

基于定量分析的吴树文翻译文体研究

王子睿 刘善钰



摘要:本研究从语料库翻译文体学研究的理论角度出发,尝试使用编程语言 R 对翻译家吴树文的翻译作品与其他翻译者的翻译作品进行翻译文体的计量分析。“翻译的文体”是原作者文体、译者文体的糅合物,文学的翻译文体无法脱离译者文体的影响,语料库研究方法可以通过定量分析的方法客观地分析译者的文体特征。本研究发现,与参照语料库相比,吴树文的翻译作品在词汇丰富度与成语使用数量方面显得更为丰富。此外,本研究使用了无监督分类法分析了不同作品的 N 元词与虚词的使用特点,聚类分析图与对应分析图的结果显示吴树文的多个翻译文本聚类到相近的位置,可以说明吴树文具有明显区别于其他译者的译者文体风格。

关键词:语料库翻译学;计量文体学;文本挖掘;R 语言

A Comparative Study of Translator Wu Shuwen's Translation Styles Based on Quantitative Methods

WANG Zirui LIU Shanyu

Abstract: This study approaches the topic of corpus translation stylistics from a theoretical standpoint, and aims to conduct a quantitative analysis of translation styles. Specifically, the study focuses on comparing the works of the translator Wu Shuwen to those of other translators, using the programming language R. “Translation style” is defined as a combination of the original author’s style and the translator’s style, with the translator’s style having a significant impact on the resulting literary translation style. Through the use of corpus research methods, the study is able to objectively analyze the stylistic characteristics of translators through quantitative analysis. The results of the study indicate that Wu Shuwen’s translation works display greater lexical richness and more frequent use of idioms when compared to the reference corpus. In addition, the study employs unsupervised classification analysis to examine the usage characteristics of N-grams and function words in different works. The clustering and corresponding analyses show that Wu Shuwen’s multiple translation texts cluster in

close proximity, highlighting the unique translator style that sets him apart from other translators.

Keywords: corpus translation stylistics; quantitative stylistics; unsupervised text classification; R language

1. 引言

吴树文是我国著名的日本文学翻译家,有《春琴抄》《田园的忧郁》等经典文学翻译著作,同时他也是中华诗词学会的成员。其丰厚的文学功底造就了其译作富有文采,句式凝练,大量使用文言连词、行文古朴典雅的文体特点。本研究尝试基于定量研究方法,以其他译者的重译本为参照库对吴树文的翻译文体进行计量对比分析。

语料库研究作为一种实证性的定量研究方法,拓宽了传统翻译研究的范畴。通过语料库与计量方法在翻译研究中的应用,研究者可以以更客观、清楚的方式呈现译者的翻译风格。本文应用语料库翻译文体学视角,运用语料库与计量统计法对翻译家吴树文的译者文体风格进行考察与分析。同时,使用基于编程语言 R 的文本挖掘技术对语料库的文体进行了多指标探索,旨在回答以下问题:第一,译者吴树文是否具有显著的译者风格特征?与重译本参照语料库有何差异?第二,本研究限定的 2 种二元词和虚词是否可以对译者的文体风格进行区分?吴树文的译作是否具有相似的翻译文体特征?

2. 文献综述

2.1 文体与翻译

文体学与翻译学都是拥有悠久历史的人文学科。文体学连接语言学与文学,关注文本主题意义与美学效果密切相关或偏离常规的语言特征描写与阐释(申丹,2002),而翻译学则着重观察在翻译过程中译者受到译入语、译出语的语言习惯影响而不自觉产生的译者“痕迹”。著名的日本文学翻译家林少华(2009)也曾指出,翻译只能是原作者文体和译者文体或者说文体的翻译和翻译的文体相妥协相融合的产物。

随着语料库语言学与计算机技术日新月异的发展,人文学科之间的跨学科研究空前发展,翻译学与文体学与时俱进地吸收了语料库语言学的基于数理统计的研究范式(王克非,2006;张德禄,2007),衍生出了语料库翻译学与语料库文体学等重视量化分析、实证研究的新生学科(卢卫中等,2010;雷茜等,2016;胡开宝,2018)。利用语料库进行文学与翻译研究,可以科学地观察出文体学家观察不到的某些文体差异,对作者或者译者的一些难以捉摸的、习惯性的语言特征进行描述、分析、对比,从而较为信服地说

明个人文体的存在,也能够有力地证明林少华(2009)所说的翻译是“相妥协相融合的产物”。

与此同时,由于人文学科的跨学科发展与研究方法的互相借鉴,基于计量与语料库方法的文体学与翻译学也有学科共融之势。在这种学科相互融合、研究范式互相接近吸收的背景下,翻译研究、文体学和语料库语言学相互关照,催生了语料库翻译文体学的诞生和成长(王峰等,2017)。

2.2 基于计量方法的语料库翻译文体学

语料库翻译文体学的发展主要得益于计算机文本处理能力的进步与人文学科的跨学科化。黄立波(2009;2014)指出语料库翻译文体学通过语言对比这一共性特征,将传统文体学研究 with 翻译研究加以结合,以基于真实语料和词频信息的统计数据作为分析文体的重要参考。Saldanha(2011)对译者风格研究提出了2种不同的诠释方式,即“原文本型译者风格”研究和“目标文本型译者风格”研究。前者主要关注译者如何在翻译文本中表现原文中的某些语言特征,而后者则主要关注译者特有的表达方式。相比之下,国内译者风格研究起步稍晚,且主要采用原文本型研究方法,仅少数研究将目光投入译者的文体研究并采用了目标文本型研究方法(刘泽权等,2011;黄立波等,2012)。

虽然一些翻译学研究者开始接受基于计量方法的语料库翻译学研究模式,并且以各种翻译实践中的案例,以定性与定量相结合的方式展开了各种翻译研究,但是此类研究因受限于技术、工具的不足,加之研究者采用的计量方法往往局限于平均词长、平均句长、标准化类符/形符比(STTR)等基本形式参数,其结果难以对译者的总体文体特征进行说明。以往的工具也不具备可视化能力,无法更加直观地观察数据与分析的结果,这就促使研究者对使用工具、指标、定量研究方法进行更新。

语料库翻译文体学需要对既往的研究方法与理论框架进行反思,对此,在综述研究中,黄立波(2018)指出译者风格研究理念主要来源于早期关于作者权归属和计量风格学研究,译者风格研究不同于仅关注局部语言特征的计量风格研究,基于语料库的翻译文体研究应当拓宽思路,向语义、语用、修辞、社会与文化参数拓展,借鉴语料库文体学、计量语言学、计算语言学等相邻领域的研究方法,将定量统计与定性分析有机结合起来,拓宽翻译文体或风格研究的范围。胡开宝(2018)指出当代数字人文研究愈来愈重视文本深度挖掘和智能分析等方法的应用,强调数据的可视化,翻译研究也该吸收先进技术,深度分析翻译本质和翻译规律。数字人文视域下的翻译研究呈现数字化、实证性、文本挖掘技术的融合,微观描写与宏观解释并重的趋势。在个案研究中,詹菊红和蒋跃(2017)尝试基于机器学习中的支持向量机算法对《傲慢与偏见》的2个译本进行了译者的判别分析,有效地发现了译本之间在语言形式参数上的差异。孔德璐(2021)尝试以机器学习中的支持向量机、朴素贝叶斯、聚类分析等作为算法,证明了《苔丝》的张若谷译本具有独特的文学译者特征。

在语料库翻译学研究领域,众多学者将理论方法积极与文体学、叙事学等理论相结

合,实践方面紧跟数字人文潮流,引入定量分析与文本挖掘技术,做出了许多具有创新性又极具跨学科意识的研究。本研究在既往的研究范式上作出新的尝试,使用无监督学习中的聚类分析和对应分析对部分文首和文末的二元词与 65 个虚词的使用频率进行译者风格的探索。

3. 文本语料处理与文体特征比较

3.1 实验设计

首先在语料库方面,本研究建立了 2 个语料库,收录吴树文 6 部翻译作品作为观察语料库,简称 W 库;另收集 6 部岳远坤、曹曼、章蓓蕾的重译作品作为对比语料库,简称 R 库。为了保证文学作品本身的文体特点不影响数据的分析,本研究保持 R 库收录的译出语作品与 W 库一致。

在分词器上,为了得到更高的分词精度和使用新词发现功能,本研究利用张华平和商建云(2019)开发的 NLPPIR 分词器对 W 库、R 库分别进行分词、词性赋码,使用 Benoit 等(2018)开发的 R 开源包 `quanteda` 包进行语料库管理和语料清洗,删除部分影响数据准确性的字母、数字、换行符等信息。语料库 W 库、R 库收录的作品如表 1 所示。

表 1 语料库内收录作品

W 库		R 库	
作品	译者	作品	译者
《阿绢兄妹》	吴树文	《阿绢和她的哥哥》	岳远坤
《春琴抄》	吴树文	《春琴抄》	曹曼
《后来的事》	吴树文	《后来的事》	章蓓蕾
《门》	吴树文	《门》	章蓓蕾
《三四郎》	吴树文	《三四郎》	章蓓蕾
《田园的忧郁》	吴树文	《田园的忧郁》	岳远坤

在文体参数方面,首先采用 `quanteda` 包中的 `R(Guiraud)` 指数和 `U(Dugast)` 指数来计算 W 库与 R 库的词汇丰富程度。其次,在成语使用率方面,由于计算机赋码技术对长文本的识别精度不高,本研究将搜狗词库收录的 54089 条四字成语嵌入分词工具,并且使用 `quanteda` 包中的 `tokens_select` 函数计算每个文本中成语的出现频率。在句长方面,用 R 语言自定义两个函数分别统计句长与句段长。

为了检验译者翻译作品中文首和文末的语言表达和虚词使用是否可以证明译者文体风格的存在,本研究选取了 2 步实验进行验证。首先抽取了“,_单词”“._单词”与“单词_,”“单词_。”2 种模式的二元词,测算距离后通过 Sébastien Lê 等人(2008)开发的

FactoMineR 包进行 K 均值聚类分析可视化观察不同译本之间的相似程度。其次抽取了 65 个虚词通过 FactoMineR 包绘制对应分析图来观察不同译者的翻译作品与虚词使用之间的关系。

3.2 无监督学习方法

本研究使用了聚类分析(Cluster Analysis)和对应分析(Correspondence Analysis) 2 种无监督学习方法对译者的文体特征进行分析。在数字人文研究中,有不少研究将文本挖掘技术引入传统文学研究中来,例如刘颖和肖天久(2014)针对金庸和古龙的小说中不同词类的使用频率进行 K-means 聚类,发现二者文体存在显著的差异。叶雷(2016)利用聚类方法分析《红楼梦》,证明了聚类分析方法在作者文体识别研究中的有效性。对应分析也称关联分析、R-Q 型因子分析,对应分析的基本思想是将一个列联表中的行与列的各元素以低维空间的形式表示出来,以降维的思想达到简化数据结构的目的,使列联表中的数据在图上直观、明了地显示出来。

4. 文体参数对比

4.1 词汇丰富度

为了把握吴树文的基本的文本特征,本研究尝试使用编程语言 R 及相关开源包,从词汇丰富度、成语使用、句长与句段长方面考察吴树文翻译文本语料库(W 库)与重译文本语料库(R 库)的差异。

首先,形符(Token)即一个分词后的单词单位,而类符(Type)是指不重复的形符总数。词汇丰富度指标可以体现译本词汇的丰富程度,观察译者的用词情况。但是由于形符与类符计算方式截然不同,所以本研究尝试同时使用两个指标,即采用了 quanteda 中的 textstat_lexdiv 函数计算 R(Guiraud)指数和 U(Dugast)指数,将其结果分语料库统计,然后以箱线图的形式输出观察。如图 1(a)图、(b)图所示,W 库的 R 指数与 U 指数普遍高于参照 R 库,呈现出翻译家吴树文较参照语料库而言用词更丰富多样的特点。

4.2 成语的使用率

四字成语在音韵上具有节奏美,词汇角度上具有整齐美,语义角度上具有意象美,适当的四字成语的使用可以增强文章的表现力,增加译文独特的韵味(李大鹏等,2015)。四字成语源于我国古代人民长久的智慧与思维的凝结,而日语受汉语影响,经由文化误读产生了大量带有民族特色的四字成语,所以日语文本与汉语中的成语有一定的不对等性,译者往往会选择加译、减译、改译等翻译策略。在翻译的过程中,经过译者的加工、过滤,原文的文体往往会被重塑,带上译者个人的风格。翻译文本中四字成语占总形符比的大小可以体现出译本的归化与异化倾向。成语是相沿习

用的,有很强的历史继承性与很深的文化烙印,使用成语会让译文带有明显的文化归化特征,文化方面的归化不仅会限制读者视野,更会让读者产生时代、民族和文化的错位感(蒋志辉等,2014)。为了探究翻译家吴树文在成语层面的翻译文体风格,本研究统计了不同译者的成语使用频率,并绘制箱线图如图1(c)图所示。可以观察出翻译家吴树文在翻译时使用了更多的四字成语,更偏向目的语读者语境,呈现言简意赅的文体特征。

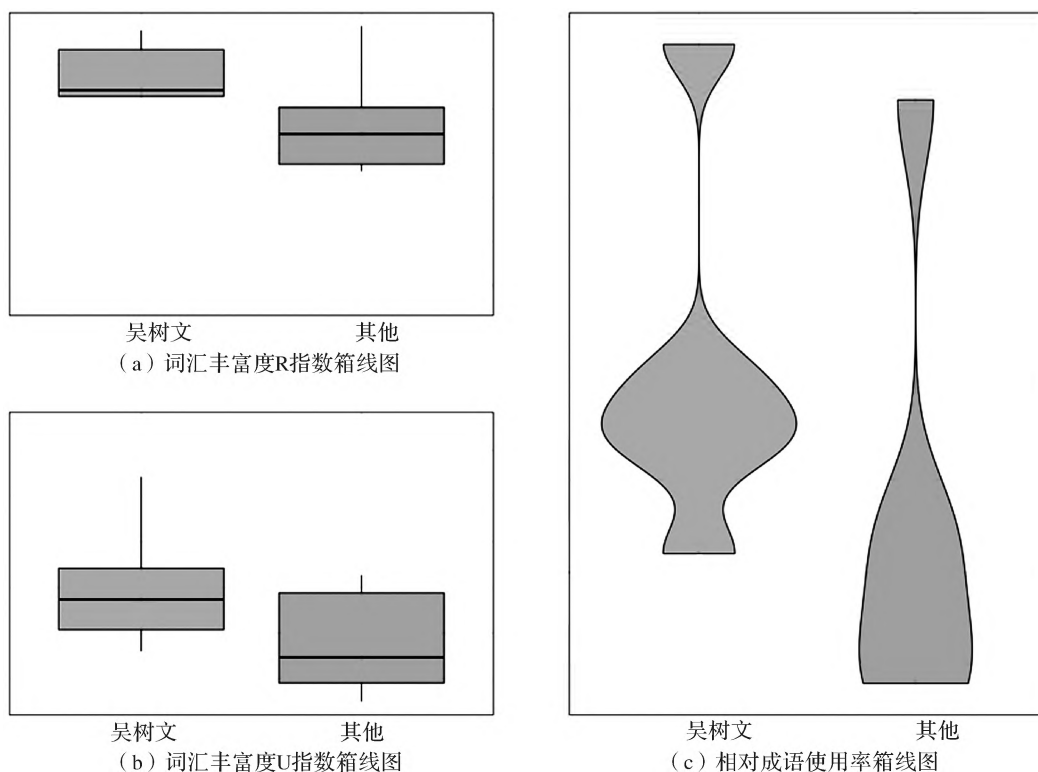


图1 不同译者词汇丰富度R指数、U指数与成语使用频率比较

4.3 句长与句段长

句长作为文体研究指标的历史已久,而句段长往往被研究者所忽视。句长是表示一句话里所含词数或字数,句可以再细分为句段,一个句段则是以问号、句号、逗号、感叹号、分号作为分割标志。实际上,在汉语中,句段长比句长更能体现出翻译语言的个性特点(秦洪武,2010;肖忠华,2012)。句长与句段长反映了作者的文体特征,翻译文本中的句长与句段长则体现了译者对译出语的结构性改造,不仅具有译出语作者的文体特征,也隐藏着译者的翻译文体特征。由于不同作者的文体特征不一致,全部翻译文本的平均句长难以反映译者个人的文体风格,为了保证语料库具有可比性,所以本研究根据W库与R库的分类以作品为单位分别统计每个作品对应的数据,如表2所示。

从表 2 可以观察到,无论是 W 库还是 R 库,日本文学作品《春琴抄》的平均句长是所有作品中最高的,这与原作尽可能省略标点符号的文体因素脱不开关系,受原作文体影响与限制,相应的译本中也呈现出了句长数值较高的文体特征。但观察句段长时,可以发现吴树文更倾向于将较长的句子拆解为句段,对原文文体进行改造。而在佐藤春夫《阿绢兄妹》(又译为《阿娟和她的哥哥》)与《田园的忧郁》中,与 R 库的翻译文本相比,吴树文保留长句的特点更明显。在翻译夏目漱石“人生三部曲”的(《三四郎》《后来的事》《门》)3 部作品时,都保持着较短的平均句长、平均句段长,体现出了翻译家吴树文在对不同文学作品翻译时采取不同翻译策略的特点。

表 2 平均句长、平均句段长、每句所含句段长数对比

译者	作品	平均句长/字符	平均句段长/字符
吴树文	《三四郎》	23.53	9.71
吴树文	《后来的事》	26.68	10.19
吴树文	《春琴抄》	36.69	10.34
吴树文	《田园的忧郁》	34.28	12.37
吴树文	《门》	29.35	10.29
吴树文	《阿绢兄妹》	27.48	10.32
章蓓蕾	《三四郎》	25.18	10.00
章蓓蕾	《后来的事》	27.41	10.25
曹曼	《春琴抄》	36.85	11.23
岳远坤	《田园的忧郁》	30.20	11.14
章蓓蕾	《门》	30.69	10.50
岳远坤	《阿绢和她的哥哥》	26.46	9.76

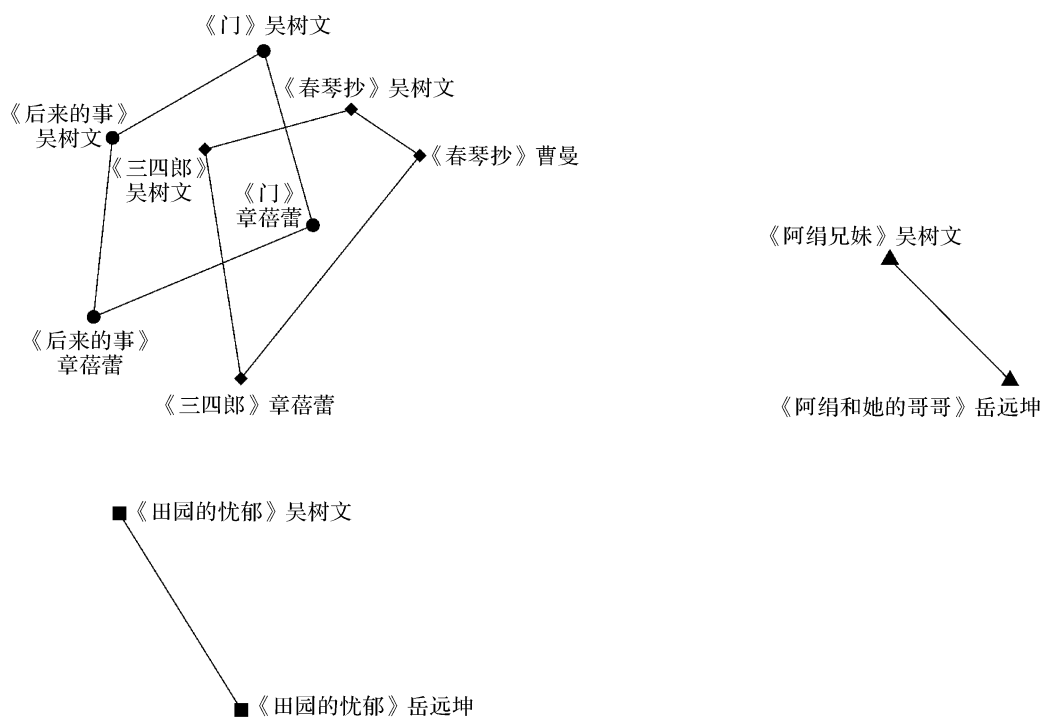
5. 基于多元统计方法的译者风格分析

5.1 基于聚类分析的译者的文首、文末文体风格

在特征的选择上,为了尽量减少原作文体特征对译者风格分析产生影响,本研究在 K 均值聚类中尝试性地选择了反应译者文首文体风格的“,_单词”“,_单词”与文末文体风格的“单词_,”“单词_。”2 种二元词模式。

本研究先将分词后的语料库转化为二元词的形式,再抽取上述的 2 种二元词模式,导入 quanteda 包处理为词频文档矩阵,再对统计数据进行标准化处理,进而对得到的结果进行 K 均值聚类分析,因语料库中包含 4 位译者,故聚类时选择聚成 4 类,如图 2 所示。

由图 2(a)可以观察到,体现译者文首文体风格的“,_单词”“,_单词”2 种二元词聚类效果并不理想,不同译者的原作被聚类到相同的簇中,例如《三四郎》的吴树文译本和章蓓蕾译本被聚类到同一个簇。但文末文体风格的“单词_,”“单词_。”的 K 均值聚类图将吴树文的 6 部翻译作品聚成一簇,如图 2(b)所示,说明选取的文末二元词可以明显地判别出吴树文的文体,而其他译者例如岳远坤、章蓓蕾也被分类到了不同簇,曹曼译《春琴抄》也没有因为译出语文体与吴树文译本相似而聚类到一簇,证明文末二元词在判别译者的文末文体风格上比文首的二元词有更好的效果。这是因为译出语在翻译到中文时往往会伴随着显化主语的倾向,位于句子前段的二元词会带有显著的原作的文体风格。



(a) 基于“,_单词”“,_单词”的 K 均值聚类

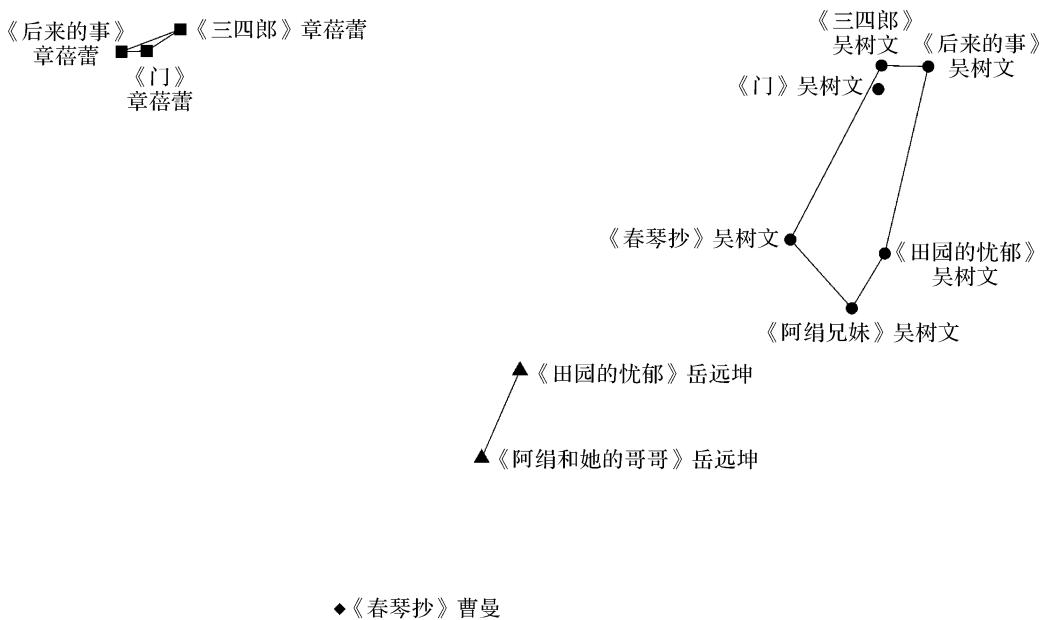


图 2 基于文首、文末二元词的 K 均值聚类

5.2 基于对应分析的虚词使用风格分析

虚词泛指仅具有语法意义但没有实质性意义的词语,不能独立成句,必须依附于实词而存在。翻译文本中的实词往往不会因翻译者的主观倾向而改变其在译入语中的意义,而虚词则不同,译者在翻译中对虚词的使用往往是无意识的体现,虚词的使用频率与文学者或译者的文学素养、翻译策略、所处时代等多种要素有关,常被作为作者判别的重要依据。刘颖和肖天久(2014)将 49 个虚词作为考察对象,利用层次聚类方法考察了金庸与古龙的小说在虚词使用上的文体差异。施建军(2011)将 44 个文言虚字频率作为特征向量,利用支持向量机算法断定《红楼梦》前 80 回和后 40 回并非同一人所作。

不同译者的虚词使用习惯不同。通过对比不同译本的虚词使用情况,可以把握译者在虚词使用上的文体特征。本研究选择了 61 个具有代表性的文言虚词来进行对应分析,以检验不同译者的翻译作品与虚词使用情况的对应关系。本文抽取的虚词为“和、跟、与、同、及、况、况且、何况、乃至、则、乃、就、于是、说到、此外、像、如、一般、比方、却、但是、然而、而、偏偏、只是、不过、至于、致、不料、岂知、原来、因为、由于、以便、因此、所以、是故、以致、或、抑、若、如果、若是、假如、假使、倘若、要是、譬如、像、好比、如同、似乎、等于、不如、不及、与其、虽然、固然、尽管、纵然、即使”。

本研究抽取了以上所述文言虚词,并进行频率计算与标准化,绘制对应分析图,将代表译本的图形调整为圆形,得到结果如图 3 所示。通过图 3 可以观察到吴树文的译本都位于图的右侧,与文言虚词的使用呈现了相应的分布特征,其他译者的译本则位于图的左侧与下侧居多,与吴树文的文体特征相比,呈现出不同的文言虚词使用

特征。

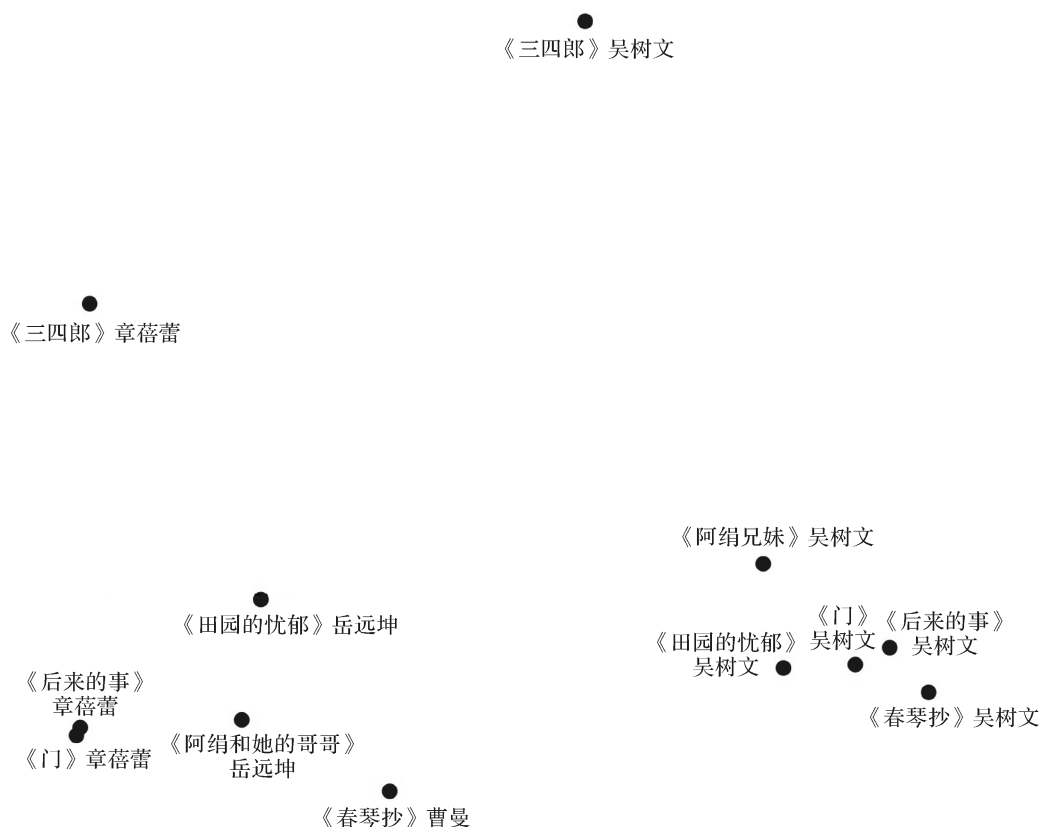


图3 不同译作虚词使用频率的对应分析

6. 实验结果分析

本研究首先从词汇丰富度、成语的使用率、句长与句段长等基础参数考察了翻译家吴树文的基本文体特征。在此基础上还尝试使用体现译者文首、文末文体风格的二元词进行了K均值聚类分析,此外还抽取了61个文言虚词进行了对应分析,以期从多元视角探讨吴树文的翻译风格。

通过实验得知,在词汇使用方面,R指数与U指数显示译者吴树文使用词汇较参照语料库更丰富、重复率更低的特点。成语使用率显示译者吴树文较参照语料库使用了更多的四字成语。句长与句段长显示吴树文在应对不同作品时对译出语的语言结构进行了不同程度的调整,印证了“翻译文体是作者文体与译者文体的融合物”的观点,句长与句段长不仅由译者操控,而且也深受译出语文体的影响。本研究为了尽可能消除译出语原作文体对译者文体判别所产生的影响,尝试性地使用了体现译者文首、文末文体风格的二元词进行K均值聚类分析,发现体现文末文体风格的二元词可以减少译出语文体特征的干扰,有效地证明了译者文末表达存在译者个人的文体特征。最后本文抽

取 61 个具有代表性的虚词,利用对应分析对 2 个语料库的 12 部作品进行了比较,对应分析图显示吴树文的 6 部作品被聚在相近的位置,说明吴树文在虚词使用方面呈现出独特的译者文体特点。

本研究为了减少译出语文体对数据的影响,在尝试进行 K 均值聚类时,没有对所有的二元词进行抽取,仅抽取了其中文首、文末的 2 种二元词,减少了其他实词对聚类分析的影响。如果不人为地制定 N 元词的选取规则,在同类实词或者译出语即原作文体的影响下,同一部作品的译本会被聚类到一簇,影响研究结果的可信性。正如林少华曾言,纯净水一般文体的翻译是没有的,翻译只能是原作者文体和译者文体或文体的翻译和翻译的文体相妥协相融合的产物。关注译者文体的翻译文体研究不仅要关注翻译文体本身,而且要把握原作者文体对翻译文体的影响,如果将译者的文体比作“脚印”,那么原作者的文体即“影子”,在运用数理化统计方法对译者的翻译文体进行研究时,应该尽量减少原作文体对数据分析产生的影响。

最后,本研究受限于电子化语料的不足,分析对象仅选取了译者吴树文的 6 部作品与重译本参照语料库的 6 部作品,在分析方法中也仅选取了聚类分析与对应分析 2 种算法,在未来应该扩大语料资源,尝试采用监督学习方法进行译者文体的判别研究。

参考文献:

- 胡开宝,2018. 数字人文视域下翻译研究的进展与前景[J]. 中国翻译,39(6):24-26.
- 黄立波,2009. 翻译研究的文体学视角探索[J]. 外语教学(5):104-108.
- 黄立波,2014. 翻译研究中的文体观:研究综述[J]. 燕山大学学报(哲学社会科学版),15(1):49-55.
- 黄立波,2018. 语料库译者风格研究反思[J]. 外语教学(1):77-81.
- 黄立波,朱志瑜,2012. 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(5):64-71.
- 蒋志辉,张焕香,2014. 论翻译中使用汉语成语的原则[J]. 英语研究(3),49-52.
- 孔德璐,2021. 基于机器学习的文学译者风格考察:以张谷若为例[D]. 上海:上海外国语大学:II.
- 雷茜,张德禄,2016. 现代文体学研究方法的新发展[J]. 现代外语(2):278-286,293.
- 李大鹏,张帅,2015. 四字结构在文学作品翻译中的运用:以《喜福会》程乃珊(2006)译本为例[J]. 郑州航空工业管理学院学报(社会科学版)(6):147-149.
- 林少华,2009. 文体的翻译和翻译的文体[J]. 日语学习与研究(1):118-123.
- 刘颖,肖天久,2014. 金庸与古龙小说计量风格学研究[J]. 清华大学学报(哲学社会科学版)(5):135-147,179.
- 刘泽权,刘超朋,朱虹,2011. 《红楼梦》四个英译本的译者风格初探:基于语料库的统计与分析[J]. 中国翻译(1):60-64.
- 卢卫中,夏云,2010. 语料库文体学:文学文体学研究的新途径[J]. 外国语(上海外国语

- 大学学报)(1):47-53.
- 秦洪武,2010. 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(上海外国语大学学报)(4):73-80.
- 申丹,2002. 论文学文体学在翻译学科建设中的重要性[J]. 中国翻译(1):11-15.
- 施建军,2011. 基于支持向量机技术的《红楼梦》作者研究[J]. 红楼梦学刊(5):35-52.
- 王峰,刘雪芹,2017. 融合 共生 互动:语料库翻译文体学理论参照体系[J]. 西安外国语大学学报,25(1):105-109.
- 王克非,2006. 语料库翻译学:新研究范式[J]. 中国外语(3):8-9.
- 肖忠华,2012. 英汉翻译中的汉语译文语料库研究[M]. 上海:上海交通大学出版社:70-73.
- 叶雷,2016. 基于计量文体特征聚类的《红楼梦》作者分析[J]. 红楼梦学刊(5):312-324.
- 詹菊红,蒋跃,2017. 机器学习算法在翻译风格研究中的应用[J]. 外语教学(5):80-85.
- 张德禄,2007. 功能文体学研究方法探索[J]. 四川外语学院学报(6):12-16.
- 张华平,商建云,2019. NLP-Parser:大数据语义智能分析平台[J]. 语料库语言学(1):87-104.
- BENOIT K, WATANABE K, WANG H, et al., 2018. Quanteda: An R package for the quantitative analysis of textual data[EB/OL]. (10-06)[2023-01-08]. <https://www.theoj.org/joss-papers/joss.00774/10.21105.joss.00774.pdf>.
- LÊ S, JOSSE J, HUSSON F, 2008. FactoMineR: An R package for multivariate analysis[J]. Journal of Statistical Software, 25(1):1-18.
- SALDANHA G, 2011. Translator style[J]. The Translator, 17(1): 25-50.

作者简介:

王子睿,男,1998年2月,山东淄博人,广西大学外国语学院硕士。研究方向为语料库翻译学、计量文体学。

刘善钰,女,1981年4月,广西百色人,博士,广西大学外国语学院讲师。研究方向为日语教育、日语语言文学、语料库语言学。