

# 基于计量文体学的谷崎润一郎小说文体变化研究

广西大学 王子睿 刘善钰

[摘要] 本研究基于数字人文视角,将谷崎润一郎的26部文学作品作为考察对象,使用词汇丰富度、句长文体测量指标和多变量分析方法的聚类分析对其文学作品进行了探究。既往的文学研究者认为谷崎润一郎在关东大地震后移居关西地区后,其文学文体理念发生了较大的变化。为了验证以上假说,本研究尝试使用计量文体学与文本挖掘的手法从语言现象、文体特征、文献作证三个角度进行论证。结果证明在1930年谷崎润一郎移居关西后其文体特征的确发生了显著的变化。

[关键词] 计量文体学;谷崎润一郎;文体研究;作者判别

## 1. 引言

计量文体学一般被认为源于作者判别研究(Authorship attribution),可以追溯到19世纪末 Mendenhall(1887)的研究。计量文体学在国内一般被称为计算风格学、语料库文体学、统计文体学等。近年来,出现了用计量方法对《红楼梦》的真假本进行作者文体判别的分析研究,也有把译者文体作为判别对象的语料库翻译文体研究。卢卫中、夏云(2010:48)将语料库文体学的研究领域分为了三个主要领域:(1)以个体语言特征的文体研究;(2)对作品主题的语言表现方式的研究;(3)作家风格研究。计量文体学采用演绎法和定量的手段,对大量语料进行量化分析,分析得出或验证某一结论,而不是仅依靠直觉或语感做判断(隋桂岚,等2003:82)。基于计量文体学的文学文体研究在国内所见不多,本文将在介绍日本计量文体学的最新成果之上,对谷崎润一郎文学创作生涯(52年)中的26部作品进行文体计量分析,尝试使用定量分析方法对文学研究者的观察与假说进行验证。

计量文体学发源于西方,但传入日本后日本研究者根据日语的语言特点创造了独特的文体研究方法,提出了 MVR<sup>①</sup>、句、逗号的使用方式等诸多文体测量指标,此外还引入了多变量分析和机器学习。由此,日本的计量文体学研究呈现出测量指标不断丰富、分析方法不

断创新、引入推断性统计和机器学习等新方法的特点(毛文伟,2021:87-88)。

谷崎润一郎(1886—1965)是日本著名的唯美主义作家,著有《春琴抄》《细雪》等著名小说。在文化研究上也颇有建树,著有日本文化评论随笔《阴翳礼赞》、近代日本文学文体评论《文章读本》等,还用现代日语翻译了《源氏物语》全文,被国内作家施蛰存称为“文体家”,一生文体风格多变。

文学界认为谷崎润一郎在关东大地震后移居关西地区后其文学文体理念发生了较大的变化。谷崎润一郎虽然不是日本关西地区人,但是自1923年移居到关西地区后,便在其作品中使用了大量的关西方言,甚至还聘用了助手把日本关东地区语言翻译成关西方言(安井寿枝,2008:35)。许多日本学者在谷崎润一郎古典回归时期的开端作品是哪一部的问题上有争议(猪口洋志,2022:79-83)。但学界暂时还没有基于计量文体学视角进行验证的相关研究。为了填补研究的空缺,并验证学者的假说,本研究尝试使用计量文体学与文本挖掘的手法对谷崎润一郎的26部文学作品进行定量分析,从文体特征、方言使用、文献佐证三个角度印证了其移居关西前后的作品中确有明显的文体差异。

## 2. 先行文献综述:日本计量文体学研究综述

日本的计量文体学可以追溯到20世纪30年代。例如波多野完治对谷崎润一郎的《金与银》与志贺直哉的《雨蛙》两篇文章的句长、符号、词性等特征进行统计分析,指出文章的文体特征与作者的性格有关联。此外,较早的研究还有安本美典(1957:4)的研究,该研究指出日本平安时代的文学巨作《源氏物语》并非同一人所作,也对宇治十帖的作者问题进行了计量研究。随着计算机的普及与发展,青空文库、国立国语研究所等大规模语料库也开放使用。日本传统的人文研究领域也引入了统计学的验证方法与语料处理技术,使跨学科研究走向繁盛,促进了语料库语言学、计量文体学、文献学、自然语言处理等领域相关研究的发展。

中尾桂子(2010:4-5)指出,日语的文本挖掘分析主要集中在两个领域,一是对语言现象进行分布、概念上的定量分析研究,例如计量国语学、计量语言学、语料库语言学等研究;二是关注作者或作品文体特征的判别、作品感情分析,通常将数个文本作为研究的对象,此类研究被称为计量文体学。此外,随着日本自然语言处理技术的发展,日本研究者还开发了一系列的文本挖掘工具,如日本同志社大学的金明哲教授基于R语言开发的MTMineR、立命馆大学教授樋口耕一(2004)开发的文本挖掘分析工具KH coder等,可以说,日本的文本挖掘工具非常丰富,这使文本挖掘与统计学在文学、经营学、教育学以及政治学中得以广泛应用(齋藤朗宏,2012:3-6)。近年来也有国内学者开始学习使用此类软件,并且应用到了文献计量研究、话语分析研究中。在这一方面,金明哲教授研究室做出了丰富的成果,例如芥川龙之介作品的创作时间推测研究(金明哲,2009:89-103)、宇野浩二的文体变化研究(劉雪琴,金明哲,2017:128-143),还有涉及作者判别的研究(財津亘,金明哲,2018:417-425)。

目前国内学界对西方和日本计量文体学的关注度还不够,借鉴日本的文体学计量方法

分析日本文学的研究所见不多,故本文将对此进行尝试。

### 3. 谷崎润一郎文体研究

中日学界关于谷崎润一郎文体的研究非常丰富。谷崎润一郎一生文体风格多变,有用全汉字写的《诞生》、用平假名写的《盲目物语》、极少使用标点符号的《春琴抄》、用关西方言写的《卍》等作品。在对谷崎润一郎文体的相关研究中,比较一致的观点是:谷崎润一郎移居关西以后,文体风格回归日本传统,在小说的主题中使用大量的日本古典元素,其对话描写排斥欧化、极其细致、强调音韵美等。谷崎润一郎虽然前期属于恶魔主义作家,但是移居关西后回归了古典文学,将传统的日本美学作为创作源泉,创作了《春琴抄》等作品,可以说这一时期的谷崎润一郎创作了许多艺术性极高、具有古典叙事美学的小说。

谷崎润一郎(1934:21)曾在《春琴抄后语》中用《源氏物语》帚木篇举例形容当时自己所追求的文体,“虽然很难区分(帚木篇)对话的引语和叙述的部分,也很难知道文中这段话是谁说的,可是只有这样才能体现出日本文学的美”。他又称“在《卍》里考虑到读者的阅读习惯加上了引号,在《芦刈》中又把它去掉了”。为了达到这种美学境界,谷崎润一郎试图将会话融入叙事中,减少不必要的直接引语,由此可见在古典回归时期,其对文体进行了多种改造。1934年的《文章读本》是谷崎润一郎移居关西后不久发表的一部作品,该作品透露出其对语言和创作观念,是谷崎氏将目光从西洋文化投向日本传统文化的重要节点。叶渭渠(2005:77-93)也指出谷崎润一郎是因为东西方文化中有着不可调和的矛盾,对具有含蓄性和余情性的日本传统文学的文体风格受西方近代文学和文体传入的影响而日渐式微表示遗憾。为了不失去理想的故土,以《痴人的爱》《孔》《各有所好》等作品为开端,谷崎氏告别恶魔主义,开始关注乡土,以新浪漫主义回归古典和传统。谷崎润一郎曾在《饶舌录》(谷崎润一郎著、汪正球译,2000:60-61)中言及东方与西方文化:“东方式的趣味,事物的思考方式,体质,性格……并不单单只限于文学艺术……存在着跟西方不同的独特的事物。”这也说明了谷崎润一郎对文体的改造源于其对东西方文化差异的认知,尝试在逐渐西化的日本文学文体中找到日本传统文学的美学意义。

猪口洋志(2022)也指出古典文学时期的作品不仅在取材上使用古典主题,在叙事方面这些作品都引入了一个极具作者特点的叙述者。例如,以日本后南朝时期为题材的历史小说《吉野葛》,故事一开始出现了一个让读者以为是作者本人的“步行者”,讲述着自己20多年前的见闻。实际上从《卍》开始,谷崎润一郎就开始有意识地采取“物语式”的讲述方式,与近代小说隐藏叙事者采用“描写”的方式相比,这种小说中往往会有具象化的叙事者出现(诸飞燕,2012:82)。

从以上先行研究可以推断,谷崎润一郎在移居关西后,其认为东方文化有着不输于西方的独特的文体与美学价值,将目光投向了古典日本文学的叙事方式和文体结构,于是在自己的新创作中不断尝试,导致了其古典回归时期作品的文体特征多变。

4. 文章特征判别

4.1 研究设计

首先收集语料,从青空文库筛选出谷崎润一郎所有电子化的文学作品,除去随笔和翻译文本(《三人法师》《厌客》两部),删除日语的所有注音假名,共计获得26部作品。以1923年关东大地震后谷崎润一郎移居关西为界限,将其分为 pre 和 post 两部分,根据其年代整理如下表所示。

表1 语料库内收录作品(以移居关西为界限)

移居关西前 (1909—1923)	《刺青》《The Affair of Two Watches》《少年》《帮間》《秘密》《金色の死》《恐怖》 《二人の稚児》《途上》
移居关西后 (1924—1961)	《痴人の愛》《蘿洞先生》《卍》《蓼喰う虫》《吉野葛》《盲目物語》《紀伊国狐憑漆掻語》 《覺海上人天狗になる事》《蘆刈》《春琴抄》《猫と庄造と二人のおんな》《細雪01上卷》 《細雪01中卷》《細雪03下卷》《少将滋幹の母》《鍵》《瘋癲老人日記》

以年代为参数,将词汇丰富度与句长绘制为折线图,观察其在不同时期的具体文体特征变化。其次以1923年为界限将作品分为两类,即1923年前和1923年后,从中挑选特征词,使用箱线图和非参数检验来观察二者在部分关西方言使用上是否存在显著性差异。

本文以基础文体参数作为变量,对26部作品进行聚类分析,绘制热图,观察在语言特征上作品之间的近似关系。

4.2 词汇丰富度与句长

首先,为了测量谷崎润一郎作品中用词的历时变化,本研究采用 R 指标和 U 指标来观察其作品的文体特征,绘制折线图如图1所示。

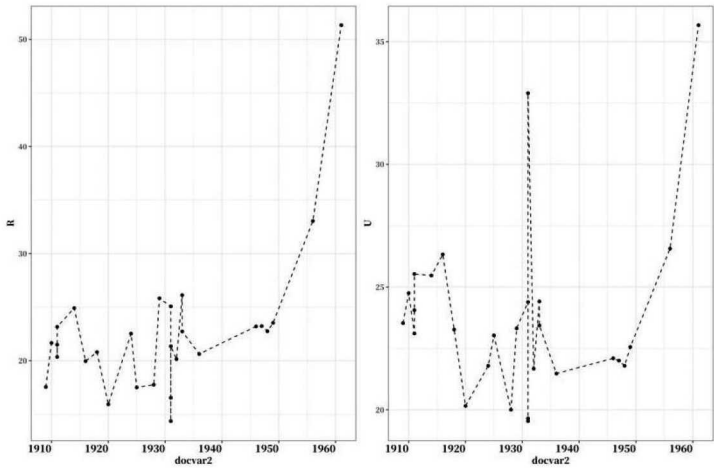


图1 作品词汇变化性分析(以年代为单位)

通过图1可以发现:谷崎润一郎早期作品词汇变化性较为平稳、一致,在整体中的水平较低。到了1930年前后,谷崎润一郎的作品发生了较大的词汇使用变化。此后到1960年完成的最后一篇《疯癫老人日记》,其词汇使用呈现出更为丰富的趋向。由此可以推测出,在1930年左右谷崎润一郎的文体观发生了变化,而其年老时的文体观也发生了较大的变化。

在句长和句段长方面,本研究使用了自定义函数自动计算出每个文学作品的年代,以年代为 X 轴绘制了句长和句段长的点图,红色线为句长值的拟合曲线,蓝色线为句段长值的拟合曲线。

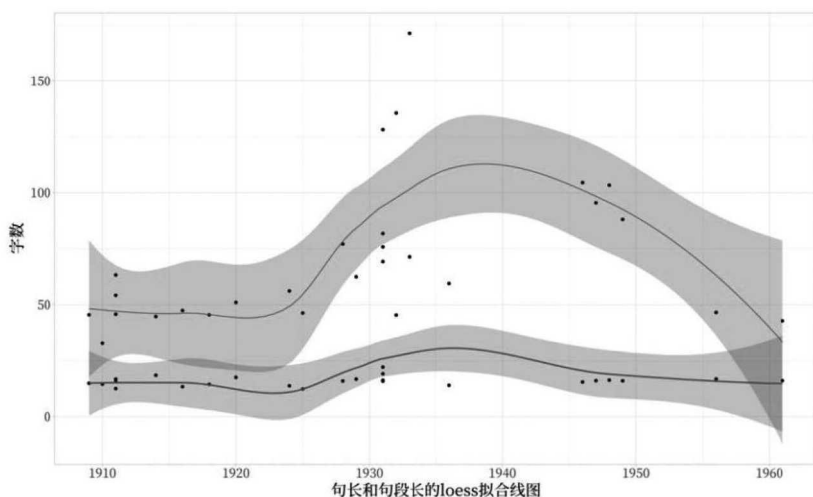


图2 句长和句段长分析(以年代为单位)

通过图2可以看出在1920年中期到1940年左右,谷崎润一郎的作品在句长和句段长上出现了较大幅度的变化,此后句段长一直保持平稳,而句长在1933年左右变长后一直保持比移居关西前高的水平。直至1960年,其在晚年时代又回归到了移居关西前的水平。使用 R 语言中的 `wilcox.test` 非参数检验方法对句长数据进行显著性检验,  $p=0.0014$ ,说明两者在统计学意义上有非常显著的差异。采用同样的方法对句段长也进行显著性检验,  $p=0.16$ ,说明移居关西前后不具备统计学意义上的显著性差异。图3则显示出了其作品中反映的句长和句段长变化。

通过图3可以看出,其文体在1928年写作《卍》时发生了较明显的变化,《卍》正如作品的名字一样,讲述的是两组男女之间倒错的男女关系,《卍》全文使用了关西方言,由柿内园子以第一人称的叙事角度展开,而且采用了口语表达,让小说情节本身具有了叙述者强烈的个人感情色彩。这样通过口语叙事的文体特征使得文中的句子也随之变长。1931年的《盲目物语》和1932年的《芦刈》都有句子长的特点,但前者的句段却没有明显增多,可以说明谷崎润一郎在1932年以后对文体的尝试更加大胆了,在新文体小说中极力减少了句号和顿号(日语中的顿号作用同中文中的逗号)的使用,且句长和句段长在《春琴抄》中达到了最高值,此后

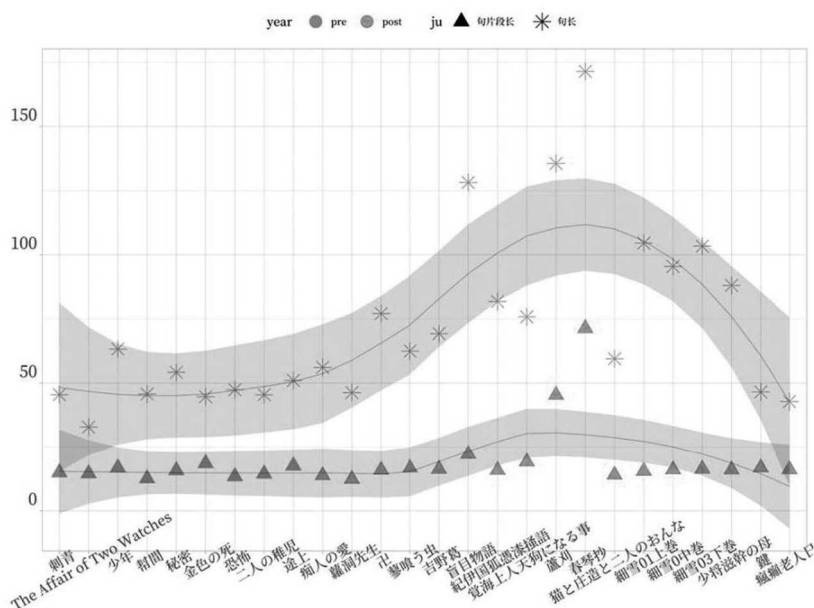


图3 句长和句段长(作品单位)

在句段方面一直保持着较为稳定的水平。而在句长方面,谷崎润一郎保持着自古典回归时期以来的长句特征,例如《细雪》的平均句长达到了100字符左右。而在其晚年时,句长和句段长又回归到了移居关西前的平均水平。谷崎润一郎在《文章读本》中将文章分为用语、调子、文体、体裁、品格、含蓄六个要素,在调子一节中他讲述了不加标点符号是为了追求文章语调的一贯性,模糊句子之间的界限(谷崎润一郎,1934:103-105),说明在1930年左右其对文体就有了明确的观点,并阐释了这一时期的写作极力节省标点的原因。

#### 4.3 关西方言使用频率

为了考查谷崎润一郎的作品在移居关西后是否呈现出与移居前不同的语言特点,根据先行研究的结论,假设其移居关西前后其文体特征不具有显著性差异( $H_0$ 假设),备择假设则为移居前后其文体特征具有显著性差异( $H_1$ 假设)。首先,本研究人工定义了12个具有代表性的日本方言语法词,10个方言人称词,随后计算文本的方言词的词频并进行标准化,将其分为移居关西前和移居关西后并绘制箱线图,由于卡方独立性检验显示数据不符合正态分布,因此本研究决定对数据进行分组,采用 `kruskal.test` 非参数检验显著性测试,最终绘制出的箱线图,如下图所示(橙色为移居关西前,蓝色为移居关西后,ns 表示不具有显著性,一个星号代表的是5%水平的显著性;两个星号代表的是1%水平的差异,意味着差异非常显著;三个星号代表0.1%水平的显著性,意味着差异极其显著)。

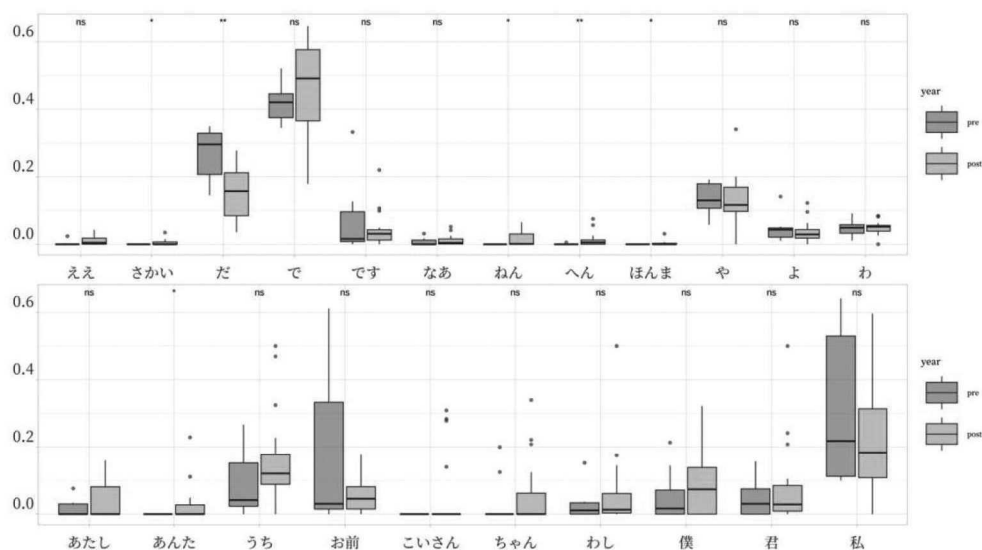


图4 关西方言使用频率(移居关西前后对比)

可以观察到上图的さかい( $p=0.03$ )、だ( $p=0.004$ )、ねん( $p=0.01$ )、へん( $p=0.006$ )、ほんま(0.01)前后出现了明显的差异,其中在だ的使用频率上前后具有非常显著的差异,前期だ的使用率显著高于后期。在へん(系关西地区方言,表示否定)的使用频率上前后具有极其显著的差异,后期关西方言へん的使用频率显著高于前期,这一数据也说明了谷崎润一郎在移居关西后形成了关西方言的文体风格,验证了既往的文学者观察到的文体现象。

在人称词的使用上,移居关西后あんた的使用频率显著比前期要高,あんた是日本关西地区女性对他人的称谓。其原因是在古典回归时期,谷崎润一郎刻画了许多生动的关西地区女性形象,例如《春琴抄》中的春琴、《细雪》中的四姐妹、《己》里的光子和柿内、《猫和庄造和两个女人》的品子和福子等,这些人物所说的都是关西地区方言,也从侧面说明了谷崎润一郎在移居关西后的作品中夹杂着许多关西女性的日常对话。

#### 4.4 基于多种文体数值的聚类分析

文本聚类是文本风格研究中的一种重要方法。聚类将文本集合分组成多个类,同一个簇中的小说内容具有较高的相似性,而不同簇中的小说内容差异较大(刘颖、肖天久,2014:145)。孙萌,刘晓芳(2022:153)使用词性占比、MVR、品词数、指示词、句长等指标对太宰治的26篇作品进行聚类分析,发现了《人间失格》在其作品群中的异质性。本研究采用了两种词汇丰富度、句长、句段长、部分关西方言使用频率作为文体特征的参数,使用R的heatmap.2函数进行将数据标准化并制成热图。如图5所示,X轴是本研究所选取的特征值,Y轴是作品名称,颜色越深则代表X轴的文体数值越能体现Y轴作品的特征,也可以说是Y轴作品在X轴的文体特征上更加突出。相近的Y轴作品则在文体上具有较高的相似性。

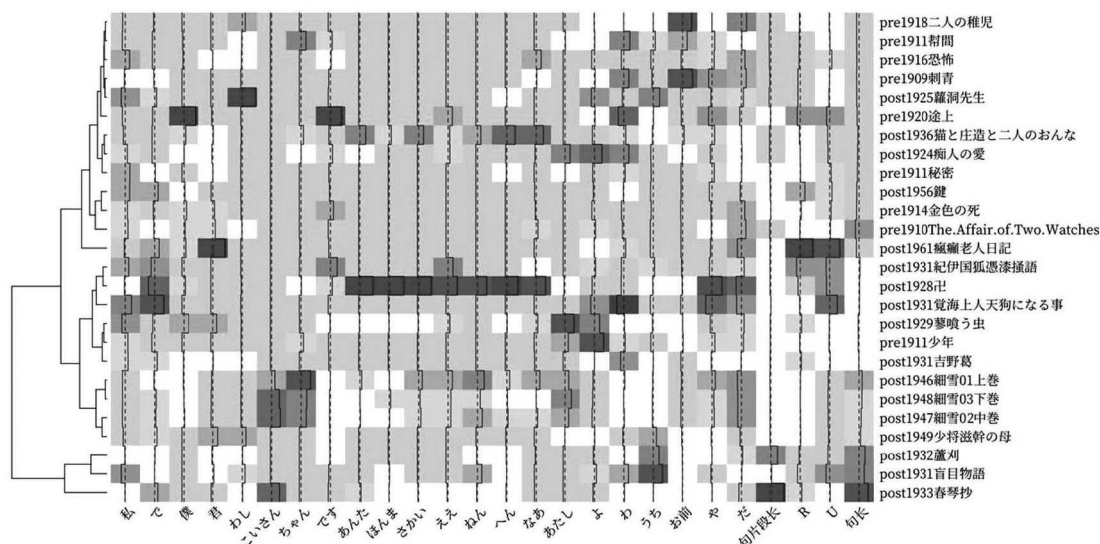


图5 基于多种文体参数的聚类分析热图(标准化数据)

从图5可知,谷崎润一郎的《细雪》上、中、下卷因为是相近年代创作的同一部作品,被聚类到相近的位置,说明该作品内部的文体风格极为相似。《细雪》是谷崎润一郎回归日本古典、翻译现代日语版《源氏物语》的直接产物。另外,其移居关西前的作品也被全部聚类到相近位置,体现出其移居关西前的作品之间具有较高的相似性。此外,在移居关西后的近20余年发表的大部分作品也被聚到了相近的簇,呈现出相似的文体特征。其中因为《卍》使用了大量的关西方言,所以对应的关西方言文体特征的红色程度较深。而从句长和句段长参数可以看出,从1931年到1947年,即《细雪》完成前的这一时期的作品,体现出句子长度整体偏长的文体特征。而在关西人称代词方面,《春琴抄》与《细雪》都大量使用了关西方言中“小妹”的说法,说明在此时谷崎润一郎着重塑造的是关西女性人物形象,体现出了这一时期的小说主要侧重于关西风土的描写,使作品体现出了方言语体特征。其晚期作品《癡癡老人日記》则体现出词汇丰富的特征,但因其远离了移居关西后的作品群,可以观察到其晚期作品又回归至早期创作风格的倾向。

## 5. 结语

本研究从计量文体学的视角出发,对谷崎润一郎移居关西前后的作品进行了计量分析,发现在1930—1940年(古典回归时期)谷崎润一郎的文体发生了较大的变化,这一时期其作品的词汇丰富度和句长、句段长数据都显示其文体风格都呈现出了与前后不同的特征。通过对数据的分析还可以发现,谷崎润一郎晚年的作品中也产生了词汇丰富度比早中期都高、句长和句段长文体特征回归到早期创作风格的现象。本研究还对前后期的部分关西方言用词进行了非参数检验,后期的部分关西方言用词明显多于前期,这与其在移居关西后塑造了大



量的关西女性的人物形象有着密切的关系。通过聚类热图可以发现,其早期作品群之间、古典回归时期的作品群之间均体现了较高的相似性,说明可以以移居关西为时间界限,对谷崎润一郎的文体风格时期进行划分。长篇小说《细雪》三部因为文体特征一致被聚类到一个小簇,同时与古典回归时期作品群体现出了近似的文体特征。通过对小说语料库进行文体的数理化测量,可以验证既往文学研究者的观点,谷崎润一郎在1923年移居关西后将近三十年的时间里,其文学创作与文体观念发生了变化,在1935年左右发生了最明显的文体变化。而后其以翻译《源氏物语》全文为契机,吸收了《源氏物语》中的古典日本文学的文体美学,使其文体观念有所转变,并对其新的创作产生了影响,唯美主义顶峰之作《细雪》也因此而诞生。最后在晚年时又有了回归早期恶魔主义的倾向。

本研究因语料不足,仅仅搜集到了谷崎润一郎的26部小说,考虑到题材不同,所以未将其随笔和散文纳入语料库中。在未来的研究中,笔者将扩大语料范围,采用新的文体指标进行多角度的测量,继续从实证角度验证既往研究者的假说。

#### 注释

- ①MVR系最早由樺島忠夫、寿岳章子(1965)提出的一种文体指标,值越高则代表文章越偏向样态描写,越低则代表文章越偏向动态描写。

#### 参考文献

- [1] Mendenhall T C. The characteristic curves of composition[J]. Science, 1887(214s): 237-246.
- [2] 安本美典. 宇治十帖的作者——文章心理学による作者推定[J]. 文学 語学, 1983, 4.
- [3] 安井寿枝. 谷崎潤一郎の作品における関西方言の変遷: 己, 蓼食ふ虫, 蘆刈, 春琴抄, 夏菊, 猫と庄造と二人のをんな, 細雪[J]. 甲南大學紀要. 文学編, 2008, 153: 35-55.
- [4] 波多野完治. 文章心理学[M]. 新編. 東京: 大日本図書株式会社, 1974.
- [5] 財津亘, 金明哲. 文末語の使用率に基づいた筆者識別[J]. 計量国語学, 2018, 31(6): 417-425.
- [6] 谷崎潤一郎. 饶舌录[M]. 汪正球, 译. 1版. 北京: 中国文联出版社, 2000.
- [7] 谷崎潤一郎全集: 卷 21[M]. 2版. 東京: 中央公論社, 1974.
- [8] 金明哲. 文章の執筆時期の推定——芥川龍之介の作品を例として[J]. 行動計量学, 2009, 36(2): 89-103.
- [9] 刘青梅. 试论谷崎潤一郎作品与《源氏物语》间的关联[J]. 日本问题研究, 2009, 23(2): 53-57.
- [10] 刘颖, 肖天久. 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), 2014, 29(5): 135-147+179.
- [11] 劉雪琴, 金明哲. 宇野浩二の病氣前後の文体変化に関する計量的分析[J]. 計量国語学, 2017, 31(2): 128-143.
- [12] 卢卫中, 夏云. 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), 2010, 33(1): 47-53.
- [13] 毛文伟. 日本的语料库文体学研究: 进展、问题及展望[J]. 外国语(上海外国语大学学报), 2021, 44(3): 82-90.

- [14] 隋桂岚, 孙利望, 李学勤. 语料库、统计学与文体分析[J]. 辽宁工程技术大学学报(社会科学版), 2003, 5(6): 81-84.
- [15] 孙萌, 刘晓芳. 语料库文体学视阈下日本近现代文学的异质性特征研究——以太宰治《人间失格》为例[J]. 解放军外国语学院学报, 2022, 45(4): 152-159.
- [16] 樋口耕一. フリーソフトウェア「KH Coder」による計量テキスト分析: 手軽なマウス操作による分析からプラグイン作成まで[J]. 研究報告人文学とコンピュータ(CH), 2015, 2015-CH-107(9): 1-2.
- [17] 叶渭渠. 谷崎润一郎传[M]. 北京: 新世界出版社, 2005.
- [18] 齋藤朗宏. 日本におけるテキストマイニングの応用[J]. The Society for Economic Studies The University of Kitakyushu Working Paper Series, 2012, 2011: 12.
- [19] 中尾桂子. 日本語テキストの計量的分析における統計手法[J]. 大妻国文, 2010, 41: 1-26.
- [20] 诸飞燕. 略论谷崎文学的“物语”性[J]. 日语学习与研究, 2012(4): 80-85.
- [21] 猪口洋志. 谷崎潤一郎の戦略: 古典回帰時代の語りと視点の変遷[J]. 國文學, 2022, 106: 79-97.

#### 作者信息

姓名: 王子睿

性别: 男

单位: 广西大学

研究方向: 计量文体学、语料库翻译学

电子邮箱: oushieiwzr@gmail.com

姓名: 刘善钰

性别: 女

单位: 广西大学

研究方向: 语料库语言学、应用语言学

电子邮箱: liushanyu408@gxu.edu.cn

## Research on the stylistic changes of Tanizaki Runichiro's novels

### ——An approach Based on Quantitative Stylistics

Guangxi University Wang Zirui, Liu Shanyu

**Abstract:** Based on the perspective of digital humanities, this study examines 26 literary works by Tanizaki Runichiro. It utilizes stylistic measurement indicators, such as lexical richness, and employs multivariate analysis methods, including correspondence analysis and cluster analysis, to analyze his literary works. The literary circle believes that Tanizaki

Runichiro's literary style concept has changed greatly since he moved to Kansai after the Great Kanto Earthquake. To verify the aforementioned hypothesis, this study aims to employ the techniques of quantitative stylistics and text mining to illustrate the three perspectives of language phenomena, stylistic characteristics, and literary evidence.

**Key words:** Quantitative Stylistics; *Tanizaki Runichiro*; Stylistic; Author discrimination