

基于语料库的《封神演义》哪吒篇四个英译本的译者风格研究

杨杰浩

武汉工程大学外语学院, 湖北 武汉

收稿日期: 2023年11月21日; 录用日期: 2024年2月20日; 发布日期: 2024年2月27日

摘要

本文通过自建《封神演义》哪吒篇(第十二回至十四回)汉英双语平行语料库, 分析、考察四个译本的译者风格。分析结果得出顾译文本内容相较其他三个译本较全面, 使用句子数量最多, 受源语影响较大, 可读性最强。玛译词汇密度最低, 类符最低, 句子长度最短, 句子数量居中, 说明玛译为贴近儿童受众群体, 有意识降低文本阅读难度, 易读性高。不知译的句子数量最少, 平均词长最长, 平均句长最长, 标准类符/形符比最高, 说明该译本用词变化程度高, 文本阅读难度也随之提高, 对读者受众有较高要求, 译文易读性最低。沃译句子数量是不知译的2倍, 比玛译略多, 标准类符/形符比比顾译高而低于其他两个译本, 平均词长最短, 说明沃译在选词上、用词变化丰富程度上较低于除玛译之外的其他两个译文。

关键词

《封神演义》, 双语平行语料库, 译者风格

A Corpus Based Study on Translator's Styles of Four English Versions of *Feng Shen Yan Yi*

Ericsson Yeung

School of Foreign Languages, Wuhan Institute of Technology, Wuhan Hubei

Received: Nov. 21st, 2023; accepted: Feb. 20th, 2024; published: Feb. 27th, 2024

Abstract

This paper analyzes the translation styles of four versions through the bilingual parallel corpus of Chinese and English from chapter 12th to 14th in *Feng Shen Yan Yi*. The results show that Gu Zhizhong's translation is more comprehensive than that of the other three, is influenced by the source

language and is most readable. Martens translated the lowest density of words, the lowest number of types, the shortest sentence length, the number of intermediate sentences, indicating that Martens was more closer to the children, consciously reduce the difficulty of reading. The version of the unknown translator has the lowest number of untranslated sentences, the longest average words, the longest average sentence length, and the highest standardised type/token ratio, indicating that the translation has a high degree of variation in terms, resulting in an increase in the difficulty of reading the text, which requires a high level of readability from the reader. Werner has twice the number of sentences than that of unknown translator, slightly more than that of Martens, the standardised type/token ratio is higher than that of Gu Zhizhong and lower than that of the other two translations, and the average length of words is the shortest, indicating that word-translated is less abundant in choice of words, and less varied in use than in the other two translations except Martens.

Keywords

Feng Shen Yan Yi, Bilingual Parallel Corpora, Translator's Style

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. 《封神演义》哪吒篇的汉英双语平行语料库简介

1.1. 语料库研究背景

中国共产党十九大指出文化乃国家、民族之灵魂。中国优秀传统文化的人文精神、道德规范，结合自主创新让中华文化展现独特风采。全球化背景之下，中国提出“走出去”国家战略，我国的文化软实力得到加强，国际话语权得到提升，这对与世界各国进行交流、树立文化自信具有重要意义[1]。

《封神演义》的翻译在时间长河中经历不同历史时期、不同传播媒介，小说中的中国传统文化精髓和内涵值得域外传播，我国《大中华文库》工程已将此小说囊括在内。本文基于自建的《封神演义》二译本平行语料库，对文字数据进行客观分析，希望揭示其译者风格，展现其小说语言特征，帮助中国古典文化更好走出国门。

1.2. 语料库介绍

《封神演义》哪吒篇的汉英双语平行语料库的汉语文本采用许仲琳编著、新世界出版社、外文出版社、湖南人民出版社出版的《封神演义》第十二回至十四回，英文文本分别为我国新闻学家以及翻译家顾执中的 *Creation of the Gods* 的第十二回至十四回(简称“顾译”)、某位不知名的译者在 *Investiture of the Gods* 中的第十二回至十四回(简称“不知译”)、英国汉学家沃纳(E. T. C. Werner)在 *Myths and Legends of China* 中的哪吒专题(简称“沃译”)、美国儿童作家玛顿斯(Frederick H. Martens)在 *The Chinese Fairy Book* 中的哪吒专题(简称“玛译”)。四个译本的词数分别约为1.2万、3千、4千、3千。需要指出的是，不知名译者的版本并未正式出版，但传播率极高，文本在书稿网站上下载并进行文本预处理工作；而沃译版本则是将哪吒的故事囊括进中国的神话小说文本内，作为其中一节；而玛译版本则是关于中国神话故事的儿童读物，面向群体较前三种不同。本文将检索和统计的结果进行对比研究，以期客观描述四个译本的语言使用和译者风格上之异同，得出《封神演义》全译本的译者风格。

1.3. 建库过程

源语文本和译语文本经过 OCR 识别和 Emeditor 降噪清洗，经人工校对后进行句子对齐和词性标注，最终形成可供检索的《封神演义》哪吒篇的汉英双语平行语料库。

1.3.1. 语料收集

首先，通过互联网在线书店购买了顾译、沃译，在开源平台下载了不知译和玛译文本。其次，利用 OCR 光学识别 ABBYY Reader 和文本格式转换，将收集的 PDF 格式文本预处理建成 DOCX 和 TXT 格式文本，为后续语料标注做好基础准备。

1.3.2. 语料降噪

由于上一步语料采集的源语文本以及顾译全译本并非由原始电子文本转成 PDF，并且文本语言为繁体文言文，因此内容字符之间存在太多噪音，因此语料降噪使用功能强大的 Emeditor 进行噪音的清洁以及人工校对。

1.3.3. 语料对齐

语料对齐是语料库建设的关键环节，其操作目的是在双语文本中互相对应其源语文本和目的语文本，借此进行句法和语篇分析[2]。第一步利用 ABBYY Aligner 进行句子层面对齐。由于源语文本和译语文本句对不同，故将 TXT 文件导入该软件后，由软件进行自动对齐，但一般对齐后的准确度无法达到研究要求，这时则需要人工手动对齐，通过合并和删除部分句子，使源语文本和译语文本句对一致。

1.3.4. 词性标注

本文采用 Tree Tagger 3.0 对四个译本进行单词标注，主要词性标注见表 1。研究过程中需要这四类实词：动词、名词、形容词、副词，这几种因素将在后续词汇密度统计时考察。

Table 1. POS annotation

表 1. 主要词性标注

附码	含义	附码	含义	附码	含义
JJ	Adjective	RB	Adverb	VVG	Gerund or present participle of lexical verb (living)
JJR	Comparative adjective	RBR	Comparative adverb	VVN	Past participle of lexical verb (lived, shown)
JJS	Superlative adjective	RBS	Superlative adverb	VVP	Present tense (other than 3rd person singular) of lexical verb (live)
NN	Common noun, singular or mass	VV	Lexical verb, base form (e.g. live)	VVZ	Present tense (3rd person singular) of lexical verb (lives)
NNS	Common noun, plural	NPS	Proper noun, plural	VVD	Past tense verb of lexical verb (e.g. lived)
NP	Proper noun, singular				

2. 语料库与译者风格

在语料库翻译学中，译者风格研究是其非常重要的组成部分之一。中外有不少学者都在从事译者风格的研究。译者风格是指通过一系列语言或非语言特征所表现出来的、与其他译者不同的个性特征[3]。译者的翻译风格具体表现为以下内容：对文本类型和翻译策略的选择；以及译者所使用的序文、后语以及脚注；文本内部释义等方法。译者风格也称译者翻译风格，指译者对翻译文本的选择、对翻译策略和方法的运用、对翻译文本语言使用的个性化特点[2]。以往相当长的一段时间里，翻译活动都被视为模仿，

译者只重现源语文本的风格，而随着描写性译学的发展，译者风格的研究才引起学界的关注，如译文中除了源语文本的作者外，还有译者的呼声[4]。通过语料库研究，学者大范围分析、统计语料，并分析译文本词汇特征、叙事结构和句法结构等几方面、从副文本视角和其他几个维度讨论译者风格在作品中的具体体现，审视译者本人的主动性[5]。通过语料库描述分析语言习惯，可以捕捉到对一些捉摸不定而又不起眼的语言特征，更有效地证明译者主观性[6]。

而在量化分析英语翻译语料库(Translational English Corpus, TEC)中的文学分库和英国国家语料库(British National Corpus, BNC)中的文学体裁分库里 say/tell + that 的搭配时，结果表明 TEC 里 that 使用多于 BNC [7] [8]。而 BNC 人称代词的使用频率显著高于 TEC [8]。以上研究结果都是源于英语源语与英语译语组成的可比语料库(comparable corpus)，没有译语和源语的对比，没有同一源语文本中不同译文的对比。

以类符/形符比、平均句长和语气助词等因素研究余光中和张藹玲两人不同版本的海明威《老人与海》译者风格，研究发现，两人均选择短句来复现海明威原作风格，但余光中版本的句子长度比张藹玲两更短，且其类符/形符比数值更大，这证明其用词丰富程度比张藹玲高[9]。

本文试图通过对《封神演义》哪吒篇(第十二回至十四回)中英文平行语料库中顾译、沃译、不知译和玛译进行词汇和句子层面的语内及语际考察，对比分析四个译本与 TEC、BNC 的语言使用情况，进而揭示四个译本的译者风格。

3. 词汇层面

本文借助 Antconc、WordSmith (6.0)等语料库工具，分别获得四个英译本中的词汇层面基本数据，包括类符(type)、形符(token)、类符/形符比(type/token ratio)、标准化类符/形符比(standardised type/token ratio)、平均词长、高频词等。四个译本的数据统计信息详见表 2~4。本文分析讨论标准类符/形符比、高频词、词长分布、词汇密度等方面，从这四个维度探讨语用特征，进而分析得出四位译者的不同风格。

3.1. 类符/形符比

类符指的是文本中出现不同的词，同时排除重复出现的数量并忽略大小写的情况；形符指的是文本中所有出现的词[3]。类符/形符比高低与词汇使用情况的丰富和多样成正比，即是说，当比值越高时说明作者用词越丰富，比值越小则说明其词汇范围窄，词汇变化少。同时，当文本长度大小不同时，类符/形符比的数值因类符聚集的均匀程度不同而受到影响，故标准化的类符/形符数值则更为可靠，具有参考性[7]。

Table 2. Word and sentence statistics based on four translations of Nezha chapters in *Feng Shen Yan Yi*

表 2. 《封神演义》哪吒篇四译本词句统计

	顾译	沃译	不知译	玛译
形符	12,463	4267	3050	3029
类符	2009	1134	898	821
类符/形符比	16.13	26.58	29.49	27.17
标准化类符/形符比	16.37	26.95	29.37	27.13
平均词长	4.12	4.05	4.63	4.12
句子数目	984	222	93	180
平均句长	12.65	19.22	32.74	16.79

观察表 2 可以得到以下信息, 在形符数量方面, 顾译最多, 大约为 1.2 万, 沃译次之, 不知译与玛译数量基本持平; 在类符数量方面, 顾译最高, 沃译次之, 不知译较沃译低, 玛译最少。不知译与玛译的形符相近, 类符上不知译高于玛译, 说明在用词上不知译的丰富程度比玛译高。顾译的形符远超出其他三个英译本, 说明其内容完整度最高。在类符/形符比方面, 不知译数值最高, 为 29.49, 玛译 27.17 高于沃译 26.58, 顾译最低, 仅为 16.13。在标准化类符/形符比方面, 数值大小的排序也与类符/形符比的顺序一致, 为不知译 > 玛译 > 沃译 > 顾译, 且顾译大约为不知译的 1/2。这说明, 在词汇使用丰富程度上, 顾译不及其他三位译者, 且大幅落后于不知译, 沃译与玛译持平, 不知译的词汇变化最多。英语翻译语料库(TEC)中小说类分语料库的标准化类符/形符比为 44.63 [10], 四个译文文本都与此标准有着较大差距, 原因在于此次选择的语料库文本来源于《封神演义》当中一个专题, 语料库内容较小。在平均词长上, 不知译为 4.63, 顾译与玛译相同, 沃译最低但也与顾译与玛译基本持平。英语翻译语料库小说类分语料库的平均词长为 4.36 [10], 四个译文文本都与此标准接近, 且不知译大于此标准, 体现了四个译本作为小说文体在平均词长上的共性。在句子数目上, 顾译 > 沃译 > 玛译 > 不知译。在译文文本单词数量相当的情况下, 玛译的句子数量是不知道译的 2 倍, 平均词长是不知道译的 1/2, 且类符少于不知译, 表明玛译有意识地缩短句子长度, 增加句子数量, 且降低词汇复杂程度, 以整体降低阅读难度, 这也符合玛译作为儿童作家的写作习惯。沃译平均词长少于不知译, 句子数量为不知译的 2 倍, 平均句长比不知译少 42%, 说明沃译使用大量且较短的句子以及较低的词汇丰富程度, 降低文本阅读难度。这也印证上述类符/形符比的考察发现。值得一提的是, 在文本对齐过程中, 发现存在大量原文句子无对应译文现象, 省去大量细节描写, 原文内容缺失。这也是在分析数据时不将显化列入考察对象的原因。

3.2. 词汇密度

词汇密度指的是实词数量与总词数量的百分比, 用以衡量文本信息量大小[3] [11]。英语实词有四类, 包括名词、实义动词、形容词和副词[12]。英语译语语料库的词汇密度显著低于英语源语语料库, 由此判断译者这种做法, 或有意识或无意识, 使译语更容易被读者接受。Laviosa 的文章则证实了这一推断, 并总结出译语词汇密度较低的特点, 即译文词汇的丰富性较弱[13]。

通过分析四个经过 Tree Tagger 软件词性赋码之后的译文文本, 可以计算得出顾译、沃译、不知译和玛译的词汇密度分别约为 67.6%、67.0%、68.7%和 63.6%。由此可见, 不知译的词汇密度最高, 顾译次之, 沃译居中, 玛译最低, 这也基本印证上述有关标准类符/形符比的分析结果。Laviosa 对英语译语和源语语料库的研究结果是: 译语的词汇密度为 52.87%, 源语的词汇密度为 54.95%。由此可见, 四个英译本的词汇密度均远高于 Laviosa 得出的译语词汇密度数值, 表明四个译本作为翻译文本的语言特征较源语文本相对不明显, 其中玛译最容易读懂, 沃译次之, 不知译由于其词汇的变化丰富程度高可能相对不容易读懂。

3.3. 词长分布

表 3 数据显示, 四个译本中 3 个字母的单词所占比例最大, 除了不知译(约 19%), 其他译本都在 25% 左右。从大往小按百分比排序, 依次为 3 个字母、4 个字母(16%~19%)、2 个字母(15%~17%)、5 个字母(10%~12%)和 6 个字母(8%~10%)。其中, 沃译占比第二的为 2 个字母, 与其他译本稍有不同。四个英译本中, 词长排列前五位的高频词之和除不知译版本(75.42%)外其他都占总词数的 80%以上。这表明四个英译本的叙事、口语文体特点比较明显。顾译、沃译与玛译在词汇长度的使用方面差别不大, 不知译所使用的 2、3 个字母都少于其他 3 个译本, 而使用的 5、6 个字母则略长于其他三个英译本。很明显的看到, 不知译从 7、8、9、10 个字母的使用都多于其他三个译本。由于词语长度越长, 阅读难度则越大,

这从侧面体现不知译的书面语程度较其他三个译本高，而作为叙事文体的表征则略低于其他三个译本。玛译使用 2、3、4 个字母居多，8、9、10 个字母显著少于其他三个译本，这说明玛译的单词长度选择偏好短的单词，从而降低阅读难度。这些研究发现与上述的词汇密度高度一致。

Table 3. Percentage of word length based on four translations of Nezha chapters in *Feng Shen Yan Yi*

表 3. 《封神演义》哪吒篇四译本主要词长分布百分比

	顾译	沃译	不知名	玛译
1-字母	5.18	4.80	4.00	3.07
2-字母	17.03	19.70	15.38	16.87
3-字母	24.26	25.69	19.25	26.61
4-字母	19.30	16.92	17.44	18.62
5-字母	11.79	10.33	12.79	11.89
6-字母	8.87	9.09	10.56	9.74
7-字母	5.66	5.62	7.70	7.30
8-字母	3.93	3.78	4.79	3.10
9-字母	1.90	1.99	3.50	1.52
10-字母	1.11	1.03	2.03	0.73

Table 4. High-frequency word statistics based on four translations of Nezha chapters in *Feng Shen Yan Yi*

表 4. 《封神演义》哪吒篇四译本高频词统计

次序	顾译			沃译			不知名			玛译		
	词	频次	%	词	频次	%	词	频次	%	词	频次	%
1	THE	632	5.07	THE	248	5.81	THE	183	6.00	THE	210	6.93
2	AND	425	3.41	OF	135	3.16	TO	134	4.39	AND	139	4.59
3	TO	363	2.91	TO	132	3.09	HIS	116	3.80	OF	92	3.04
4	YOU	296	2.38	HIS	131	3.07	NEZHA	72	2.36	TO	81	2.67
5	HE	223	1.79	AND	128	3.00	AND	61	2.00	HIS	76	2.51
6	A	210	1.68	A	97	2.27	OF	61	2.00	HE	69	2.28
7	I	206	1.65	HE	89	2.09	A	50	1.64	NOTSCHA	60	1.98
8	HIS	181	1.45	NO	68	1.59	IN	47	1.54	A	59	1.95
9	NEZHA	180	1.44	CHA	67	1.57	WITH	47	1.54	HIM	45	1.49
10	OF	169	1.36	HIM	62	1.45	THAT	43	1.41	IN	43	1.42
11	WAS	151	1.21	IN	56	1.31	S	39	1.28	IT	39	1.29
12	HIM	139	1.12	I	55	1.29	AS	37	1.21	DSING	30	0.99
13	IN	129	1.04	YOU	46	1.08	HE	36	1.18	HAD	30	0.99
14	LI	129	1.04	WAS	45	1.05	LI	35	1.15	LI	30	0.99
15	ON	117	0.94	THAT	41	0.96	WAS	35	1.15	WAS	30	0.99
16	JING	115	0.92	LI	39	0.91	JING	33	1.08	WITH	29	0.96

另一方面,由表4可得四个译本中使用频率最高的前五个词大致相同,都是虚词(the, and, to, of)以及人称代词(his),由此可看出四位译者在前五个人称代词的使用方面基本一致,除了个别人称代词。在1998年英语语料库(Bank of English)的统计中,英语母语的前提下频率最高的前五个词依次为the, of, to, and和a[14],而在英语翻译语料库中,前五个词依次为the, and, to, of, a[10]。由此可见,四个译本在前五个高频词的使用方面与英语语料库、英语翻译语料库相近,比如沃译与英语语料库相近,玛译与英语翻译语料库相近,其他两个译文的高频词排序与英国国家语料库的排列顺序有较大出入。此外,四个译本中词频排列前16个词中的共有人称代词(his, you, he, I),加上虚词和短语的大量使用,能够很好说明译文文本非常明显的属于小说类文体。

4. 句子层面

通过对比句子层面的平均句长以及标点符号使用情况,可以得出四个译本的语言使用特色。

4.1. 平均句长

类符/形符比、平均句长是体现译者风格的一般因素[10]。通过对比英语源语和英语译语可比语料库,得知译语的平均句长显著高于源语[13]在英汉翻译中,无论是英汉文本或是汉英文本,译文词汇数量均呈现上升趋势,这是英汉双语平行语料库不同于译文词汇数量的变化情况的地方[15][16]。从表2的平均句长来看,不知译、沃译、玛译、顾译的句长呈递减趋势,不知译的平均句长为32.74个单词,沃译、玛译、顾译分别为19.22、16.79、12.65个单词。翻译叙事文体译语文本的句子明显长于源语文本,英语译语及源语语料库平均句长的统计结果分别为24.1和15.6个单词[13]。由此,就句长而言,沃译是最接近译语的特点,而不知译超出译语语料库统计的平均句长水平,玛译与顾译则偏离最远。由于译语在某种程度上受源语影响,《封神演义》的语言属于古文言文,语句短小精悍,顾译大量使用较短的句子以贴合源语作为意合语言的特征。

4.2. 标点符号

在翻译过程中,译者风格一定程度上表现于标点符号上,而这刚好体现译者的偏好和对文本的理解。

如表5所示,四个英译本标点符号之使用情况与原文差异存在较大不同,由此可知四个译本没有被原文格式所束缚,能够将重心置于传达原文主旨。四个译本中的问号、感叹号使用频率之和分别为7.5、3.02、0、5.04,以上数据表明顾译情感在四者中更加强烈,更好揭示文中不同角色的性格特点和内心情感,相较之下,玛译的问号和感叹号使用低于顾译,同时高于沃译,这表明语调相较平稳,情感内敛含蓄。不知译为0,说明译文倾向客观描述事件,不重情感表达。由此可见,四位译者在标点符号的使用情况上体现各自译文不同的风格特点。

Table 5. Punctuation marks comparison of the four translations of Nezha chapters with the original text in *Feng Shen Yan Yi*
表5. 《封神演义》哪吒篇四译本与原文标点符号比较

标点符号	顾译		沃译		不知译		玛译		原文	
	频次	频率%	频次	频率%	频次	频率%	频次	频率%	频次	频率%
句号	811	31.20	206	21.54	93	30.39	158	31.91	624	19.15
问号	91	3.50	19	1.98	0	0	7	1.41	65	1.99
感叹号	104	4.00	10	1.04	0	0	18	3.63	107	3.28

5. 篇章层面

传统语料库软件(如 Wordsmith, AntConc)为语言库研究提供参数,而语料库的数据大体依赖于此,并且未能突破其限制,统计结果也趋向于表现译本的翻译共性[17];传统语料库存在缺陷,如语言参数单一、数据笼统等[18]。因此计量风格学也开始应用于译者风格研究中。

文本风格影响因素之一包括文本的可读性,可读性用于衡量文本阅读难度、以及文本在目标语读者中的接受度[19]。统计分析数据、归纳译者风格,从译者自身因素和非译者自身因素归因译者风格[20]。

5.1. 可读性和易读性

5.1.1. 弗莱士阅读易读性

弗莱士阅读易读性(Flesch Reading Ease Score)被广泛采用于文本易读性测量。弗莱士阅读易读性的依据是弗莱士阅读易读性公式 $RE = 206.835 - 1.015 \times ASL - 84.6 \times ASW$ 。其中,RE 为易读性,ASL 为平均句长(即单词数除以句子数),ASW 为单词平均音节数(即音节数除以单词数)。通过计算公式得出在 1 至 100 区间的评分得数,其中 100 是最高可读性分数,90~100 分代表非常容易、80~90 分代表容易、70~80 分代表相对容易,即分数越高,文本的可读性越强。见表 6。

Table 6. Flesch reading ease score

表 6. 弗莱士阅读易读性表

弗莱士阅读易读性	阅读难度
90~100	很容易(Very easy)
80~90	容易(Easy)
70~80	较容易(Fairly easy)
60~70	标准(Standard)
50~60	较难(Fairly difficult)
30~50	难(Difficult)
0~30	很难(Very difficult)

5.1.2. 弗莱士 - 金凯德年级水平

Table 7. Flesch Kincaid grade level

表 7. 弗莱士 - 金凯德年级水平表

数值	年级水平	阅读易读性
5.0~5.9	5 年级	很容易
6.0~6.9	6 年级	较容易
7.0~7.9	7 年级	容易
8.0~9.9	8~9 年级	会话英语
10.0~12.9	10~12 年级	难
13.0~15.9	大学	较难
16.0~17.9	大学毕业	很难
18.0+	专业	极难

弗莱士 - 金凯德年级水平(Flesch Kincaid Grade Level)是一种广泛使用的可读性公式,用于评估文本的大致阅读等级水平,按美国中小学年级水平评定文本的得分。分值越高表示作者的英语写作水平越高,同时对读者的英语阅读水平要求也随之提升,见表 7。其计算公式为 $39 \times ASL + 11.8 \times ASW - 15.59$,共分 12 级,数值越高,代表该文本需要接受更高受教育年限才能理解,比如当分数为 8.0 时,表示一个八年级水平的人能够理解该文本的内容。

5.1.3. 自动可读性指数

自动可读性指数(Automated Readability Index)最早应用于评价空军手册和技术性文件的可理解性,从单词难度及句子难度的比率计算得出,其计算公式为 $4.71 \times (c/w) + 0.5 \times (w/s) - 21.43$,c 代表字母数、w 代表单词数、s 代表句子数。自动可读性指数得出的数值近似于理解文本所需的年级水平,平均数值为 7,高于 7 代表难度高。见表 8。

Table 8. Automated readability index

表 8. 自动可读性指数表

年级	年纪
幼儿园	5~6 岁
1 年级	6~7 岁
2 年级	7~8 岁
3 年级	8~9 岁
4 年级	9~10 岁
5 年级	10~11 岁
6 年级	11~12 岁
7 年级	12~13 岁
8 年级	13~14 岁
9 年级	14~15 岁
10 年级	15~16 岁
11 年级	16~17 岁
12 年级	17~18 岁
大学	18~22 岁

5.1.4. Coleman-Liau 指数

Coleman-Liau 指数(Coleman-Liau Index)起初用于评估美国公立学校使用的教科书的可读性,计算过程不关注单词的音节数,而是单词长度。其计算公式为 $CLI = 0.0588L - 0.296S - 15.8$,其中,L 为每 100 个单词的平均字母数,S 为每 100 个单词的平均句子数。得出数值预估理解文本的年级水平。

5.1.5. 冈宁灰雾指数

冈宁灰雾指数(Gunning Fog Index),或迷雾指数,主要揭示看懂文本内容所需要的受教育年限。其公式为 $FOG = 0.4 (ASL + PHW)$, $ASL = \text{平均句长(即单词数} \div \text{句子数)}$, $PHW = \text{长词百分比(即长词} \div \text{单词数} \times 100)$,长词是 3 个及以上音节的单词。得出数值参考受教育年限表。

5.1.6. SMOG 指数

SMOG 指数(SMOG Index)原用于测算杂志文本的可读性,公式: $\sqrt{NOP} + 3$,计算过程选取 30 及以

上的句子，计算音节数多余 3 的单词总数开根，得出数字再加 3。计算结果与理解文本需要的受教育年限成正相关关系。

Table 9. Readability comparison of four translations of Nezha chapters in *Feng Shen Yan Yi*
表 9. 《封神演义》哪吒篇四译本各项易读性比较

对比项	顾译	沃译	不知译	玛译
弗莱士阅读易读性	77.48	73.47	40.95	73.92
自动可读性指数	5.36	7.25	17.15	6.94
弗莱士 - 金凯德年级水平	5.43	7.1	15.39	6.99
Coleman-Liau 指数	7.71	7.96	11.75	7.66
冈宁灰雾指数	15.01	18.68	31.43	15.55
SMOG 指数	12.59	15.22	24.1	13.5

通过北外 Huge Mind Readability Analyzer 的数据分析易读性，结果由表 9 所示，弗莱士阅读易读性方面，顾译的易读性最高，玛译次之，沃译居中，不知译最低，前三者译文都处于较容易阅读一级，而不知译译文处在困难一级，阅读难度情况与考察词汇层面时结果一致。弗莱士 - 金凯德年级水平方面，顾译的阅读难度在 5 年级水平，玛译在 6 年级水平，沃译在 7 年级水平，不知译在大学水平，不知译难度高于其他三个译本。自动可读性指数的结果与金凯德年级水平相同，其岁数分别对应 10~11 岁、11~12 岁、12~13 岁和 18~22 岁。Coleman-Liau 指数结果显示与前面不同，四个译文所需教育年限除不知译外都所有提升(7 年级水平)，而不知译文本下降至 11 年级水平。冈宁灰雾指数结果与理想结果 7 相去甚远，顾译与玛译约为 15，沃译约为 18，不知译为 31，说明，不知译的阅读难度是前两者的 2 倍，难度骤然提升，所需教育水平也随之提升。SMOG 指数表明顾译与玛译文本阅读难度相差不大，沃译略难，不知译难度大，年级分别是 6 年级、7 年级、8 年级和 7 年级。

6. 结论

通过语料库数据对比分析，得出，《封神演义》哪吒篇四个英译本在标准类符/形符比、词汇密度、高频词分布的考察呈现高度一致的结果，但作为译语的译语特征不明显。但是，由于现实中各位译者本身从事职业、自身所处当时的社会背景、个人翻译的时间长度、目的各异所导致四个英译本呈现各自独特的风格特点。顾译文本内容相较其他三个译本较全面，使用句子数量最多，且较多属于一对一对应翻译，受源语影响较大。玛译词汇密度最低，类符最低，句子长度最短，句子数量居中，说明玛译为贴近儿童受众群体，有意识降低文本阅读难度。不知译的句子数量最少，平均词长最长，平均句长最长，标准类符/形符比最高，说明该译本用词变化程度高，文本阅读难度也随之提高，对读者受众有较高要求。沃译句子数量是不知译的 2 倍，比玛译略多，标准类符/形符比比顾译高而低于其他两个译本，平均词长最短，说明沃译在选词上、用词变化丰富程度上较低与除玛译之外的其他两个译文。本文分析以及讨论的四个英译本的语内特征维度计较有限，尚需更全面的分析角度展现源语以及译语的语料库全貌，比如语境、文化层面。通过不同的易读性公式计算得出，顾译的易读性最强，玛译次之，沃译居中，不知译最弱，与词汇层面相结合，玛译最具特色。

参考文献

- [1] 裴永刚, 索煜祺. 中华优秀传统文化走出去的创新路径探析——以数字出版中游戏出海为视角[J]. 出版广角, 2022(15): 47-52.

- [2] 胡开宝. 语料库翻译学[M]. 上海: 上海交通大学出版社, 2017.
- [3] Baker, M. (1995) Corpora in Translation Studies: An Overview and Some Suggestions for Future Research. *Target*, No. 2, 223-243. <https://doi.org/10.1075/target.7.2.03bak>
- [4] Hermans, T. (1996) The Translator's Voice in Translated Narrative. *Target*, No. 1, 23-48. <https://doi.org/10.1075/target.8.1.03her>
- [5] 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(4): 75-82.
- [6] 张美芳. 利用语料库调查译者的文体——贝克研究新法评介[J]. 解放军外国语学院学报, 2002(3): 54-57.
- [7] Baker, M. (2000) Towards a Methodology for Investigating the Style of a Literary Translator. *Target*, No. 2, 241-266. <https://doi.org/10.1075/target.12.2.04bak>
- [8] Olohan, M. (2003) How Frequent Are the Contractions. *Target*, No. 1, 59-89. <https://doi.org/10.1075/target.15.1.04olo>
- [9] Yu, C.-H. (2009) Similarity and Difference in Translator's Style: A Case Study of the Two Translations of Hemingway's Work. In: Hu, K.B. (Ed.), *Translation and Interdisciplinary Studies*, Foreign Language Teaching and Research Press, Beijing, 191-201.
- [10] Olohan, M. (2004) *Introducing Corpora in Translation Studies*. Routledge, New York. <https://doi.org/10.4324/9780203640005>
- [11] Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. Pearson Education Limited, London.
- [12] 朱珊. 《狂人日记》译者风格: 一项基于语料库的研究[J]. 外国语文, 2021(5): 119-128.
- [13] Laviosa, S. (1998) Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. *Meta*, 43, 557-570. <https://doi.org/10.7202/003425ar>
- [14] Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9781139524773>
- [15] 王克非. 英汉/汉英语句对应的语料库考察[J]. 外语教学与研究, 2003(6): 410-416.
- [16] 王克非, 秦洪武. 英译汉语言特征探讨——基于对应语料库的宏观分析[J]. 外语学刊, 2009(1): 102-105.
- [17] 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.
- [18] 许明, 蒋跃. 《阿Q正传》译入译出文本的风格计量学对比[J]. 外语研究, 2020(3): 86-103.
- [19] 仲文明, 王靖涵. 少年儿童翻译文学的译本风格计量研究——以 *Silent Spring* 三译本为例[J]. 外语翻译, 2023(1): 20-27.
- [20] 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(2): 12-18.