

# 基于平行语料库的《红楼梦》 英译本文体风格研究

□李敏杰 朱 薇 [中南民族大学 武汉 430074]

**[摘要]** 已有的《红楼梦》英译研究多属赏析性研究,主观性较强,在深度和广度上都远远不够。基于《红楼梦》汉英平行语料库,研究者可以统计霍译本和杨译本的形符数、类符/形符比、平均词长、高频词、平均句长等,发现两者的不同特点。总体上,较之杨译本,霍译本显化特征更明显,这有利于西方读者的理解和接受;杨译本更强调译作对原作的重视,同时更强调译作的文学性,因而表达方式更加丰富、生动。

**[关键词]** 红楼梦; 语料库; 形符/类符比; 平均词长; 平均句长

**[中图分类号]** I045

**[文献标识码]** A

**[文章编号]** 1008-8105(2012)03-0090-05

《红楼梦》是中国古典小说发展的最高峰,其思想内容博大精深,艺术成就精湛无比。自20世纪初王国维、蔡元培、胡适等人相继发表研究《红楼梦》的文章和著作,开创了红学研究的小说批评、红学索隐、红学考证三大学派以来,新红学研究已逾百载。上世纪70年代,英国汉学家霍克斯先生和我国翻译家杨宪益夫妇先后推出了《红楼梦》的英文全译本。之后,国内外许多学者对《红楼梦》的这两个英译本展开研究,形成了一股不亚于“红学”研究的英译研究热潮。

闫敏敏分析了1980~2003年20多年当中刊登在13种外语类期刊上的有关《红楼梦》英译的文章,发现研究的焦点主要在诗词、文化内容、书名、人物语言、警句和习语、回目、人名、修辞等方面。这说明《红楼梦》英译研究的内容越来越广泛且越来越深入了。然而,比起《红楼梦》这一文学巨著和高潮迭起的红学研究,《红楼梦》英译研究的深度和广度还远远不够。从整体上看,“《红楼梦》英译研究基本上还处于比较零散的状态,系统性还是远远不够的”,且大多“采用译文比较的方法,即比较-分析-结论的模式,从总体上看研究方法还是比较单一的。”<sup>[1]</sup>这些研究多属赏析性研究,即研究者多从文学价值观、审美观出发,对译文作出个人评价。因此,这种评价手段与研究者的文学观念、

美学素养、价值观等息息相关,主观性较强,其结论难免失之偏颇,在一定程度上缺乏可靠性。究其原因,主要在于:1)《红楼梦》篇幅长、内容广,单个研究者很难有足够的时间和精力进行大规模的、系统的研究;2)缺乏现代科学技术的支持,因而很难进行定量统计与分析。

在国外,随着翻译研究视域的不断拓宽,实证的方法在翻译研究中逐步受到重视。在这种注重实证和描写的趋势下,翻译理论家将目光投向语料库翻译研究。Mona Baker第一个倡导把语料库应用于探究翻译现象,并展开基于语料库的翻译研究。之后,许多学者如Kenny和Laviosa等利用语料库研究翻译文本的特点;Ebeling用翻译语料库研究语言之间的异同;Bowber和Zanettin et al指出语料库可以作为帮助翻译和查找术语的资源库。受此影响,国内学者如王克非、柯飞等运用语料库对翻译的普遍性进行研究。

为了对《红楼梦》英译本的文体特征进行宏观分析,笔者自建了《红楼梦》中英文平行语料库。语料库中的中文文本采用中国艺术研究院红楼梦研究所校注、人民文学出版社1996出版的《红楼梦》。此版本以庚辰本为底本,是一个“学术价值很高、接近曹雪芹原稿的珍贵本子。”(冯其庸语)英文文本选用迄今为止最有影响力的两个全译本,分别为

**[收稿日期]** 2011-09-13

**[基金项目]** 中南民族大学中央高校基本科研业务费专项资金项目“基于平行语料库的《红楼梦》英译研究”资助(ZSQ10031)

**[作者简介]** 李敏杰(1973-)男,博士研究生,中南民族大学外国语学院副教授;朱薇(1981-)女,硕士,中南民族大学外国语学院教师。

英国汉学家霍克斯和闵福德(David Hawkes & John Minford)翻译、企鹅出版社出版的The Story of the Stone(以下简称“霍译本”)<sup>[2]</sup>, 和中国翻译家杨宪益及夫人戴乃迭翻译、外文出版社出版的A Dream of Red Mansion(以下简称“杨译本”)<sup>[3]</sup>。需要指出的是, 虽然在1830年至今年近180年间, 先后出现了多种英译本, 但大多属片段翻译或节译本, 只有霍译本、杨译本以及另一译本: 邦斯尔神父译本属真正的全译本。邦斯尔神父译本目前尚未正式出版, 由其儿子将译本手稿捐给香港大学图书馆, 供人们网上免费阅读(网址为<http://lib.hku.hk/bonsall/hongloumeng/index1.html>), 因此本语料库未包括此译本。

基于平行语料库的《红楼梦》英译研究, 我们可以对语料进行自动处理, 对译本的文体特征、语言特色进行定量分析。Tognini-Bonelli认为, 翻译语料库能揭示语言之间的对应与差异(cross-linguistic correspondances and differences), 而这是单语料库无法完成的<sup>[4]</sup>。例如, 我们可以对两个译本的形符数、类符数、类形符比、词汇密度、词频、词长、句数、平均句长、搭配模式等进行统计, 发现译本间的差异, 并对照分析原文, 从认知、语用和翻译策略等角度分析这些差异的内在动因。

笔者借用语料库常用工具软件WordSmith4.0中的Word List工具, 对两个英译本进行统计, 得到两译本中词汇的基本数据, 包括类符/形符比(type/token ratio, TTR)、标准化的类符/形符比(standardised TTR)、平均词长(mean word length)、平均句长等。

以下将从词汇、句子、段落层面进行探讨, 对比分析两个英译本的语言特征和文体风格。

## 一、词汇层面

在词汇层面上, 我们可以对比两个译本的形符(tokens)、类符(types/distinct words)、类符/形符比(type/token ratio, TTR)、标准化的类符/形符比(standardised TTR)、平均词长(mean word length)、词长的标准差(word length std. dev.)等, 揭示两个译本的用词特点。

### (一) 类符/形符比(TTR)

形符是指语料库中出现的所有词, 而类符指语料库中所有不同的词, 排除重复、忽略大小写和屈折变化(含单复数、时态、语态)。Mona Baker认为, 类符/形符比是不同的词语与全文词汇总量的比值, 可以用以测量写作者或特定语料库的用词范围

(range)和多样性(diversity)<sup>[5]</sup>。一般而言, 类符/形符比越高, 表明写作者词汇使用越丰富; 反之, 类符/形符比越低, 则表明写作者的不同词汇使用量越小。不过, 在一定时期内, 语言的词汇量是有限的, 如果语料库容量不断扩大, 则形符数不断增加, 而类符数不一定增加, 导致语料库容量越大, 形符/类符比反而越来越小, 因而不同容量的语料库的形符/类符比缺乏可比性。为了避免这一问题, Mona Baker指出, 在比较的文本长度不一时, 较之类符/形符比, 标准化的类符/形符比更可靠, 即按照一定长度(通常为1000个形符)分批计算文本的形符/类符比, 再求其平均值<sup>[5]</sup>。

表1

项目	霍译本	杨译本	BNC
形符数	852,399	632,506	102,467,488
类符数	22,061	18,106	166,962
类符/形符比	2.59	2.86	0.16
标准化类符/形符比	42.77	44.16	41.20

由表1可以看出, 霍译本的形符数为852,399, 而杨译本的形符数为632,506。形符数是显化特征的重要标志。大量研究表明, 译语有趋于显化之特征。显化的概念最早由法国学者Vinay&Darbelnet提出, 指翻译中“将原语中隐含的信息在译语中加以明示, 此类信息在原文中可以根据语境或情境获得。”<sup>[6]</sup>美国翻译理论家Nida et al也指出, 好的译文往往比原文长, 因为在翻译过程中, 译者往往会通过明示原文含蓄的信息而适当增加译文的冗余度<sup>[7]</sup>。因而, 从形符数量上看, 霍译本的显化特征较杨译本更明显。究其原因, 作为英国人, 霍克斯深知本国读者对中国语言文化的了解程度, 因而在翻译过程中处处考虑西方读者的接受能力, 努力为他们扫除障碍, 这势必需要更多的解释性语言。而杨宪益则认为, 我国人民应该了解外国的文化遗产, 外国也应该了解中国丰富的文化遗产。为了翻译中国文化遗产, 忠实传达中国文化的价值, 杨宪益强调“翻译时不能作过多的解释。译者应尽量忠实于原文的形象”, 否则, “就不是翻译, 而是改写了。”<sup>[8]</sup>两译本之间的这种差异可以从以下例句看出:

刘姥姥说: “这倒不然, 谋事在人, 成事在天。咱们谋到了, 靠菩萨的保佑, 有些机会, 也未可知。……”(第六回)

霍译: “I wouldn't say that,” said Grannie Liu, “Man proposes, God disposes. It's up to us to think of

something. We must leave it to the good Lord to decide whether He'll help us or not. Who knows. He might give us the opportunity we are looking for..."

杨译: "Don't be so sure," said Granny Liu. "Man proposes, Heaven disposes. Work out a plan, trust to Buddha, and something may come of it for all you know..."

两译本对原文划线部分进行了不同的处理。除对“菩萨”一词有不同的翻译外,两个译本在组句方面存在较大差异。霍译本使用三个句子翻译划线部分,句子地道、流畅,口语化特征明显,更易于为西方读者接收。而杨译本只使用了一个句子来翻译此句,口语化特征不如霍译本明显。

从标准化类符/形符比看,霍译本为42.77,杨译本为44.16,而BNC为41.20。可以看出,杨译本的标准类符/形符比最高,其次分别为霍译本和BNC。这表明:1)两个译本作为文学作品,其语言较之英语国家常用语言丰富;2)在一定程度上,杨译本的用词范围更加宽泛,表达方式更加丰富、生动;3)霍译本的用词相对狭窄,但更接近英语本族使用者的用词变化。另外,杨译本的标准类符/形符比,更接近于英语翻译语料库(TEC)中小说类分库的标准类符/形符比44.63<sup>[9]</sup>,这表明杨译本作为翻译小说而存在的特征较霍译本明显。

### (二) 平均词长

平均词长是文本中类符的平均长度。一般而言,平均词长较长则表明文本中的长词较多,文本更复杂深奥;反之,词长较小,则文本越浅显易懂。WordSmith软件在作词频统计时,能够自动计算文本的平均词长。以下为两个译本和BNC的统计数据:

表2

	霍译本	杨译本	BNC
平均词长	4.19	4.23	4.54

霍译本的平均词长为4.19,杨译本的平均词长为4.23,而BNC的平均词长为4.54。这表明:从总体上看,两个译本的用词复杂程度接近,均略低于本族英语使用者。究其原因,在于两个译本作为翻译文本,目的在于使英语读者能够理解、接受这部中

国文学作品。若使用大量复杂的词汇,则势必影响英语读者的阅读与接受。因而,两个译本体现出了其作为翻译作品的共同特性。

此外,为了更详尽地表现文本中不同长度词语的分布,WordSmith软件还能自动计算各长度词语出现的频率。下表显示的是两个英译本和BNC中各长度词所占比例:

表3 两译本和BNC主要词长分布百分比

词长	霍译本	杨译本	BNC
1-letter words	3.19	2.81	4.71
2-letter words	17.08	15.63	17.25
3-letter words	25.51	24.87	20.39
4-letter words	20.93	21.21	17.67
5-letter words	11.06	12.39	10.77
6-letter words	7.46	8.60	7.94
7-letter words	6.15	6.72	7.29
8-letter words	3.77	3.90	5.13
9-letter words	2.33	2.12	3.75
10(+)-letter words	2.52	1.76	5.1

由统计结果可以看出:1)在两个译本中,由3个字母组成的词所占比例最多,均在24%以上,其他依次为4字母词(20%以上)、2字母词(15%以上)、5字母词(11%以上),这一排列次序与BNC相同;2)两个译本中3、4个字母的词汇所占百分比明显高于BNC,而长难词(如7个以上字母的词汇)明显低于BNC;3)杨译本中1、2、3、4个字母的词汇所占百分比均低于霍译本,而5、6、7、8个字母的词汇均高于霍译本。由此,我们可以得出结论:1)总体上看,两个译本均倾向于使用日常的、通俗易懂的词汇,以便于西方读者阅读;2)与霍译本相比,杨译本更注重译作作为文学作品而存在的事实,因而选择了一些文学性强、较复杂深奥的词汇。

### (三) 高频词表

WordSmith提供的WordList功能,除能计算词频统计外,还能按照字母顺序、使用频率提供普通词频表。表4为两译本和BNC的高频词列表:

表4

次序	霍译本			杨译本			BNC		
	词汇	频率	百分比	词汇	频率	百分比	词汇	频率	百分比
1	THE	26850	4.25	THE	36921	4.33	THE	6,074,315	5.93
2	TO	21652	3.42	TO	28868	3.39	OF	3,062,469	2.99

(续表)

3	AND	17890	2.83	AND	25264	2.96	AND	2,629,938	2.57
4	A	11371	1.80	OF	19724	2.31	TO	2,613,090	2.55
5	OF	10839	1.71	A	16866	1.98	A	2,189,709	2.14
6	YOU	8968	1.42	YOU	12481	1.46	IN	1,963,194	1.92
7	HER	8715	1.38	IN	12100	1.42	THAT	1,121,864	1.09
8	IN	8470	1.34	HER	11740	1.38	IT	1,058,066	1.03
9	SHE	7926	1.25	THAT	9943	1.17	IS	997,109	0.97
10	HE	6351	1.00	SHE	9677	1.14	FOR	890,134	0.87
总计		129032	15.14		183584	29.02		22,599,888	13.14

由列表可以看出: 1) 两个译本中使用频率最高的前3个词完全相同, 前5个词与BNC语料库完全相同, 为the, to, and, a, of, 但与后者排列次序有较大差异; 2) 两个译本中, 前10个高频词中, 且均含人称代词you, her, she, 这说明译语文本属于小说体, 叙述人物、事件较多, 且《红楼梦》人物中女性较多。

## 二、句子层面

词是可以独立使用的最小语言单位, 而句子则是由词或词组构成的能够表达一个完整意思的语言单位。因此, 基于语料库的翻译研究, 不能仅仅停留于词的层面, 同时必须考虑句子层面的特征。Olohan认为, 与形符/类符比一样, 平均句长(每句所含词数)也是译者风格的一般标记。利用WordSmith软件统计, 我们可以获得两个译本的句子总数和平均句长:

表5

	霍译本	杨译本
句子总数	57,506	49,635
平均句长	14.82	12.73

由表5可知, 杨译本的平均句长低于霍译本。这表明了两者明显的风格差异, 霍译本喜欢用较长的句子, 而杨译本句子相对较简短。这表明, 霍译本趋于显化, 可能与其为英语读者着想而在翻译中使用更多的修饰性、解释性语言有关。而杨译本受汉语句子、句段和篇章表达习惯的影响较大, 相比之下更难于为西方读者理解, 这也与前文的类符/形符比相印证。两个译本在平均句长上的差异, 可以由下例看出:

当时街坊上家家箫管, 户户弦歌。当头一轮明月, 飞彩凝辉, 二人愈添豪兴, 酒到杯干。雨村此时已有七八分酒意, 狂兴不禁, 乃对月寓怀, 口号

一绝云: .....

霍译: The sounds of music and singing which could now be heard from every house in the neighbourhood and the full moon which shone with cold brilliance overhead seemed to increase their elation, so that the cups were emptied almost as soon as they touched their lips, and Yu - cun, who was already a sheet or so in the wind, was seized with an irrepressible excitement to which he presently gave expression in the form of a quatrain, ostensibly on the subject of the moon, but really about the ambition he had hitherto been at some pains to conceal:...

杨译: The sound of flutes and strings could be heard from all the houses in the neighbourhood; everywhere was singing; and overhead the bright moon shone in full splendour. The two men became very merry and drained cup after cup. Yucun, eight-tenths drunk, could not suppress his elation. He improvised a quatrain to the moon and declaimed it:...

英汉语言分属不同的语系。英语重形合, 因而长句在英语中很常见。而汉语重意合, 强调句子间的内在逻辑关联, 句子一般不长。上例中, 原文为三个句子。霍译本使用大量连接词, 使之成为一个长句。虽然句子逻辑严密, 但失去了原文的轻松活泼, 风格上略显滞重。相比之下, 杨译本使用四个句子来翻译, 句子较短, 简洁明快, 受汉语语言表达习惯的影响更明显。

由以上分析我们可以看出, 在形符数、类符/形符比、平均词长、高频词、平均句长等方面, 霍译本和杨译本具有不同的特点。总体上, 两个译本均倾向于使用日常的、通俗易懂的词汇, 以便于西方读者阅读; 较之杨译本, 霍译本显化特征更明显, 这有利于西方读者的理解和接受; 杨译本更强调译作的文学性, 因而用词范围更加宽泛, 表达方式更

加丰富、生动, 而霍译本的用词相对狭窄, 但更接近英语本族使用者的语言。

### 参考文献

- [1] 闫敏敏. 二十年来的《红楼梦》英译研究[J]. 外语教学, 2005, 26(4): 64-68.
- [2] CAO Xue-qin. The story of the stone[M]. HAWKES D, MINFORD J, trans. London: Penguin Books Ltd., 1973-1986.
- [3] CAO Xue-qin. A dream of red mansions[M]. YANG Hsien-yi, YANG Gladys, trans. Beijing: Foreign Languages Press, 1994.
- [4] TOGNINI-Bonelli E. Corpus linguistics at work[M]. Amsterdam: John Benjamins Publishing Co, 2001.
- [5] BAKER M. Towards a methodology for investigating the style of a literary translator [J]. Target, 2000, 12(2): 241-266.
- [6] KLAYDY K. Explication. BAKER M. Routledge Encyclopedia of translation studies[M]. Shanghai: Shanghai Foreign Language Education Press, 2004: 81.
- [7] NIDA E A, TABER C. The theory and practice of translation[M]. Leiden: E. J. Brill, 1969: 164.
- [8] 任生. 杨宪益的文学翻译思想散记[J]. 中国翻译, 1993, (4): 33-35.
- [9] OLOHAN M. Introducing Corpora in Translation Studies[M]. New York:Routledge,2004: 80.

[10] BAKER M. Corpus linguistics and translation studies: implications and applications[C]//BAKER M. Text and technology: In honour of John Sinclair. Amsterdam & Philadelphia: John Benjamins, 1993:233-250.

[11] BOWKER M. Corpus-based application for translator training: Exploring the possibilities[C]//GRANGER S, LEROT J, PETCH-Tyson S (ed). Corpus-based approaches to contrastive linguistics and translation studies. Amsterdam & New York: Rodopi, 2003.

[12] EBELING J. Contrastive linguistics, translation, and parallel corpora[J]. Meta, 1998, 43 (4):602-615.

[13] KENNY D. Creatures of habits? What translators usually do with words? [J] Meta, 1998, 43 (4): 515-523.

[14] LAVIOSA S. The corpus-based approach: a new paradigm in translation studies [J]. Meta, 1998, 43 (4): 474-479.

[15] ZANETTIN F, SILVIA B, STEWART D. (ed). Corpora in translator education[C]. Manchester: St Jerome Publishing, 2003.

[16] 曹雪芹, 高鹗. 红楼梦[M]. 中国艺术研究院红楼梦研究所校注(第二版). 北京: 人民文学出版社, 1996.

[17] 柯飞. 翻译中的隐和显[J]. 外语教学与研究, 2005, 37(4): 303-307.

[18] 王克非, 秦洪武. 英汉译语言特征探讨——基于对应语料库的宏观分析[J]. 外语学刊, 2009, (1): 102-105.

## A Parallel-Corpus Based Study of Stylistic Features of Honglougeng's English Versions

LI Min-jie ZHU Wei

(South-central University for Nationalities Wuhan 430074 China)

**Abstract** Studies on English versions of *Honglougeng* are mostly commentary, which are usually subjective, so these studies are limited in depth and breadth. With the Chinese-English parallel corpus of *Honglougeng*, researchers can get the numbers of tokens, type/token ratio, mean word length, high-frequency word lists, mean sentence length, etc, and discover their different stylistic features. On the whole, Hawkes' version demonstrates more explicitation traits than Yangs' version, which helps western readers comprehend and accept this literary work. Instead, Yangs' version is more faithful to the source text, and emphasizes the literariness of target text, so it is more expressive.

**Key words** Honglougeng; corpus; type/token ratio; mean word length; mean sentence length