



# 基于语料库的人机译本语言计量特征对比研究

——以 *Peter Pan* 四译本为例

袁艳玲, 肖一美, 黄晶晶

(南华大学语言文学学院, 湖南衡阳 421001)

**摘要:** 机器翻译如今深刻影响着翻译的各个领域。为观察文学领域中人工译本与机器译本的语言特征差异, 基于自建 *Peter Pan* 英文原文及其两个人工译本、两个机器译本的英汉平行语料库, 运用 Python、CRIE、Emeditor 和 ABBYY Aligner 等工具, 从词法与句法两个层面对人机译本的 17 个语言计量特征进行对比研究。研究结果表明: 在文学作品翻译中, 机器翻译尚不如人工翻译, 机器翻译处理词汇与句子不灵活, 主要表现为代词“显化”明显和行文缺乏逻辑。两个人工译本相比, 任溶溶译本语言富有韵律, 阅读难度较低; 杨静远译本语言丰富, 流畅自然。两个机器译本相比, 谷歌译本较有道译本词汇表达更多样; 有道译本更贴近中文表达。

**关键词:** 语料库; 语言计量特征; 人工译本; 机器译本; 对比

**中图分类号:** H315.9

**文献标志码:** A

**文章编号:** 1673-3851 (2024) 06-0299-07

## Research on the linguistic features of human translation and machine translation based on corpus: Taking *Peter Pan*'s four translations as an example

YUAN Yanling, XIAO Yimei, HUANG Jingjing

(School of Languages and Literature, University of South China, Hengyang 421001, China)

**Abstract:** Machine translation has a profound impact on various fields of translation today. To observe the differences in linguistic features between human translation and machine translation in the field of literature, the authors built *Peter Pan*'s English-Chinese parallel corpus of the original English text and its two human translations and two machine translations. Tools such as Python, CRIE, Emeditor and ABBYY Aligner were used to conduct comparative research on 17 linguistic structures of the four translations from the lexical and syntactic levels. The research results show that in the field of literature, machine translation is not as good as human translation, with the former being less flexible than the latter in processing words and sentences, mainly inflected in features of evident "explicitation" of pronouns, and lack of logic in writing. It is found from the comparison of the two human translations that Ren Rongrong's version is full of rhythm and easy to read, while Yang Jingyuan's version is rich in language and fluent in text. It is found from the comparison of the two machine translations that the Google version has more diverse vocabulary expressions than the Youdao version, while the Youdao version is closer to Chinese expressions.

**Key words:** corpus; quantitative characteristics; human translation; machine translation; comparison

机器翻译即计算机翻译,指利用计算机将一种语言符号转为另外一种语言符号<sup>[1]</sup>。目前,人机翻译对比研究为一大热点,学者们从不同层面展开人机翻译对比。如,梁君英等<sup>[2]</sup>从词汇、句法、篇章等多维度系统分析了人工翻译与机器翻译的异同,总结了人类智能翻译能力的三大优势。李奉栖<sup>[3]</sup>运用错误积分法,从五个维度对比研究英语专业翻译学习者与神经机器翻译的英汉翻译质量,研究结果表明机器更擅长处理术语翻译,人工译者则更擅长处理文化差异问题。Sheng等<sup>[4]</sup>通过对比《劝学》和《论构建人类命运共同体》的人工译本和机器译本,发现人工译者能够更加巧妙地处理政治和意识形态问题,而神经机器翻译系统缺乏灵活性和主观性。

然而,将语料库与计量语言学结合,对比文学领域中机器翻译与人工翻译的相关文献不多,且现有少量研究仅仅将国内的“百度翻译”的译本与人工译本作比较<sup>[5-7]</sup>,并未体现国外机器翻译水平。为了更客观地展现人工译本与机器译本的语言特征差异,本研究结合语料库翻译研究和计量语言学方法,选取儿童文学作品 *Peter Pan* 原文及其四译本即任溶溶、杨静远两个人工译本以及有道翻译(中国)、谷歌翻译(美国)的两个机器译本为研究文本,从词法与句法两个层面、17个语言计量单位对人工及机器译本语言特征的差异进行比较研究。本研究旨在回答以下两个问题:一是,人工译本与机器译本分别有哪些语言特征?二是,人工译本与机器译本的语言特征有哪些差异?对比人工译者与机器翻译的语言特征差异,可发现机器翻译存在的问题与局限性,这有助于指导机器翻译系统的改进,提高翻译质量,优化翻译流程。

## 一、研究设计

### (一)语料选取

本研究以儿童文学作品 *Peter Pan* 及其四译本为语料,建立一对四的英汉双语平行语料库。其中, *Peter Pan* 的英文文本采用苏格兰小说家及剧作家詹姆斯·马修·巴利(James Matthew Barrier)所作、2018年由 Wordsworth 出版社出版的 *Peter Pan*; 两个人工译本为:任溶溶翻译、2017年由少年儿童出版社出版的《小飞侠彼得·潘》(以下简称“任译本”)和杨静远翻译、2020年由南京大学出版社出版的《彼得·潘》(以下简称“杨译本”);两个机器译本为:有道翻译自动生成译本(以下简称“有道译本”)

和谷歌在线翻译系统自动生成译本(以下简称“谷歌译本”)。根据各类语料库检索工具对英文原著以及四个译本进行统计,任译本、杨译本、有道译本和谷歌译本的字数分别为 75812 字、73844 字、69492 字和 68608 字。四个译本的字数相差不大,具有可比性,符合语料选择的标准。

### (二)语言特征选定

本研究所考察的对象为词汇与句子层面的语言特征。基于以往研究,本研究总结归纳了人工译本与机器译本对比研究常用的语言计量单位。由于 *Peter Pan* 属于文学领域中的儿童文学,叠词在儿童文学汉译中的灵活运用,可增强源语故事的感染力,提升儿童汉语语言文化认知和故事审美体验,是儿童文学汉译语言研究不可或缺的一部分<sup>[8]</sup>,因此本研究也考察叠词的使用频率。最终确定 17 个语言计量单位,用于比较分析人工译本与机器译本的语言特征差异。其中,词汇层面共 13 个语言计量单位,分别为:标准类符/形符比、词汇密度、高频词、叠词、名词比例、动词比例、形容词比例、副词比例、数词比例、量词比例、连词比例、代词比例、助词比例;句法层面共 4 个语言计量单位,分别为:句长、陈述句比例、疑问句比例和感叹句比例。

### (三)统计工具与统计方法

本研究采用 Python、中文可读性指标自动化分析系统(Chinese readability index explorer, CRIE)、Emeditor 和 ABBYY Aligner 四个工具对上述 17 个语言计量单位进行数据统计。对词汇层面的数据首先运用 Python 进行分词和词性标注,统计类符/形符比(Type token ratio, TTR)、标准化类符/形符比(Standardized type token ratio, STTR)、词长分布、词汇密度、高频词等数据;再通过 CRIE 获取名词、动词、形容词、代词等数量,各类词数量与总词数相比即可得出各词汇比例。对叠词的统计稍有不同,先利用 Python 进行分词之后,通过 Emeditor 输入正则表达式提取检索目标,再进行人工筛选,最终统计出各类叠词具体数量。而对句子层面的数据则先运用 ABBYY Aligner 软件将四个译本与英文原文对齐,再利用 CRIE 计算出句子数目、字数和平均句长,最终通过 Python 编程实现对标点符号的识别及符号间单词数量的统计,得出陈述句、疑问句与感叹句的句子数目,将各类句式数目与总句子数目相比便可得出各类句式比例。

二、词汇层面的考察

基于各类工具所获取的数据,本研究将从标准类符/形符比、词汇密度、实词及虚词比例、高频词、叠词等数据对比四个英译本的语言计量特征,并探讨不同译本之间的语言计量特征产生差异的原因。

(一)词汇丰富度

Baker<sup>[9]</sup>提出,类符(type)指文本中不同的词,排除重复并忽略大小写;形符(token)指文本中所有出现的词。文本中类符/形符比值越高,说明作者的词汇使用越丰富、用词越富于变化<sup>[10]</sup>。标准类符形符比是按照一定长度(如 1000 形符)切分文本,再逐个计算各切分文本的类符/形符比,最后取其平均值。STTR 常用于比较不同长度的文本,因为类符聚集的均匀程度不同会影响类符/形符比值,所以采取标准化的类符/形符比值描述词汇丰富度更可靠<sup>[10]</sup>。有关词汇密度的统计数据如表 1 所示。

表 1 四个译本的词汇丰富度统计数据

译本	语言特征				
	形符/个	类符/个	TTR/%	STTR/%	字数/个
任译本	58856	6376	10.83	33.09	75812
杨译本	57826	7402	12.80	35.68	73844
谷歌译本	54515	6049	11.09	34.72	68608
有道译本	55727	5979	10.73	33.03	69492

从表 1 来看,杨译本的 STTR 高于两个机器译本,而任译本的 STTR 与有道译本持平但低于谷歌译本,即杨译本(35.68) > 谷歌译本(34.72) > 任译本(33.09) > 有道译本(33.03)。该结果说明,人工译本与机器译本在词汇丰富度方面差异不大,二者均无明显优势。程勇等<sup>[11]</sup>指出词汇丰富度越高的文章,阅读难度越高,而词汇越单一、丰富度越低的文章,其阅读难度也就越低。杨译本的 STTR 稍高于其他三个译本,说明在相同字数的情况下,杨译本的词汇使用更为多样,其主要原因是人工译本的语言特征会因不同译者的翻译目的及主体性发挥而不同。细读杨译本,不难发现杨静远在翻译过程中运用了一系列翻译技巧,力求译文语言活泼生动且优美动人。为达到此翻译目的,杨译本有意采用丰富多样的词汇将故事呈现给孩子们,故而其译文词汇丰富度会高于其他译文。而任溶溶在翻译儿童文学作品时有其个性化的翻译原则,即忠于原文、口语化、情境化等。任译本的字数最多,但其类符远低于杨译本,这是由于任在翻译过程中充分考虑到儿童读者群体,秉持忠于原文为本的原则,有意识地减少

类符数量,以降低文本的阅读难度。神经机器翻译依赖数据驱动,其翻译结果受到语料库大小和质量的影响<sup>[12]</sup>。谷歌译本较有道译本的词汇丰富度高,其词汇丰富度超过了任溶溶译本,表明两个神经机器翻译系统之间存在差别,谷歌翻译系统所具备的语料库庞大,在挖掘中英双语语料与词汇丰富度方面有优势,其词汇选取与运用接近人工翻译。

(二)词汇密度

词汇密度有两种计算方法,一种是直接以类符/形符比(TTR)作为词汇密度,另一种则是统计实词形符在总形符数中的比例。由于语料库容量和选材会对类符/形符比(TTR)有影响,不能真实反应词语变化度,因此以类符/形符比作为词汇密度考察会有一定的误差,故本研究选择第二种方法。

经计算,任译本、杨译本、有道译本和谷歌译本的词汇密度分别是 0.79、0.81、0.77 和 0.77,说明两个人工译本的词汇密度均高于两个机器译本。王克非<sup>[13]</sup>提出了词汇密度低反映信息量和文本难易度相对降低。机器译本词汇密度低于人工译本,即在相同长度的句子中机器译本所运用的实词少于人工译本,其所承载的信息量少于人工译本,阅读难度也低于后者。谷歌译本与有道译本的词汇密度完全相同,说明谷歌翻译系统与有道翻译系统对于词汇的处理差别不大。人工译者对文本的处理方式因翻译原则不同而有差异,故而二者的词汇密度也会有所区别。任译本的词汇密度低于杨译本,即译本阅读难度低于杨译本,这与词汇丰富度数据分析所得结论相同。

(三)词频

为了进一步考察词汇在各译本中的使用差异,本研究将词频考察分为两点:一是各类词的词频差异,二是高频词。运用语料库工具共统计了名词、动词、形容词、副词、数词、量词、连词、代词、助词的频次比例,以及各译文前十位高频词。以下将从不同词性的词频差异与高频词差异,分析不同译本的语言特征。

1. 使用各类词的词频差异

四个译本使用各类词的比例见表 2。以下将对名词、动词、副词、连词和代词比例进行分析,形容词、数词、量词和助词比例差距不大,故不作分析。

从表 2 可见,就名词而言,任译本与谷歌译本的名词比例持平,而杨译本与有道译本的名词比例一致,这表明人工译本与机器译本在名词的使用频率上差异不大,机器翻译能准确识别和处理名词。就



动词而言,其使用频率远高于其他词性,且两个人工译本的动词比例均高于机器译本。汉语倾向于多采用动词,人工译者在翻译各类词时不仅会将动词译为动词,还会将其他词类转化为动词,故而人工译本

的动词比例会高于机器译本。这也表明机器翻译在词性转化功能上还有待进一步优化,目前尚无法贴近地道的汉语表达。

表 2 四个译本使用各类词汇的词频统计数据

译本	名词 比例/%	动词 比例/%	形容词 比例/%	副词 比例/%	数词 比例/%	量词 比例/%	连词 比例/%	代词 比例/%	助词 比例/%
任译本	17	27	5	9	3	3	2	16	10
杨译本	18	28	5	10	3	3	2	14	10
谷歌译本	17	25	5	6	3	2	3	15	10
有道译本	18	26	5	7	3	2	3	16	11

人工译本与机器译本各词汇使用比例差异较大的还有副词。英语中的副词很丰富,使用频率较高。在汉语中,副词既可以修饰动词、形容词,也可以修饰整个句子,作状语和补语。人工译本的副词比例远高于机器译本,说明机器翻译对副词的识别和处理能力不够,使得机器译本的篇章连贯性不强。而从连词的角度来看,连词是体现语义逻辑关系的重要形式,使语篇成为连贯的有机整体。英语中的连词比汉语中的连词种类数量多,使用频率也高很多。机器译本的连词比例高于人工译本,可见其篇章逻辑性与连贯性都不如后者,进一步说明机器翻译对文本中词汇处理的灵活性不如人工译者。

英语中代词使用频繁,而汉语则与之相反。受到了英语源语影响,汉语译文代词使用频率会增加。王克非等<sup>[14]</sup>认为,与原创汉语相比,汉语译文具有虚词与指代方式“显化”等特征。机器译本和任译本

的代词比例均高于杨译本,表明杨译本采用其他形式取代英文中反复出现的代词,使得译文通顺流畅。任译本的代词比例与有道译本相同,可能是任溶溶秉持以忠实取向为本的原则,没有优化处理译文代词。而机器翻译可能无法运用其他方法来替换原文中的代词,将源语中的代词一一译出,造成译入语代词“显化”明显。

2. 高频词

高频词是指在一个文本里使用频率特别高的词语,通常通过一个词语在整个文本中所占的百分比或者该词语在整个词频中所占的前后位置来判断<sup>[15]</sup>。本研究运用 Python 统计四个译本的词汇频率,以下列举了前十位的高频词,各高频词的占比可在一定程度上反映文本的独特性。四个译本使用排名前十位的高频词统计如表 3 所示。

表 3 四个译本使用排名前十位的高频词

排名	任译本			杨译本			谷歌译本			有道译本		
	高频词	频次/次	占比/%	高频词	频次/次	占比/%	高频词	频次/次	占比/%	高频词	频次/次	占比/%
1	的	2718	3.59	的	2590	3.51	的	2770	4.04	的	2992	4.31
2	他	1642	2.17	了	1465	1.98	他	1617	2.36	他	1599	2.30
3	了	1241	1.64	他	1330	1.80	了	1231	1.79	了	1320	1.90
4	她	893	1.18	是	799	1.08	她	957	1.39	她	964	1.39
5	他们	796	1.05	她	777	1.05	在	831	1.21	在	878	1.26
6	在	776	1.02	他们	671	0.91	他们	829	1.21	是	818	1.18
7	是	736	0.97	在	670	0.91	是	799	1.16	他们	790	1.14
8	我	618	0.82	彼得	611	0.83	我	584	0.85	我	581	0.84
9	说	530	0.70	我	549	0.74	你	496	0.72	你	502	0.72
10	你	485	0.64	说	514	0.70	彼得	393	0.57	说	451	0.65

四个译本的前十位高频词中出现频率较高的主要是代词(他、她、他们、我、你)。任译本、杨译本、有道译本和谷歌译本使用的前十位高频词中,代词累计出现频次比例分别为 5.86%、4.50%、6.53%和 6.39%,要特别指出的是机器译本的代词比例高出

杨译本两个百分比。这进一步表明,机器翻译的代词“显化”尤为突出。具体可见例 1:

例 1 英文原文:She flew away with Peter in the frock **she** had woven from leaves and berries in the Neverland, and **her** one fear was

that **he** might notice how short it had become; but **he** never noticed, **he** had so much to say about **himself**.<sup>[16]190</sup>

**任译本:**她穿着在梦幻岛的时候用树叶编织镶浆果的斗篷和彼得一起飞走,她怕彼得会注意到斗篷变得那么短,可是他根本没有注意,关于他本人,他要说的事情太多了。<sup>[17]167</sup>

**杨译本:**温迪和彼得一起飞走时,身上穿着她在永无岛时,用树叶和浆果编织成的罩褂,她生怕彼得看出这罩褂已经变得多么短了;可是彼得根本没注意,他自己的事,他还说不完呢。<sup>[18]158</sup>

**谷歌译本:**她和彼得一起飞走了,穿着她在梦幻岛用树叶和浆果编织的连衣裙,她唯一的担心是他可能会注意到它变得有多短;但他从来没有注意到,他对自己有很多话要说。

**有道译本:**她穿着在永无岛用树叶和浆果编织的连衣裙,和彼得一起飞走了。她唯一担心的是,他会注意到这件连衣裙已经变得很短了。但是他从来没有注意到,他有那么多关于他自己的话要说。

从例 1 可见,杨译本用人名代替了英语原文中频繁出现的人称代词,使得译文更加符合中文的阅读习惯。而谷歌和有道翻译无法根据上下文语境来表达逻辑关系以省略不必要的表达,在译文中显化代词,完全对应于英文原文,有些拗口。而且对于“he had so much to say about himself”的翻译,谷歌翻译给出的译文为错译,有道译本对这句话的处理便优于前者。各词汇比例与高频词分析可以交叉印证,机器翻译无法识别上下文语义,无法根据语境准确完成双语转换,译文相较人工译文较为生硬拗口。

(四)叠 词

*Peter Pan* 是一本儿童读物,考虑到读者因素,译本风格应遵循原文风格,行文生动,富有童趣,从而吸引广大儿童。叠词是汉语中一种特殊的词汇现象,表现为两种形式:一是由两个相同音节构成的叠音词,二是由两个同词根重叠组成的重叠词<sup>[19]</sup>。根据叠词的表现形式,本研究共统计了 12 类叠词,详见表 4。

由表 4 可见,机器译本所运用的叠词数量大致为人工译本的一半。可见机器翻译不善于将词译为叠词这类特殊词汇,即其在该类词汇的处理上尚不如人工译者。从任译本和杨译本来看,任译本所运

用的叠词数更多,该译本应更具趣味性,更富感染力,可体现出任溶溶注重读者接受,译文富有童趣。两个机器译本之间的叠词数量也有差异,这可能是由于翻译系统中双语语料的不同。具体可见例 2 和例 3。

表 4 四个译本使用各类叠词的频次					次
叠词类型	任译本	杨译本	谷歌译本	有道译本	
AA	73	50	27	32	
AAA	2	0	1	0	
AAB	85	60	36	46	
ABA	5	8	8	3	
ABB	32	47	13	15	
AAAA	1	0	0	0	
AABB	60	65	22	27	
AABC	25	32	32	20	
ABAB	17	8	4	4	
ABAC	50	51	28	43	
ABCB	8	6	1	1	
ABCC	11	15	5	3	
合计	369	342	177	194	

**例 2 英文原文:**One day when she was two years old she was playing in a garden, and she plucked another flower and ran with it to her mother.<sup>[16]3</sup>

**任译本:**她两岁的时候,有一天在花园里玩,摘了一朵花,拿着它噔噔噔跑到妈妈那里。<sup>[17]5</sup>

**杨译本:**她两岁的时候,有一天在花园里玩,她摘了一朵花,拿在手里,朝妈妈跑去。<sup>[18]14</sup>

**谷歌译本:**两岁时的一天,她在花园里玩耍,摘下另一朵花,跑到妈妈那里。

**有道译本:**温迪两岁的时候,有一天她在花园里玩,她摘了另一朵花,带着它跑向妈妈。

**例 3 英文原文:**They could hear Nana barking, and John whimpered, "It is because he is chaining her up in the yard", but Windy was wiser.<sup>[16]23</sup>

**任译本:**他们听到南娜在汪汪叫,约翰抽抽嗒嗒地说:“都因为爸爸把它用铁链拴在院子里了。”可是温迪更聪明。<sup>[17]23</sup>

**杨译本:**他们听得见娜娜的吠声,约翰呜咽着说:“这都是因为他把她拴在院子里了。”可是温迪知道得更多。<sup>[18]22</sup>

**谷歌译本:**他们能听到娜娜在叫,约翰呜咽着说:“那是因为他把她锁在院子里了,”但

温迪更聪明。

**有道译本：**他们听见娜娜在叫，约翰呜咽着说：“那是因为他把她拴在院子里了。”但温迪更聪明。

例 2 中，任溶溶将“ran”译为“噎噎噎跑”，其他三个译本均译为“跑”。“噎噎噎”生动形象地表现出两岁的温迪想将花朵送给妈妈的喜悦与迫切心情，整个句子充满了画面感。例 3 中，“barking”和“whimpered”的译文不太相同，任溶溶将二者分别译为“汪汪叫”和“抽抽嗒嗒”，杨静远则分别译为“吠声”和“呜咽”，谷歌和有道译文均为“叫”和“呜咽”。“汪汪叫”和“抽抽嗒嗒”两个叠词均摹声，使得原文描述的狗狗“娜娜”和约翰的形象更加生动，读起来也朗朗上口。从例 2 和例 3 来看，任译本多处运用叠词，译本富有童趣，在韵律、形式和语义上更具有美感。由于机器翻译往往选取最高频的译法，故谷歌和有道的译文一致，而人工译者会根据不同的语境选择恰当的词汇，因此两个人工译者的译文也有不同之处。

三、句法层面的考察

本研究将句子标志设为问号、句号、感叹号和省略号，统计了英文原文与四个译本的句子数和平均句长，并统计了陈述句、疑问句和感叹句的使用比例。通过对平均句长与句子特征的考察，可对比分析出四个译本在句子层面的语言特征。

(一)平均句长

小说原文与四个译本的句子特征统计结果如表 5 所示。可见，两个机器译本的平均句长略低于原文句长但基本一致，而两个人工译本的平均句长则差距明显。胡开宝等<sup>[1]</sup>指出机器翻译的一大特征就是以句子为翻译单位。两个机器译本平均句长相同，可见机器翻译对句子本身改动不大，在句子层面操纵较少。比较两个机器翻译的翻译方法，谷歌翻译偏向于直译，而有道翻译可对句子进行拆译和合译，从而导致谷歌译本与有道译本的平均句长尚有细微差别。任译本句子数量高于原文，平均句长低于原文，这可能是因为任溶溶考虑到读者受众为儿童以及他们文学接受能力的特殊性，故而减少句子所载的信息量，降低阅读难度。杨静远追求译文与原文形式上的对等，句子数目与原文相差不大，但其平均句长略高于任译本。这表明句子层面分析所得结论与以上词汇运用分析结论相同，杨译本的阅读难度稍高于任译本。综合而言，机器翻译拆分与合

并处理句子能力尚不如人工译者，无法像人工译者那样根据原文的含义对句子作适当的改动。

(二)句子特征

从表 5 可见，小说原文的陈述句、疑问句和感叹句比例分别为 0.90、0.06 和 0.04，谷歌译本与有道译本的比例与原文各类句式比例完全相同，这凸显出机器翻译可准确识别出标点符号，按照原文句对句的翻译。任译本的陈述句均高于小说原文和杨译本，疑问句数量则低于杨译本，而感叹句少于小说原文。可见，任溶溶灵活使用标点符号，使得陈述句数量多于原文，而感叹句数量少于原文。而杨译本陈述句比例与原文相等，疑问句与感叹句比例与原文有点偏差，杨静远在翻译句子时仍会适当改变句子标点符号。总之，相较于两种机器译本而言，任和杨的译本对原文句型的依赖较少，善于灵活转换句型。这进一步印证了机器翻译无法识别和处理原文的语义与情感，无法根据原文含义创造性地改动文本句子，搭建更为连贯的篇章，故而机器翻译对文本句子处理的灵活性不如人工译者。

表 5 句长和各类句子特征统计

考察项目	小说原文	任译本	杨译本	谷歌译本	有道译本
句子数目/个	5536	5951	5643	5485	5660
平均句长/词	8.57	8.49	8.80	8.33	8.30
陈述句/个	4986	5440	5065	4936	5110
疑问句/个	318	325	367	318	320
感叹句/个	232	186	211	231	230
陈述句比例/%	90	91	90	90	90
疑问句比例/%	6	6	7	6	6
感叹句比例/%	4	3	3	4	4

四、结 语

本研究以儿童文学作品 *Peter Pan* 及其四译本为语料建立一对四的英汉双语平行语料库，从语料库语言学及计量语言学的角度，运用 Python、CRIE 和 Emeditor 等统计工具从 17 个语言计量单位对人工译本与机器译本的语言特征进行了对比研究。本研究发现：从词汇方面来看，机器译本较人工译本在词汇丰富度上没有明显优势；机器翻译词性识别精确，译文有忠实于原文的特点，但是机器翻译存在词性转化功能不佳，导致译本中代词“显化”明显、译文不贴合汉语表达习惯等问题。而人工译者根据自身翻译目的并发挥主体性，可使译本在词汇丰富度上不逊色于机器译本，甚至可能略胜一筹；同时人工译者可巧妙转化动词与代词等，善于翻译叠词这类特

殊词汇,从而使译文更贴近译入语表达习惯、满足读者需求。从句子层面来看,机器翻译可准确识别标点符号,但其以句子为翻译单位,缺乏对表达内容进行加工整合的能力,故机器译本有行文不流畅、逻辑性不强等缺点。人工译者可恰当地切分或整合句子并灵活使用标点符号,因而其译本更加自然流畅、更具逻辑性。两个人工译本以及两个机器译本内部之间也有差别。人工译者由于自身的经历与背景不同,其译本会呈现出不同的风格。有道和谷歌翻译虽同属于神经机器翻译,但由于记忆库与术语库等方面不同,二者对文本的处理仍存在着差异,各有特色。

文学作品翻译不是将源语言转化为目标语言的简单工作,其应具备艺术性与创造性。即使机器翻译发展到以神经网络为基础,译文质量得到了一定程度的提升,但其仍面临语义理解、情感识别和信息抽取等方面的困难,机器翻译应针对这几个问题加以改进。因此,现阶段的机器翻译在文学领域仍难以超越人工翻译。当然,本研究所建的语料库规模不大而且仅从语言特征层面分析比较四个译本。为了更加全面地对比人工译本与机器译本的异同,后续需持续深入地开展相关研究。

参考文献:

[1] 胡开宝,李翼. 机器翻译特征及其与人工翻译关系的研究[J]. 中国翻译, 2016, 37(5):10-15.

[2] 梁君英,刘益光. 人类智能的翻译能力优势:基于语料库的人机翻译对比研究[J]. 外语与外语教学,2023(3):74-84.

[3] 李奉栖. 人工智能时代人机英汉翻译质量对比研究[J]. 外语界, 2022, 211(4): 72-79.

[4] Sheng A, Kong Y. Neural machine translation and human translation: A political and ideological perspective [J]. Babel, 2023, 69(4): 483-498.

[5] 韩红建,蒋跃. 基于语料库的人机文学译本语言特征对

比研究:以《傲慢与偏见》三个译本为例[J]. 外语教学, 2016,37(5):102-106.

[6] 蒋跃,张英贤,韩纪建. 英语被动句人机翻译语言计量特征对比:以《傲慢与偏见》译本为例[J]. 外语电化教学,2016(3):46-51.

[7] 蒋跃. 人工译本与机器在线译本的语言计量特征对比:以 5 届韩素音翻译竞赛英译汉人工译本和在线译本为例[J]. 外语教学, 2014, 35(5): 98-102.

[8] 韩洋. 基于语料库的儿童文学汉译中叠词的应用研究:以李文俊的儿童文学翻译为例[J]. 外语电化教学, 2019(3): 15-21.

[9] Baker M. Corpora in translation studies: An overview and some suggestions for future research[J]. Target, 1995,7(2): 223-245.

[10] Baker M. Towards a methodology for investigating the style of a literary translator[J]. Target, 2000(2):241-266.

[11] 程勇,徐德宽,董军. 基于语文教材语料库的文本阅读难度分级关键因素分析与易读性公式研究[J]. 语言文字应用, 2020(1): 132-143.

[12] 章钧津,田永红,宋哲煜,等. 神经机器翻译综述[J]. 计算机工程与应用,2024,60(4):57-74.

[13] 王克非. 语料库翻译学探索[M]. 上海:上海交通大学出版社, 2012: 81.

[14] 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.

[15] 冯庆华. 思维模式下的译文与词汇[M]. 上海:上海外语教育出版社, 2012: 5.

[16] James Matthew Barrie. *Peter Pan* [M]. Hertfordshire: Wordsworth Edition Ltd,2018.

[17] 巴里. 小飞侠彼得·潘[M]. 任溶溶,译. 北京:少年儿童出版社, 2017.

[18] 巴里. 小飞侠彼得·潘[M]. 杨静远,译. 北京:中国画报出版社, 2013.

[19] 周莉,刘娅,崔家平.《黄帝内经》叠词基本特征及英译研究[J]. 时珍国医国药, 2021, 32(8): 1958-1960.

(责任编辑:陈丽琼)