

# 基于计量语言学的《汤姆·索亚历险记》 两译本语言特色研究

程然

辽宁师范大学

**摘要:**本研究对比《汤姆·索亚历险记》两个不同译本,自建平行语料库,运用语料库方法从计量语言学的角度探究译本的语言特色在面向译本不同受众人群时所呈现的差异。经过研究发现:两个译本都比较通俗易懂。刘桂兰的译本词汇更为简单,朱建迅、郑康译本的TTR指数较高,词汇较为丰富多样;朱郑本较刘本有更突出的显化倾向,对人称代词的使用较多,对话较多。

**关键词:** 计量语言学;《汤姆·索亚历险记》;语料库;词汇

【DOI】10.12252/j.issn.2096-6288.2022.09.126

## 一、引言

Hermans (1996: 27) 认为,译本中隐藏着“译者的声音”。传统的译本研究多基于对例句的分析,挑选例句进行研究,主要依赖研究者个人的直觉感受。较窄的考察范围和过多人感性因素的介入在一定程度上影响了研究深度和结论的可行性。<sup>[1]</sup>Baker (2000: 245) 提出建议将语料库语言学与翻译学相关的理论和方法结合起来,从而去探究翻译风格。<sup>[2]</sup>黄伟、刘海涛 (2009) 提出将语言计量研究成果应用于文本聚类研究的方法,以语料库和统计方法进行语体特征计量研究。<sup>[3]</sup>本文从译文面向的受众人群不同来选择译本,选取译文出版社2010年出版,朱建迅、郑康译本;上海人民美术出版社2001年出版,刘桂兰译,为青少年版本,出自世界文学名著宝库系列丛书。

本文根据蒋跃、董贺 (2015) 提出的13个汉语计量特征中的型例比 (又称类型比,即TTR值)、词长、句长、各类惯用语的比例等,来探究不同版本在词汇特点、句子复杂性、语篇可读性的差异。词汇复杂度、丰富度的不同,会影响目标人群的对译本接受程度,词汇相对复杂且密集的文章会更适合接受水平更高的人群,反之词汇较为简单的会更适合年龄层次较低的人群阅读。句子复杂性也是译者在考虑译本受众时的一个因素。可读性是检验文本难易程度的重要的参考。

## 二、语料库的创建和检索的方式

### (一) 译本选取

本研究语料选用的是豆瓣评分较高的、评价人数较高的两本,分别是译文出版社于2010年出版,朱建迅、郑康译,为经典译林系列丛书中的一本。另一个是上海人民美术出版社2001年出版,刘桂兰译,为青少年版本,出自世界文学名著宝库系列丛书,该版本简单易懂,为

青少年学生所喜爱。

### (二) 语料库来源

1. 自建语料库的基本步骤如下:首先分别将找到的朱郑本和刘本的PDF版本转写为Word版本,清除语料中的空行、尾注、垃圾字符等。接下来再对其进行分词、词性标注,利用的是肖航分词软件,为了降低数据的误差,会进行二次校对,形成最终的自建语料库。运用语料库检索软件进行进一步的检索。

2. “兰卡斯特现代汉语语料库”(LCMC),该语料库是由英国兰卡斯特大学语言学系的Tony McEnery和肖忠华于2003年建成并向研究人员提供。LCMC的容量为100万词(以每1.6万个汉字对一个英文单词折算,包含了15个文体类型,共收入500篇字数约2,000词(3,200个汉字)的样本,语料为1989年到1993年中国大陆的出版物,还包括5种类型的小说(胡显耀2007),在本文中作为对比语料库使用。

### (三) 语料库检索

对语料库的检索工具有很多,本文运用AntConc (4.1.0) 语料库软件作为首要的检索软件,对语料进行检索分析和统计,AntConc是日本早稻田大学Laurence Anthony博士实验室研发的基于Microsoft Windows操作系统的多平台的语料库分析工具包,由最初一个相对简单的索引程序(Concordance program) 发展为目前一个非常有用的文本分析工具。

## 三、检索结果分析

### (一) 类符/形符比

文本的词汇丰富度一般通过类符/形符比来实现,可以反映文本的词汇量和用词的变化(杨慧中,2002: 153),即TTR值,通过对比TTR值,能够在一定程度上揭示译本的词汇丰富程度由于本文所选取的两个译本容

表1 原文及两译本类符/形符比统计结果

译本	字数	类符数 (token)	形符数 (type)	标准类符/形符比
朱建迅、郑康译本	134, 583	85, 863	14, 410	6.47
刘桂兰译本	107, 155	60, 033	13, 591	3.73

量存在差异。<sup>[4]</sup>本文使用AntConc (4.1.0) 对分词后的语料进行统计, 结果如表1所示:

由表1可知, 经过AntConc (4.1.0) 对数据计算, 朱郑译本的标准类符形符比为6.47, 而刘译本的标准类符形符比为3.73。一般来说, TTR值越大, 说明译者对于用词方面更加的宽泛, 用词的丰富程度相对较高。就这两个译本而言, 朱郑的译本比值相对较高, 刘译本的比值相对较低, 说明朱郑译本在用词方面更加注重词汇的丰富程度。译者考虑该译本面对的是全体大众, 所以译者为了照顾不同年龄层面读者的阅读体验, 会适当的丰富翻译时用到的词汇, 传递更加丰富的内容, 以区别于专门针对中小学的学生出版的青少年版本。而相较于朱郑译本来说, 刘译本在词汇的范围上还是数量上都略少, 这是因为译本出版面向的目标人群不同, 译者也会适当的调整自己在翻译时用到的词汇难度, 以便于读者能够更好地理解。但是, 在某方面也可能受到译者翻译水平的限制或原著影响等原因, 从而导致译本的标准类符形符比相对较低。

例1:

朱建迅、郑康本: 那孩子转身跑去, 在距花一两英尺处停住脚步, 一只手掩着眼睛向街的另一头望去, 似乎发现那里正在发生什么有趣的事情。他随即又拾起一根干草, 脑袋尽量往后仰, 同时把草放到鼻子上, 竭力保持它的平衡。他费劲地扭动身子, 脚步缓缓地向三色堇花挪移。最后他的光脚踏在花上, 用灵巧的脚趾牢牢夹住, 就这样带着他的宝贝一只脚一蹦一跳地离开, 拐过弯就不见了人影。

刘本: 汤姆赶忙跑过来, 在这枝花近旁站住, 他用眼悄悄地向四周观察, 看到周围并没有人注意到他, 他这才不动声色地伸出右脚, 用脚趾夹住了那枝花, 然后迅速地用手拿起来, 轻轻地塞进短衫, 贴近胸口, 装作没事一样离开。

原文: The boy ran around and stopped within a foot or two of the flower, and then shaded his eyes with his hand and began to look down street as if he had discovered something of interest going on in that direction. Presently he picked up a straw and began trying to balance it on his

nose, with his head tilted far back; and as he moved from side to side, in his efforts, he edged nearer and nearer toward the pansy; finally his bare foot rested upon it, his pliant toes closed upon it, and he hopped away with the treasure and disappeared round the corner.

这段主要讲的是汤姆要捡起来艾米·劳伦斯扔掉的玫瑰花, 所以较多突出在对汤姆的动作描写上。在这个例子当中我们可以观察到, 和原文对照后, 相较于刘本, 朱郑本对于原文的描写是更为还原的, 而刘本则更倾向于省略, 这表现为以下三个方面:

1) 朱郑本在这一段中用了十四个动词, 刘本用了十个动词, 朱郑本更贴近一字一词的翻译, 而刘本更加的简略, 比如, 朱郑本“那孩子转身跑去”, 刘本译成“汤姆赶忙跑过来”, 而原文翻译是“The boy ran around and stopped ……”。

2) 刘本省略掉了一些修饰部分, 如, 原文中“his pliant toes closed upon it”, 朱郑本译成“用灵巧的脚趾牢牢夹住”, 用“灵巧的”来修饰脚趾, 而刘本却并未对脚趾做修饰, 更加简洁。

3) 情节上的省略, 如原文中最后“disappeared round the corner”, 朱郑本直接翻译了出来而刘本则进行了情节上的省略了, 直接译作主人公离开, 原因可能是该情节的删除并不影响故事的发展方向。

## (二) 高频词比较

本文利用AntConc软件的Word list功能做出两个译本的前十位的高频词的一个词表, 对其进行对比。Word list是语料库的索引工具中的基本功能之一, 它是按照类符的频率高低列出的类符表。通过比较高频词出现的频率来判断两译本在可读性方面的差异, 另外引入“兰卡斯特现代汉语语料库”(LCMC), 作为对比语料库。以下是根据AntConc中高频词列表的得出数据。

朱郑译本排名前十的词分别是的他、我、在、你、汤姆、是、个、他们、就。

刘译本排名前十的词分别是的了、汤姆、他、就在、是、也、把、都。

LCMC对比语料库排名前十的词分别为的、了、汤姆、他、就在、是、也、把、都。

根据数据可以看出,三个语料库前十位的高频词相差无几,几乎都是助词、介词、连词和代词等单字构成,虽然顺序略有差异,但基本上都大同小异,单字频率在语料库中出现频率较高现象这也是汉语词表较为常见的现象。<sup>[5]</sup>“的”这个结构助词在两个语料库当中都是频率最高的,这也符合汉语当中对于虚词的使用较为频繁的特征。其次,值得注意的是两个译本当中对于人称代词的使用,朱郑本中依次为“他,我,你,他们”,而刘本则为“他,他们”,“我”在刘译本的词频表中更是排在了第11位,由此可以看出,刘本对于第一人称的使用频率不高,主要偏向用第三人称叙述。另外在这里,汤姆·索亚是故事的主人公,但是在文中出现的频次可能会受到翻译的影响,有时翻译成“汤姆”有时是“汤姆·索亚”,这就对频次表的排序造成了影响,但是他的频次相对来说还是较高,可以看出是故事的主人公。

### (三) 句子复杂性比较

通过对兰卡斯特自带汉语语料库的高频语气词进行检索,可看出排在前四位的语气词分别是“吧”“啊”“吗”“呀”,这也是现代汉语中经常使用的语气词,在日常生活中也经常使用,符合现代汉语的特点。

从语气词的统计来看,可以看出朱郑本对于语气词“啊”和“呀”的使用较少,且“啊”在文中间断分布,但是语气词“吧”和“吗”的频率还是较高的。但是刘译本对于语气词的使用频率与LCMC的频率高的语气词排序是大致相当的。“吧”是朱郑译本和刘译本使用频率都很高的词。

从语气词能反映一个人说话的状态、心情还有态度,增强文章的气势和感染力。在语气词的使用方面,刘本几乎每个语气词在文章当中分布的都是相对较为均匀的,照顾到了文章的叙述、人物形象的塑造,并且也是很大程度还原原文真实性的一个体现,说明在翻译时译者注意到了这方面,青少年在阅读故事的时候相对来说会更容易被人物之间的对话吸引,从而产生对人物形象一个整体的理解,提高了青少年的阅读兴趣,可读性更强,也让故事变得更容易理解和品味。

其次,译者考虑到了译本主人公在整个故事当中所起到的一个作用,汤姆·索亚是这个故事的主人公,天生淘气,所以他说出来话的语气应该充满活力、充满疑问与好奇,这一方面可以通过人物之间的对话展示出来,另一方面语气词的加入,也让故事的人物更加的鲜活生动。

例3:

刘本:汤姆在学校附近拦住穿戴整齐的同伴:

“喂,毕利,你有黄字条吗?”

毕利骄傲地回答:“当然有哇。”

“你肯将纸条换给我吗?”

这个例子是一段对话,有问有答,出现了两个“吗”和一个“哇”,可以看出对话中常出现语气词。

### 四、结语

本文运用语料库检索工具,基于计量语言学特征对《汤姆·索亚历险记》的两个中文译本进行比较,目的在于探究不同译本在面对不同的阅读人群时的时候是否会在词汇特点、句子复杂性、可读性方面存在差异。通过研究发现,两译本在词汇方面存在使用差异,反映译本所面向的阅读人群的差异。不同年龄段的人群对于阅读的需求是不同的,较低年龄的人群的阅读能力还不够强,所以用词简单、对话简单、故事脉络清晰的译本是最佳选择,既丰富了儿童的想象力,也在可接受范围内提高了他们的阅读能力,也认识了一些新鲜的词汇。但是在面对更多需求的阅读者的时候,译者在翻译的时候往往用词更加丰富、故事感更强、人物形象也更加的立体,可读性也更强。本文仅限于对两个译本中的词汇丰富度、句子复杂性进行比较,研究有待深入;其次,由于是通过软件进行的分词、标注,在产出数据方面会产生一定的误差,希望本文可以给译本比较研究提供一些思路。

### 参考文献

- [1]徐欣.基于多译本语料库的译文对比研究——对《傲慢与偏见》三译本的对比分析[J].外国语(上海外国语大学学报),2010,33(02):53-59.
- [2]Hermans, T. The translator's voice in translated narrative[J]. Target, 1996(1).
- [3]黄伟,刘海涛.汉语语体的计量特征在文本聚类中的应用[J].计算机工程与应用,2009,45(29):25-27+33.
- [4]徐欣.基于多译本语料库的译文对比研究——对《傲慢与偏见》三译本的对比分析[J].外国语(上海外国语大学学报),2010,33(02):53-59.
- [5]张继东,朱亚菲.基于语料库的《追风筝的人》两译本风格对比研究[J].外语电化教学,2020(05):50-57+8.
- [6]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(06):16-21+92.