论和探索。有基于语料库的新闻翻译研究，如董帅[4]等；有口译方面研究，如徐翠[5]等；还有从语料库出发，对比国内外小说、古代典籍、政治文本等不同类型译本的翻译风格，如赵轩[6]、郑一菲[7]和赵颖[8]等。

3. 研究设计

本研究通过选取传阅度较高的李文俊与梅静二位译者的译本，自建语料库进行对比分析。本研究主要从词汇层面(类符/形符比、词汇密度与特殊词汇等)与句式层面(平均句长、平均句段长等)两个角度入手，考察两位译者在《秘密花园》这部儿童文学作品上翻译风格的差异。

3.1. 语料库建立

本研究之英文语料选取 1998 年美国哈珀柯林斯出版社出版的弗朗西丝·霍奇森·伯内特经典儿童小说 *The Secret Garden 100 Anniversary* (《秘密花园 100 周年版》[9]，下称“英文原文”)，中文语料选取 2018 年北京中信出版社的李文俊汉译本《秘密花园》[10] (下称“李译”)以及 2017 年云南美术出版社的梅静汉译本《秘密花园》[11] (下称“梅译”)。将它们收集整理后按照如下步骤建立个人小型研究语料

库：

- 通过 OCR 文字识别软件 ABBYY FineReader 将书本素材转换为文字格式，并经过人工处理校正和删去一些图书信息、作者或译者简介、乱码以及特殊符号等内容，在仅保留小说章节标题和正文内容的情况下将其存储为便于相关语料软件识别的 txt 格式纯文本文档；
- 为便于语料的识别和使用，通过文字编辑软件 EmEditor 对于 txt 格式的文本文档进行清洁、提纯和整理；
- 通过语料库分词和词性标注程序 CorpusWordParser 分别对一个英文原文文本和两个汉译本文本进行分词和词性的标注。由于汉译本中人名、地名或是某些专有名词的特殊性可能导致分词或词性标注困难，为了减少分词后人工处理的工作量，在分词之前便通过软件的“自定义词表”功能对部分进行预先处理，如图 1 所示：

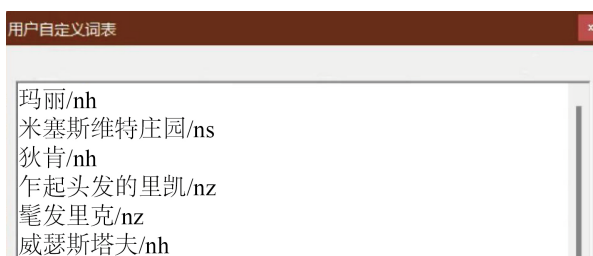


Figure 1. Customised Word List function of Corpus Word Parser

图 1. Corpus Word Parser 软件的“用户自定义词表”功能

- 将软件自动分词并标注完的文本经过人工再处理后另行保存，提高语料精度便于之后研究步骤中的进一步使用和处理。

3.2. 语料检索与分析

本研究在语料库之检索与分析上使用语料库检索软件 AntConc 和文字编辑软件 EmEditor，通过 AntConc 之词表、主题词、搭配等功能和 EmEditor 之正则表达式搜索功能，分别从类符/形符比、词密度、叠词和拟声词等词汇范畴，平均句长、句段长、译者注释等句式方面考察两个译本的文本特点，对两位译者的翻译风格做出一定程度的量化分析。

4. 分析过程和讨论

4.1. 词汇层面

4.1.1. 类符/形符比

类符/形符比(Type/Token Ratio, 下称“TTR”)所反映的是被考察文本中词汇使用上的丰富程度以及它们变化的大小[2]。通常情况下，二者间的比值越高，说明文本的用词越丰富，反之，越小则说明文本的用词更加简单和单调。但具体情况下，某一特定时期中某一特定语言所普遍使用的词汇量是有限的，文本内容的增加会使得语料库的容量变大，形符数也很会相应增加，但类符数在这种情况下不一定会按比例增加，甚至不会增加。这种情况就可能使得语料库当中类符与形符的比越来越小[12]。因此，按照一定长度(本研究以 1000 为计算基准)将文本分批计算类符形符比的标准类符形符比(Standardized Type/Token Ratio, 下称“STTR”)能更好地衡量语料库的词汇使用。通过将语料文本导入 AntConc 软件获得类符数、形符数后通过计算得出相关数据，如表 1 所示：

Table 1. TTR and STTR of the corpus texts**表 1.** 语料文本的类符/形符比与标准类符/形符比

	形符数	类符数	类符/形符比(%)	标准类符/形符比(%)
英文原文	82989	4779	5.76	38.01
李译	88539	8151	9.21	47.18
梅译	76615	7138	9.32	47.38

从表中可以看出,李译版本在 TTR 和 STTR 两方面均低于梅译版本,这说明梅译版本在用词上较李译更加丰富。作为一部经典的儿童读物,自然也应当从儿童文学的角度来思考译者的翻译风格。对于儿童来说,用词重复性更高的李译可能更容易接受更适合阅读,但这仅是一个方面,并不能完全判断译本的好坏。

4.1.2. 词汇密度

词汇密度可以看作是分析译文文本风格的一个重要参照。一般情况下,它指的是译文中实义词在整个文本中所占的比重,它的数值可以反映文本中所含的信息量大小及文本的难易度[13]。本研究根据王克非[12]的观点,将英文中的名词、实义动词、形容词和副词界定为实词,将助动词、情态动词、代词、介词、冠词和连词界定为虚词;将中文里的名词、动词、形容词、副词、数词和量词界定为实词,将连词、介词、助词、叹词和语气词以及代词这种半虚词界定为虚词。通过软件对语料进行搜索、筛选和计算得出表 2 内容:

Table 2. Lexical density of the corpus texts**表 2.** 语料文本的词汇密度

	形符数	类符数	词汇密度(%)
李译	47715	88539	53.89
梅译	43797	76615	57.16

通过表中数据可以看出,李译词汇密度,即实词数占总词数的比值为 53.89%,较梅译的 57.16% 更小,说明李译版本译者在翻译时使用了更多的其语法功能的虚词来扩充译文,而梅译版本译者则更注重精简表达原文意。例如:

例 1:

原文: Once she crept into the dining-room and found it empty, though a partly finished meal was on the table and chairs and plates looked as if they had been hastily pushed back when the diners rose suddenly for some reason.

李译: 有一次,她爬到餐厅里去,发现那儿空无一人,不过饭桌上、椅子上有些盘子,里面放着些没有吃完的东西,看得出不知是为了什么原因,吃着饭的人慌忙中把盘子一推,突然就站起身来离开了。

梅译: 有一次,她溜进餐厅,发现里面空荡荡的。不过,桌上还有些残羹剩饭。椅子和盘子似乎被匆匆地推了回去,仿佛吃饭的人因为某种原因,突然起身离开了。

从例 1 中可以看出,李译较梅译句幅更长,多出来的部分多为解释性的功能词。这种情况的优势在于能帮助读者更好地理解原文内容,但是出现过多的情况下会使得文字冗杂、过于拖沓,甚至增加儿童的阅读障碍。但是单凭这一个因素也无法完全考察译者翻译风格对作品阅读的影响,还需从之后的内容

中继续分析。

4.1.3. 特殊词汇

儿童文学作品有着独特的语言特色，其中典型的一点便是特殊词汇的使用。胡显耀认为，较之成年人，年龄越小的儿童对形象化语言的依赖越强[14]。对于这样的重点目标读者群体来说，译文在内容和形式上要符合儿童的认知规律和他们的心理特点[15]，叠词、拟声词、叹词和儿化音这样的特殊词汇的使用，在儿童文学作品的翻译中有着举足轻重的作用，不仅能增添译文的音律性，使文字更加工整美观，也能增添文字的趣味性，使之更加生动形象，让孩子们更爱阅读。本研究通过对语料进行搜索、筛选，经过统计后得出表3内容：

Table 3. Some special words in the corpus texts

表 3. 语料文本中的一些特殊词汇

	叠词	叹词	拟声词	儿化词
李译	1273	83	232	824
梅译	1535	89	254	696

从表中数据可以明显看出两个译本之间的区别和特点。从叠词(如“光秃秃”“蹦蹦跳跳”“得意洋洋”等)、叹词(如“唉”“噢”等)和拟声词(如“窸窸窣窣”“叽叽喳喳”等)这三方面看来，梅译均较李译数量更多；而从儿化词(如“今儿”“小不点儿”等)方面看来，李译又较梅译更多。这也体现出两个译者在这些特殊词汇上有着不同的侧重点。例如：

例 2:

原文：There were trees, and flower-beds, and evergreens clipped into strange shapes, and a large pool with an old gray fountain in its midst.

李译：这儿有宽阔的草坪，有迂回曲折的小径，路两边都是修剪过的树篱。这里有树木，有花坛，有修剪成各种形态的冬青树，还有一个很大的池塘。

梅译：这里有宽阔的草坪、弯弯曲曲的小径、修剪过的狭长绿化带、树木、花坛、修剪成各种奇妙形状的常绿植物，还有一个大池塘。

例 2 是语料中很典型的一个例子，从其中可以看出，李译会用儿化的方位词，而梅译会使用叠词来对场景进行形容。相比较而言，李译的用词会显得稍微文雅一些。但两位译者均从不同的角度强化了自身对于儿童文学作品翻译的策略，增加了译文的可读性。

4.2. 句式层面

4.2.1. 平均句长与平均句段长

平均句长指的是一个文本语料中所有句子(Sentence)的平均长度，该长度通常以句子所包含的词数或字数所表示。Olohan 指出，平均句长可以用来研究译者的翻译风格，平均句长越长，说明文本难度越大，可读性越低[16]。平均句长的计算方法为总形符数除以句子数，以句号、问号和感叹号作为句子划分标记[17]。通过 EmEditor 软件搜索语料标记后通过计算可以得出两个译本各自的平均句长，此处以表示语句终止的句号、叹号和问号为语料标记，详细数据如表 4 所示。

根据计算得出的数据显示，英文原文的平均句长为 13.17，两个汉译本的平均句长分别为 14.56 和 11.85。由此可见李译在平均句长上较为接近英文原文，但较梅译会长出不少。为了进一步分析句式层面带来的影响，还应参照划分更为详细的内容。在语料库研究中，相对于句长分析来说，句段(Syntagma)

的长度更能反映汉语翻译语言的个性特征[18]。在这一部分研究中,进一步选取了逗号、冒号等标点符号作为语料标记,依旧通过计算得出表 5 内容。

Table 4. Mean sentence length of the corpus texts

表 4. 语料文本的平均句长

	句号	问号	叹号	句子数量	形符数	平均句长
英文原文	5327	441	534	6302	83989	13.17
李译	5082	511	490	6083	88539	14.56
梅译	5351	522	593	6466	76615	11.85

Table 5. Mean Syntagma length of the corpus texts

表 5. 语料文本的平均句段长

	冒号	逗号	分号	句段数量	形符数	平均句段长
英文原文	18	3131	79	9530	82989	8.71
李译	90	7432	5	13610	88539	6.06
梅译	99	8311	14	14491	76615	5.28

从平均句段长的表中数据来看,两个汉译本的平均句段长均低于英文原文,这体现的是汉语本身的语言文字特点。但从两个汉译本之间的比较看来,李译句段数量更少,平均句段长长于梅译,这说明梅译的句法结构更为简单,每个句段所负载的信息量更少。例如:

例 3:

原文: When they sat down on the grass with Captain curled at their feet, Soot solemnly listening on a tree and Nut and Shell nosing about close to them.

李译: 当他们在草地上坐下来时,“船长”在他们脚下蜷成一团,“煤烟”则栖在了一棵树上,很严肃地静静谛听,“坚果”和“贝壳”则在他们附近嗅嗅闻闻。

梅译: 两人坐在草地上,“船长”蜷成一团,趴在他们脚边。“煤灰”栖在树上,正儿八经地听着。“果仁”和“贝壳”则在他们身边东嗅嗅、西闻闻。

例 3 中可以看出,李译明显在每个句段上都会长于梅译。分析其成因有两方面因素:一方面是梅译多以短句为主,李译较前者多使用复合句,使得每个句段负载的内容变多;另一方面,同之前的分析相结合来看,李译多在译文中增添信息,帮助读者理解内容,也好重复使用辅助性的连词、代词等将句子结构变得更加连贯。相较而言,二者都为读者的观感做出考虑,希望译本能便于读者尤其是儿童群体理解。

4.2.2. 其他因素

在对比研究两个汉译本的过程中发现英文原文和李译均有着明显更多的破折号的使用。在中文和英文中,破折号都有着近似的功能和作用,比方说解释说明、引出下文、声音的延长、话语的转折、交代来源等等。英文中可以称为破折号的标点符号有三种,分别是“连字符(hyphen, ‘-’)”“短破折号(en dash, ‘-’)”和“长破折号(em dash, ‘—’)”,由于它们作用之不同,此处选取的是与中文破折号概念相同的 em dash。在进一步的分析过程中还发现,李译较梅译有更多的译者注释,为了更好阐明这方面的内容,通过软件统计相关数据后如表 6 所示:

Table 6. Em dashes and Translator's Notes in the corpus texts
表 6. 语料文本中的破折号与译注

	破折号	译注
英文原文	483	0
李译	327	7
梅译	131	3

从表中数据可以看出，两个中文译本的破折号使用都明显要低于英文原文，而且李译的确在数量上达到了梅译的两倍不止。例如：

例 4：

原文：He looked so strange and different because a pink glow of color had actually crept all over him—ivory face and neck and hands and all.

李译：他显得如此陌生，如此特别，因为真的有一道粉红色的亮光爬满他的全身——打在他原本白里透黄的脸、颈、双手和所有的部位上。

梅译：粉红的亮光正在他周身蔓延，爬上那象牙般苍白的脸、脖子，渐渐将他整个人都笼罩起来。他变得好奇异，好陌生呀！

从例 4 中可以看出，李译在翻译风格上更喜好使用破折号来还原英文原文，而梅译更喜好改变句式，使用更贴合中文习惯的行文。对于中文环境的儿童读者来说，或许梅译更能贴合他们的阅读习惯，更加生动有可读性。至于译注方面，由于原文是儿童文学作品，文字难度低，不需要过多注释，但李译本还是更为贴心的为小读者们在较多必要的地方做出注释，帮助他们更好地了解一些日常可能尚未涉及到的内容，扩充知识面。

5. 总结

本文通过自建语料库，以语料库翻译学为研究框架，使用语料整理和检索工具，主要从文本的词汇与句式层面考量，对两位不同译者对儿童文学作品《秘密花园》的汉译本做出对比与分析。通过研究发现，两个译本均使用连贯的句式，较为简单平实的语言，虽侧重不同但更加生动的词汇表达，使得译文整体阅读难度不高。相较之下，李文俊之译本在行文和语篇结构上更加西式，语言表达更加明确清晰，缺点是句段偏长；梅静之译本表达更加中式，用词更加丰富，句子更加简练，不足之处可能在于更加丰富的词汇使用会增加儿童受众的阅读难度。从研究中我们可以看出，为了增加译文的可读性和接受程度，译者需要考量受众的整体认知水平和喜好，在此基础上发挥自身翻译风格使得译文更加具有特色。

本研究仍存在不足，今后或许可以一次性选择更多的译本纳入研究范畴，亦或是将研究的重点再次细化，从更加深刻的角度利用语料库做出译者风格上的研究。

参考文献

- [1] 陈福宇. 人性的光辉, 人格的升华——《秘密花园》人物思想解读[J]. 名作欣赏, 2012(27): 103-104.
- [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] Saldanha, G. (2011) Translator Style. *The Translator*, 17, 25-50. <https://doi.org/10.1080/13556509.2011.10799478>
- [4] 董帅, 崔校平. 新闻体翻译换言标记语内“显化”特征: 基于语料库的计量分析[J]. 外语学刊, 2022(4): 19-27. <https://doi.org/10.16263/j.cnki.23-1071/h.2022.04.003>
- [5] 徐翠, 李德超. 同传专业译员预制语块使用特征: 基于语料库的对比研究[J]. 外语教学与研究, 2021, 53(4):

582-593+640. <https://doi.org/10.19923/j.cnki.fltr.2021.04.009>

- [6] 赵轩. 基于语料库的《红楼梦》诗词翻译风格对比研究——以《好了歌》两德译本为例[J]. 开封文化艺术职业学院学报, 2021, 41(9): 45-47.
- [7] 郑一菲. 基于语料库的《瓦尔登湖》译本对比研究[J]. 散文百家(理论), 2021(2): 138-140.
- [8] 赵颖. 基于语料库的《道德经》两译本的翻译风格研究[J]. 中国翻译, 2015, 36(4): 110-113.
- [9] Frances, H.B. (1998) *The Secret Garden 100th Anniversary*. Harper Collins, New York.
- [10] 弗朗西斯·霍奇森·伯内特. 秘密花园[M]. 李文俊, 译. 北京: 中信出版社, 2018.
- [11] 弗朗西斯·霍奇森·伯内特. 秘密花园[M]. 梅静, 译. 云南: 云南美术出版社, 2017.
- [12] 王克非, 胡显耀. 基于语料库的翻译汉语词汇特征研究[J]. 中国翻译, 2008, 29(6): 16-21+92.
- [13] 吕鹏飞, 陈道胜. 基于语料库的《论语》英译本翻译风格比较研究——以辜鸿铭和亚瑟·威利两译本为例[J]. 上海翻译, 2021(3): 61-65.
- [14] 胡显耀, 李力. 高级文学翻译[M]. 北京: 外语教学与研究出版社, 2009: 209.
- [15] 贾晓玉. 基于语料库探析儿童文学翻译风格[J]. 文化学刊, 2020(8): 189-191.
- [16] Olohan, M. (2004) *Introducing Corpora in Translation Studies*. Routledge, London. <https://doi.org/10.4324/9780203640005>
- [17] 胡开宝. 语料库翻译学概论[M]. 上海: 上海交通大学出版社, 2011.
- [18] 秦洪武. 英译汉翻译语言的结构容量: 基于多译本语料库的研究[J]. 外国语, 2010(4): 73-80.