

《醒世姻缘传》作者归属的应用计量风格学研究

周 睿 马清华

【摘 要】明末清初白话小说《醒世姻缘传》的作者归属是文学史上的悬案。蒲松龄说最具影响。本文首次提出“应用计量风格学”范畴,并构建相应的研究范式,基于语言指纹(idiolect)理论,在《醒世姻缘传》和蒲松龄白话作品《聊斋俚曲集》中抽样,从对比分布、一般分布和散点分布三种统计路径考察了强程度标记在《醒世姻缘传》《聊斋俚曲集》中的使用特点。再扩大参证条件,增选学界指认的《醒世姻缘传》的另一可能作者丁耀亢的小说和剧作,以及现当代京味作家老舍与苏叔阳相应体裁的小说和剧作,以相同统计方法获取数据,并构建扩充的数据矩阵,由此剖析作者、时间、体裁对强程度标记分布的影响,由此作为甄别依据。最终判定《醒世姻缘传》的作者不是蒲松龄,也不是丁耀亢。论文以科学手段排除了学界对《醒世姻缘传》作者归属的主流推测,也为文学史上的作者归属研究提供了可靠的应用计量风格学范式,该范式对完善语言风格学理论同样有很好的学术价值。

【关键词】醒世姻缘传;作者归属;强程度标记;应用计量风格学;语言指纹

【作者简介】周睿,文学博士,上海对外经贸大学国际商务外语学院讲师,研究方向为语义学、语法学和计量风格学;马清华,南京大学文学院教授、博士生导师,主要研究方向为汉语语义学和语法学。

【原文出处】《明清小说研究》(南京),2022.4.266~287

【基金项目】本文受教育部人文社会科学基金项目“现代汉语语法复杂性计量研究”(项目编号18YJA740033)资助。

引言

《醒世姻缘传》(以下简称《醒》)是成书于明末清初的一部长篇白话小说,时间介于《金瓶梅》和《红楼梦》之间。确切创作时间不详,学界有崇祯说^①、顺治说^②、康熙中后期说^③、乾隆说^④等。兹不作推测。唯独乾隆说可以排除,因为雍正六年(1728)日本的《舶载书目》已著录此书。《醒》的语言风格比较一贯,口语色彩极浓,颇多方言成分,主要反映山东济南、历城、章丘一带方言特点^⑤,是后期近代汉语语法研究的重要语料。

《醒》未署作者真实姓名,仅题“西周生辑著”,作者身份仍是未解之谜^⑥。对《醒》作者的推断有多种,其中影响较大的是蒲松龄说,此说以胡适、孙楷第为

代表,为众多学者接受,一些较有影响的教科书、工具书也采用这一观点^⑦,理由有三:人物情节与《聊斋志异》中《江城》的悍妇故事颇为类似;部分方言土语和蒲松龄用方言色彩浓厚的白话戏曲《聊斋俚曲集》(以下简称《聊》)中的土语一致^⑧;作品中体现的地理、灾祥及创作时代与蒲松龄的情况吻合^⑨。也有学者反对蒲松龄说,理由有三:在墓表、碑阴以及行略等文献中均未被提到,也未见蒲松龄友人为其题跋;“蒲松龄说”滥觞于《梦阑琐笔》,其记载不可靠,鲍以文与蒲氏相距甚远,其言真实性差^⑩;作品中部分语言现象,如:差比句^⑪、处置式的使用^⑫,人称代词的表达^⑬以及虚词频率^⑭等与《聊》存在明显差异。可见,《醒》作者是否为蒲松龄存在争议。

在研究方法上,以往从语言角度考证《醒》的作者归属存在以下不足:(1)缺乏辩证。要么只抓住《醒》《聊》同的一面,无视另一面,由此认定作者系同一人,如:胡适^⑤、严云受^⑥;要么只抓住它们异的一面,无视另一面,由此认定作者不同,如:张树铮等^⑦、王洪秀^⑧、李焱等^⑨。能够同时关注同和异两面的学者,无一例外都着眼于异的一面,最终判定两部作品作者不同,如:刘玥^⑩、程志兵^⑪。(2)缺乏固定客观的标准。不知道相同程度多大为同一作者,相异程度多大为不同作者。

本文拟构建并利用可反映语言指纹的应用计量风格学方法,正视同异两面,并探究其关系。先在《醒》《聊》两作品间进行二元对比分析,依次从强程度标记的对比分布、一般分布和散点分布入手,展开统计与分析。然后在多个作者的多个作品间进行多样本数据的矩阵分析,在比例关系中评估体裁、年代等和作者身份对作品的影响力差幅,从而剥离非作者身份因素,对作者甄别做出客观有据、有充分说服力的判定。研究不仅为《醒》的作者归属提供了可信答案,而且为文学史上的作者甄别工作提供了适用的应用计量风格学范式。该范式对完善语言风格学理论也有很好的学术价值。

一、理论基础

Statistical Stylistics(“统计语体学、计量风格学”)这一术语至迟在上世纪中叶就已在西方学界出现^⑫。而将统计方法实际用于风格学研究则更早。国内将以语体为对象的“语体学”和以作家作品风格为对象的“风格学”明确区别开来。计量风格学认为,文本的语言特征表现了作者在写作活动中的言语特征,是其个人风格不自觉的深刻反映,并且这些特征可以通过数量特征来刻画^⑬。这些研究以往基本都旨在揭示作品本体特征,因此用的是归纳描写方法。不同的目的,选用不同的策略,否则无以确保有效性。为解决《醒》的作者归属问题,本文初步构建一套用于甄别作家身份的计量风格学研究范式,它实属应用计量风格学范畴,建立在如下认识论、方

法论和工具论基础之上。

(一)认识论基础:语言指纹理论

语言指纹(idiolect)由库尔撒德(Coulthard)在其经典论文《个人语言特点和作者身份认定》中提出,认为每一个个体的言说和写作,都有其独一无二的特点。语言指纹体现在常用词表上,不仅表现为每个人的词表所包含的语项不相一致,还表现为每个人对不同语项有着特殊的偏好^⑭。奥尔森(Olsson)进一步指出,语言指纹的“区别性”特征在应用层面更值得关注,并成功运用语言指纹解决了匿名信件的作者比较,论文抄袭者判别等问题^⑮。语言指纹理论已被广泛应用于司法领域,形成司法语言(forensic linguistics)这个交叉学科。这些都表明语言指纹理论的可靠性和有效性。通过语言分析、文本鉴别,确定文本作者身份的方法和思想已发展成校勘学(textual criticism)的重要部分。奥尔森说,文学领域也可以借鉴该原理解决作者身份问题(如:《圣经》作者身份的争议、莎士比亚著作权纠纷等)。但尽管如此,事实上似乎未见有人明确将“语言指纹”这一范畴沿用于文学研究。

(二)方法论基础:随机抽样的多样本统计、多角度分布分析和多重验证

具体技术路线是:

第一,采用随机抽样的多样本统计方法,而非穷尽统计方法,后者本质上是双样本对比。

第二,二元对比分析。在作者疑似相同的两类作品间展开对比分析。(1)对比分布。着眼于A、B两个作品在选词差异上的分布。获得有或无的对比性分布数据。(2)一般分布。先着眼于两部作品中特征词的词次以及词数的高低差异,获取一般分布数据,再采用T检验判断特征词在两部作品差异的显著度。弥补以往研究仅关注有和无的绝对数据这一缺陷,挖掘更多极有价值的相对数据。即便是绝对数据,也从多样本分布中挖掘出具有反映高度特征价值的极端对立分布和近极端对立分布(参下)。(3)散点分布。着眼于两个样本间特征词的概率分布相似

度,通过样本间的两两比较分析,计算出作品间的相似度,获得散点分布数据,并进一步构建散点分布图。分析散点分布时,采用T检验,判断特征词在作品间与作品内样本差异的显著度,由此判断不同作者作品和同一作者不同作品或同一作品前后的风格差幅。(4)数据互证。由一般分布的绝对数据得出的(近)极端对立分布结论和由一般分布的相对数据得出的结论之间,以及由一般分布数据得出的结论和由散点分布数据得出的结论之间,都须能构成互证。

第三,由于其他因素干扰,严格地讲,一般分布和散点分布只能刻画出作品风格间的差异程度,并不能确定作者身份。为评估体裁、年代等和作者身份的影响力差幅,从而剥离非作者身份因素,须将待甄别对象放置在扩充的矩阵关系中来判认。矩阵是在纵横两个方向上都具有等比例关系的数据构型。除待甄别对象外,矩阵中其他对象的参数都是确定的。纳入矩阵关系后,待甄别对象在哪个方向上适配或不适配该等比例关系,都可得出有关作者身份的结论。

这些技术手段和路线,相较于传统的单样本比较或两个作者之间的二元对比,更具科学性,它们充分发掘了数据间的系统逻辑关系,提高了推断准确性和结论可信度。

(三)工具论基础:同义特征词集的选取

解决作者归属问题的关键是从作品中提取出能反映风格的特征。关于《醒》作者归属问题,以往研究多基于部分语法现象^⑤、特殊句式^⑥、常见功能词^⑦以及实词频数^⑧等,这些特征会受文本内容和语言规范化的影响,难以很好地反映作者风格。

就词汇手段而言,什么样的词有条件用作判定文学作品作者归属的语言特征词?计量风格学常用高频词^⑨、标记词(主要是虚词)频率^⑩、词类总频率^⑪、词汇丰富度和词汇密度^⑫来描写文学作品的风格特征。但这些往往基于穷尽统计,且内在关联性不强,为提高便捷性、有效性,我们将强程度词作为判定文学作品作者归属的语言特征词。有五点理由:

第一,同义手段的选择,是一项修辞活动,表达同一语义时,不同作者使用词项的丰富度和频率也存在差异,所以同义词群中的词项选择及各词项频更能有效反映作者风格。强程度词是一个相同概念的词群。

第二,同义词群内的词项越多,选择余地就越大,其使用倾向就越容易反映个人风格。强程度概念词成员众多。现代汉语普通话里的强程度词就有不下221个^⑬。

第三,同义特征词中的词项越丰富,其多元的形成渠道和多种不同的形义特征,也给使用者的优化表达提供了较大的选择空间,能有效反映个人风格。众多强程度词在汉语中共存。首先是多种理据共存。由于词义发展的类推机制,强程度标记多来源于量大、质好、顶级和超常等意义的实词,具备这些意义的实词都有发展为强程度标记的可能^⑭。其次是单、双音节共存。受双音化影响,单音节强程度标记通过重叠,加缀或复合等方式形成双音节标记。再次是多种功能类别并存。强程度标记有副词、语气词、代词、唯补词等词性类别。

第四,非标记词复现率低,且易受作品内容影响。相比之下,语言标记复现率高,能深刻反映作者的潜在特征,因此近义词丰富的语言标记当然能更好地反映个人风格。强程度词修饰性质形容词或心理动词,其所表程度范畴跟频度、范围,都是广义情态范畴的次类,属广义情态标记^⑮。

第五,要能较大程度地体现文学趣味,更能反映作品语言风格,从而更有利于解决文学问题。强程度概念是可用于激情表达的活跃范畴。就整个词群而言,强程度词在艺术语言和口语体里有着高频分布,极高级程度副词(如“绝顶、绝伦、透顶、何其、何等、极端”等),往往出现于文艺语体中,语气类强程度多出现于口语体,口语体的强程度词密度最高,它们背后往往都有渲情表达的动机^⑯。它们用久了会失去锋芒,为代偿强程度的磨损,又不断有新的强程度词产生,所以更新极快,当代汉语的很

多新潮词语都出自强程度词。强程度词在词项数、更新速度上,都跟弱程度词极不均衡,后者要少得多、稳定得多。

二、二元对比分析

二元对比指的是直接在《醒》和《聊》两个作品之间展开的对比分析。

(一)总体设计

1. 抽样设计

本研究选取与底本差距较小的注本即华夏出版社1995年版《醒世姻缘传》和齐鲁书社2018年版《聊斋俚曲集》为统计对象,以最大程度地反映作品原貌。为提升数据客观性,反映作者特征恒定性,我们以“回”为单位随机取样,两个作品各取5个样本,每个样本2万字,共10个样本^③。若两部作品是同一作者,相似特征会不断复现,有一定的恒定性,若为不同作者,差异性将会在5个样本中体现出来。

2. 总体描述和路径设计

《醒》《聊》共10个样本中强程度标记十分丰富,有程度副词、语气副词、指示代词及唯补词等^④。程度副词是强程度词的典型,一般作状语,少数还可作补语。在《醒》《聊》10个样本中共出现了27种(52.94%)表强度的程度副词,如(《醒》《聊》后的数字表页码):

扭的生疼(《醒》757);好生不忍(《醒》526);那几日狠暖和(《醒》285);他已跑得老远(《醒》298);颇颇清秀(《醒》513);忒也莽撞(《聊》524);分外欢喜(《聊》60);可恶极,你才极可恶(《聊》465);娇儿一个最孤单(《聊》607);甚喜欢(《聊》610)

由确定或强调表强程度的语气副词,在10个样本中共出现6种(11.76%):“真、真是、真真的、真正、着实、煞实”。如:

皮囊生剥真难受(《聊》232);真是呆(《聊》43);老婆子真真的妙(《聊》7);真正蹊跷(《聊》57);着实后悔(《醒》183);煞实疼痛(《醒》20)

表示强程度的指示代词在10个样本中共出现9种(17.65%):“这么、那么、真么、这等、这般、这样、那

样”。如:

如今大爷这等严明(《醒》173);真么不凑巧(《聊》343);你这么乖(《聊》24);也不能那么热闹(《聊》52);病得这般沉重(《醒》334);这样惶惶(《醒》528);那样上等的好人(《醒》302)

“真么”只见于《聊》,它是“这么”的音变形式([tʂə]→[tʂən]),显示出同一较大方言区域内次方言间语音变化的不同^⑤。

唯补词只放在动词或形容词之后,充当补语以表示强程度。表强程度的唯补词在10个样本中共出现9种(17.65%):“X煞、X不过、X难当、X难堪、X之极、X极了、X非常、X得紧、X死”。如:

教人爱煞(《聊》271);气愤不过(《聊》76);我那侄儿苦难当(《聊》76);冻饿难堪(《聊》7);无赖之极(《聊》53);欢喜非常(《聊》350);那头更觉疼得紧(《醒》21);岂不羞死人(《醒》327)

二元对比分析依次从对比分布、一般分布、散点分布三个角度展开统计与分析。

(二)对比分布

对比分布着眼于A、B两个作品在选词差异上的分布,即A有B无、A无B有、AB皆有三类情形的分布。

《醒》《聊》共10个样本出现51种强程度词,其中《醒》样本中共出现33种(64.7%),《聊》样本共出现42种(82.4%)。如图1所示。《醒》有《聊》无的强程度词有9种(17.6%)(a)。《醒》无《聊》有的强程度词有18种(35.3%)(b)。《醒》《聊》皆有的强程度词共24种(47.1%),就这24词在两作品各5样本中有或无的分布格局而言,差异样本数为4的2词,为3的3词,为2

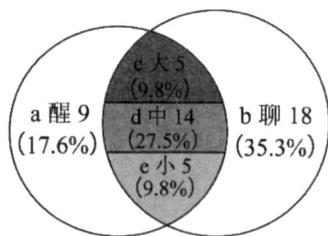


图1:《醒》《聊》强程度标记使用关系图

的7词,为1的7词,为0的5词,归纳起来看,差异样本数为3-4的大差异5个(9.8%)(c),为1-2的中差异14个(27.5%)(d),为0的小差异5个(9.8%)(e)。

由此可见,《醒》和《聊》选用的强程度标记52.9%不一致,《醒》《聊》均有的强程度标记中又有37.3%的强程度标记在分布上存在较大差异。可见,《醒》《聊》强程度词的选择上存在极大差异。

一般对比分布的“有”可以是单样本意义上的,但所有对比分布的“无”都是全样本意义上的。再根据在全样本中的覆盖率,提取两种最具特征的对立分布。一是极端对立分布,指A作品的5个样本中都有的词,在B的5个样本中都没有,其“有”“无”都是全样本意义上的。二是近极端对立分布,指A的4个样本中都有的词,在B的5个样本中都没有。其“有”是4样本意义上的,“无”都是全样本意义上的。《醒》《聊》的极端对立分布是《聊》的“忒也”(19次),近极端对立分布是《聊》的“煞”(7次)。

(三)一般分布

把强程度标记在各个样本中出现的词次和词数叫做一般分布。表1展示了《醒》和《聊》各个样本中强程度标记出现的词次和词数。对《醒》和《聊》10个样本中出现的强程度标记进行穷尽搜集与统计,发现两部作品在强程度标记的选用种类和使用频率上均存在较大差异(见表1)。为了判断强程度标记在两部作品中词次及词数的差异是否显著,对其进行T检验,主要依据T分布理论推断差异发生的概率,进而比较均值的差异是否显著,显著性P值作为主要的评判指标,若 $P < 0.05$ 则两个均值间存在显著性差异,若 $P < 0.01$ 则两个均值间存在极其显著差异,否则无明显差异^⑩。

由表1可知,《醒》每个样本中强程度标记平均出现词次为48.8,词数为16.2,《聊》中平均出现的词次为90.6,词数为24,《聊》样本中强程度词的词次和词数均显著高于《醒》(P均小于0.05),可见《聊》相对于《醒》强程度标记的种类更丰富,数量更多,且差异显著。

(四)散点分布

通过对两部作品10个样本的观察和统计,发现两部作品中强程度标记在对比分布和一般分布上存在显著差异,为验证该结论,本节通过强程度标记的散点分布,进一步探究《醒》《聊》间差异的显著度。

1. 相似度计算方法

强程度标记在文本中的分布可以理解为概率分布。本文引入JSD(Jensen—Shannon Divergence)^⑪来度量两个概率分布间的差异,JSD也称JS散度,定义如下:

$$JSD(P \parallel Q) = \frac{1}{2} D(P \parallel M) + \frac{1}{2} D(Q \parallel M) \text{ where } M = \frac{1}{2}(P+Q)$$

$$D(P \parallel Q) = \sum P(x) \log \frac{P(x)}{Q(x)}$$

$P(x)$ 和 $Q(x)$ 分别指待对比的两个样本中各强程度标记的概率分布,为了解决样本中零概率问题,本文使用拉普拉斯平滑(Laplace smoothing),即用加1的方法估计没有出现过的观测次数,具体方法是分子加1,分母加观测值总数^⑫。如在计算《醒》样本1和《醒》样本2间的JS散度时, $P(\text{极})$ 为强程度标记“极”在《醒》样本1中出现的次数(8+1)除以《醒》样本1中所有强程度标记出现的次数(64+1×51), $Q(\text{极})$ 为强程度标记“极”在《醒》样本2中出现的次数(7+1)除以《醒》样本2中所有强程度标记出现的次数

表1:《聊》和《醒》样本中强程度词的词次和词数

	样本1	样本2	样本3	样本4	样本5	均值	T检验(P值)
《醒》词次	64	37	54	28	61	48.8	0.002269
《聊》词次	101	88	102	75	77	90.6	
《醒》词数	21	14	19	10	17	16.2	0.013120
《聊》词数	29	22	22	21	26	24	

(37+1×51),同理可计算出待比较样本中其他强程度标记的概率,按照上述公式即可计算出两个样本间的JS散度。

JSD是一种通过信息值的密度分布测量其相似度的方法。JSD可以衡量两个随机分布之间的距离,当两个随机分布相同时,它们的JSD为零,当两个随机分布的差别增大时,它们的JSD也会增大,JSD越小,两个分布间的距离越小,说明两个文本间的相似度越高。JSD的值域是[0,1],相同则是0,相反为1。

根据语言风格特征判断作者归属并非为一个黑白分明,非是即否的问题,即{0,1},而应是一个连续的集合[0,1]。语言作为一个交际工具,作者个人风格会受时间、地点等多种因素的影响,因此我们不能仅仅依据某几个语言现象的异同就判断作者归属,而应根据语义范畴成系统地综合考虑,并将作者自身风格变化纳入考虑范围之内,同一位作者的风格变化应处于一个区间范围内,超出这个区间则两部作品应并非出自同一人之手。

2.《醒》《聊》样本间JSD统计

首先计算《醒》5个样本两两间的JSD(共10组数据),同理可计算出《聊》5个样本两两间的JSD(共10组数据),再计算《聊》5个样本和《醒》5个样本之间的JSD(25组数据),数据汇总于表2,方框内部为作品间的比较,外部为同一作品内部的比较。

由表2可知,同一作品内JS散度的范围是

[0.01209,0.03066],均值为0.02255;《醒》内均值为0.02228,《聊》内均值为0.02281;《醒》和《聊》作品间JS散度的范围是[0.02905,0.06041],均值为0.04251。

本文进一步对JSD均值进行T检验,《醒》内和《聊》内的平均JSD差异不显著($p=0.7913>0.05$),作品间和作品内的平均JSD差异极其显著($p=8.854\times 10^{-13}<0.01$),作品间的JSD显著高于作品内的JSD。本文将表2可视化如图2。

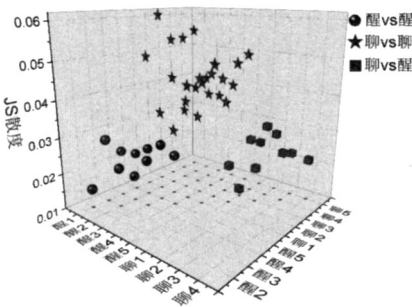


图2:样本间JSD分布图

可直观看出,《醒》和《聊》样本间的JSD,基本高于《醒》和《聊》样本内的JSD。由表2和图2结果可知,《醒》与《聊》强程度标记的概率分布差异相当大。

《醒》和《聊》虽均为明清之际以山东方言创作的白话文学作品,但二者体裁不同,分别是小说和戏曲。小说以叙述为主诉诸读者智力和想象,戏曲以直观展示为主诉诸观众视觉和听觉^④。

至此,对比分布、一般分布、散点分布三个角度的数据与分析,都得出同一结论:《醒》《聊》强程度词的使用风格差异极大。但悬殊差异形成的原因不限

表2:《醒》和《聊》样本间JSD

	醒2	醒3	醒4	醒5	聊1	聊2	聊3	聊4	聊5
醒1	0.01527	0.02802	0.02323	0.02099	0.04900	0.06041	0.05296	0.05300	0.05498
	醒2	0.02102	0.01707	0.02378	0.03325	0.04270	0.03183	0.03797	0.04031
		醒3	0.02302	0.02615	0.02905	0.03650	0.0307S	0.03678	0.03531
			醒4	0.02427	0.04259	0.04370	0.04438	0.04212	0.03938
				醒5	0.04427	0.04824	0.03688	0.04729	0.04917
					聊1	0.02091	0.01209	0.02565	0.02340
						聊2	0.01986	0.03066	0.02713
							聊3	0.02412	0.02253
								聊4	0.02178

于作者因素。严格地说,在没有比较出作者、体裁、时代差异对作品风格影响力的相对尺度之前,结论恐怕只能到此为止。要明确有据地甄别作者身份,须拓宽观察范围,分析论证作者因素和非作者因素对作品风格的影响力差幅,以便从复杂评估中剔除非作者因素的影响。下面展开多元对比分析。

三、多元对比分析

多元对比分析包括简单矩阵分析和扩充的矩阵分析。

(一)简单矩阵分析

《醒》《聊》方言背景相近,但体裁分别为小说和俚曲,为扩大参证条件,考察作者、体裁因素对强程度标记使用风格的影响力孰大孰小,可以把对比分析纳入到以作者因素为横、体裁因素为纵的矩阵中来。根据横纵两个维度上的比例关系来比较多要素对作品强程度词使用风格的影响力。

为排除作者方言背景、年代等其他因素的影响,并且为提升真实性、客观性和可证性,我们控制变量,类推《醒》《聊》的体裁差异和方言一致、年代相仿的背景特征。

优先从现当代作家中选取方言背景一致、年代相仿的两个典型“京味”双料作家(兼有小说和话剧作品)老舍和苏叔阳^⑤的作品构建矩阵,从老舍的小说《骆驼祥子》、话剧《龙须沟》及《茶馆》,苏叔阳的小说《故土》、话剧《左邻右舍》及《丹心谱》^⑥中抽样。每人每种体裁(小说/话剧)随机抽取5个样本,每个样本1万字,共20个样本。统计各样本中强程度标记的散点分布。从矩阵数据中,对比分析作者、体裁因素对强程度词使用风格影响力大小,然后将《醒》《聊》的数据关系跟矩阵分析结论进行比对,初步判断出

作者身份。

图3中,圆箭头abcd分别表示同一作者同一体裁作品5个样本两两间的JSD,每个箭头包含10组数据;双箭头①②,③④分别表示不同作者不同体裁,同一作者不同体裁作品间的JSD,①表示老舍小说5个样本和苏叔阳话剧5个样本两两间的JSD,②表示苏叔阳小说和老舍话剧间的JSD,③表示老舍小说和话剧间的JSD,④表示苏叔阳小说和话剧间的JSD;每部作品抽取5个样本,双箭头则表示两部作品样本间的JSD,共25(5*5)组数据。圆箭头和双箭头共140组数据。

表3中,分别计算不同对比类型的JSD均值,并进行T检验;通过对比6组JSD均值及其差异的显著性,分析作品体裁和作者身份对JSD的影响,进而验证运用强程度标记分布相似度判别作者身份的可行性。

由表3可知,作品间的平均JSD均高于作品内的平均JSD;不同体裁不同作者作品间JSD差异的显著度高于同一作者不同体裁作品间JSD差异的显著度,该结论得到了图3矩阵中abc和cda的双重验证。因此,这证实了基于强程度标记分布相似度判别作者身份的方法具有可行性,该方法可用于《醒》

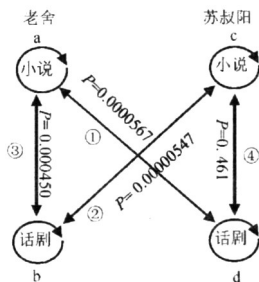


图3:老舍、苏叔阳作品的位置关系

表3:老舍和苏叔阳作品的JSD均值和T检验

类型	对象	JSD均值	T检验(P值)	类型	对象	JSD均值	T检验(P值)
同一作者	a~b	0.02072	0.000450	不同作者	a~d	0.02260	0.0000567
	③	0.03121			①	0.03506	
不同体裁	c~d	0.02381	0.461371	不同体裁	b~c	0.01915	0.00000547
	④	0.02581			②	0.03001	

和《聊》作者身份的判断。

利用此方法,根据《醒》《聊》间和《醒》《聊》内JSD均值差异极其显著($P=0.0000000000008854$),将《醒》《聊》的关系与图3中老舍和苏叔阳作品的P值矩阵进行比对,发现《醒》和《聊》应处于对角线位置(图4),说明这两部不同体裁作品的作者不同。

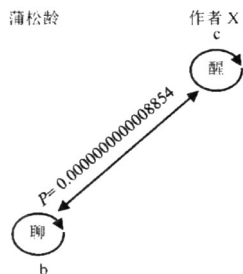


图4:《醒》和《聊》的位置关系

上述结果证实,《醒》《聊》间JSD的差异显著超出同一作者不同体裁作品间的差异,故《醒》和《聊》的作者并非同一人。从p值的数据来看,这两者不仅不是同一人,而且二者语言风格差异远大于老舍和苏叔阳之间的差异,这种差异可能源自时间跨度等更多的因素。总之,《醒》的作者并非蒲松龄。

(二)扩充的矩阵分析

1. 古今双矩阵的构建与分析

尽管现当代双料作家老舍与苏叔阳身份、作品、时代都真实可证,且矩阵只是建立在比例关系而非绝对关系上,但其当今语体毕竟都经过了西方化的大洗礼,跟古代语体有很大不同。为逼近古作品风格,进一步扩大参证条件,除《醒》《聊》外,另选取与之时代相仿的作家丁耀亢的作品,构建古作品矩阵。依据保罗比例式($a:b=c:x$)分析古矩阵数据,其作者甄别结论跟今矩阵分析结论、双矩阵分析结论相互佐证和检验,提升结论可信度和作者甄别范式的科学性和普适性。

构建古矩阵时,之所以选取丁耀亢的作品,除因为与《醒》《聊》时代相仿外,还有以下四点理由:(1)《醒》的作者归属,学界除认是淄博蒲松龄外,还有诸城丁耀亢^⑩、兖州贾岛西^⑪、章丘人士^⑫、河南人士^⑬、陕西人士^⑭、集体创作^⑮各说。丁是其中之一。(2)丁是兼有

小说和戏剧作品的双料作家,其身份明确、作品体裁吻合、规模足量、因而满足本文计量分析和增选为参证材料条件。(3)丁的方言背景吻合。尽管学界对《醒》的作者身份推测不一,但所判都属今北方方言区人,尤其是中原官话区(包括山东、河南)人。其次,不少学者都指认,方言文学色彩浓厚、有极高的口语化程度的《醒》,用的是明清时期北方话系统的山东方言^⑯。对此尚未见异议。持河南人说的在承认《醒》“有大量山东方言之外,亦尚有河南方言”^⑰,只是增加了地名证据。持陕西人说的并非基于方言证据,而是就“西周生”姓名所作的猜测。(4)虽然以下作者与《醒》小说及蒲松龄大致同时或相差不远、且兼有小说和剧作或曲作,但方言背景相距较远,故不纳入计量考察范围。如,江西饶州府人[今赣方言区]邓志谟^⑱、苏州吴县人[今吴方言区]袁于令和苏州府长洲县人[今吴方言区]冯梦龙、陕西人[今西北官话区]罗懋登^⑲。

按作者和体裁对作品进行解析,《醒》为“X说”,《聊》为“蒲曲”、丁耀亢小说为“丁说”、丁耀亢杂剧为“丁剧”,它们构成古矩阵。老舍小说为“老说”、老舍话剧为“老剧”,苏叔阳小说为“苏说”、苏叔阳话剧为“苏剧”,它们构成今矩阵。两矩阵各自的方言、时代背景相同。矩阵除首列作者关系不确定外,其他各列作者相同,首行的体裁相同或相近。第二行的语体风格古今有别,蒲曲和丁剧也有别,但它们的最大公约数是“戏”。古今两矩阵既可互相佐证和检验,也可连成更大的一个矩阵。

采用相同方法,抽样统计丁的相应作品。抽样时,从丁耀亢小说(《续金瓶梅》)和杂剧(《西湖扇》《化人游》《赤松游》《蚰蛇胆》)^⑳中各随机抽取5个样本,



图5:与《醒》《聊》相关的古今双体裁矩阵图

每个样本的字数为1万字。把所获得的丁的统计数据跟已经获得的其他各端作品的相应统计数据(均各由5个样本统计所得)归入古今数据矩阵,见表4。表4数据中仅列出对比分布中的(近)极端对立分布和散点分布数据。不仅可以通过相对尺度对比作者、体裁、时间跨度因素对作品强程度词使用风格的影响力大小,还可通过古今矩阵的对比,观测年代因素、方言背景因素对作品强程度词使用风格的影响力大小。由此为《醒》的作者甄别提供更强证据,收缩怀疑范围。

从扩充的数据矩阵中可以看出,作者、体裁、时间跨度都是影响强程度标记的对立分布和散点分布差异的重要因素,呈正相关。证据是:(1)作者相同但体裁不同(a—c),或体裁相同但作者不同(e—f),均未出现极端对立分布,且散点分布的差异显著度相对较小^⑨。(2)当体裁相似,但作者不同且跨年幅度极大时,出现极端对立分布,且散点分布的差异显著度大幅提升(d)。(3)当作者相同,但体裁不同且跨年幅度较大时,仅可出现近极端对立分布,且散点分布的差异显著度相对并不算高(b)。(4)作者和体裁都不同且跨年幅度极大时,极端对立分布和近极端对立分布的出现几率大增,且散点分布的差异显著度均有提

升(g—i)。(5)对相同作者而言,时间跨度越短,即使体裁不同,强程度标记散点分布的差异也会相对越不显著(c)。(6)若忽略时间跨度影响,近代作品强程度标记散点分布的差异显著度,相较于今人作品(如老、苏作品e、f、h、i)而言,大幅提升(如蒲、丁作品d、g),可能是因为近代文学文白并存、体裁差异(如剧、曲差异)和韵律因素影响大、规范化水平不高而致方言影响大等,因而多样化程度大增。从现代山东方言看,蒲的临淄话属山东方言西齐片,丁的诸城话属东潍片。

在各影响因素中,作者因素对强程度词使用风格的影响往往大于体裁因素对它的影响,无论古今。就差异显著度而言,在古代,作者不同、体裁相近的d远大于作者相同、体裁不同的a;在现当代,作者不同、体裁相近的f远大于作者相同、体裁不同的c。不过,同样是在现当代,作者不同、体裁相近的e和作者相同、体裁不同的b却只是同一数量级上的差异,差异小到了可视为误差的程度。之所以如此,是因为现代语体经过了西方的大洗礼(参前)、规范化水平提升,因而作者不同、体裁不同的作品公约数总体上比古代的大,换言之,现代不同体裁间差异往往小于古代不同体裁间的风格差异,如 $a > \{b, c\}$, $g > \{h, i\}$ 。

表4:作品风格主要影响参数及强程度概念词数量特征对比表

作者	体裁		时代	发表年	均跨年	对立分布(5样本)		散点分布(5样本)
						极端对立分布	近极端对立分布	P值
同	异	a 丁说~丁剧	古	1662~1653	9	—	—	0.0000299[2.99×10 ⁻⁵]
		b 老说~老剧	今	1936~(1950、56)	17	—	[老说]颇7	0.000450014[4.50×10 ⁻⁴]
		c 苏说~苏剧		1983~(1978、80)	4	—	—	0.461370768
异	似	d 蒲曲~丁剧	古	(1676-1711) ^⑨ ~1653	40.5	[蒲]忒也19	—	0.0000000000590[5.90×10 ⁻¹¹]
		e 老说~苏说	今	1936~1983	47	—	—	0.000762131[7.62×10 ⁻⁴]
		f 老剧~苏剧		(1950、56)~(1978、80)	24.5	—	[老]顶12	0.002093892[2.09×10 ⁻³]
异	异	g 丁说~蒲曲	古	1662~(1676-1711)	31.5	[蒲]忒也19	[丁]十分13;[蒲]得紧6,煞7	0.0000000000891[8.91×10 ⁻¹¹]
		h 老说~苏剧	今	1936~(1978、80)	43	[老]老10	—	0.0000567[5.67×10 ⁻⁵]
		i 老剧~苏说		(1950、56)~1983	30	—	[老]顶12	0.000005472[5.47×10 ⁻⁶]
?	同	j X说~丁说	古	?~1662	?	—	[X]稀5,这般8	0.000000669[6.69×10 ⁻⁷]
		k X说~蒲曲		?~(1676-1711)	?	[蒲]忒也19	[蒲]煞7	0.000000000008854[8.85×10 ⁻¹³]
		l X说~丁剧		?~1653	?	—	[X]稀5,这般8	0.00000000436[4.36×10 ⁻⁹]

i}。古代戏剧/戏曲体裁间的风格差异也往往大于现代相同体裁间差异,如 $d > \{e, f\}$ 。

作者因素是影响强程度词使用风格的主要因素这一结论,跟二元对比分析所得结论结合起来,进一步加强了简单矩阵分析中论述“X不是蒲松龄”的证据。

2. 古矩阵分析

依据保罗比例式($a:b=c:x$)分析古矩阵数据。若“X说一蒲曲”呈常态分布,则X可能是蒲,若呈反常分布,则X不可能是蒲。同样,若“X说一丁说”呈常态分布,则X可能是丁,若呈反常分布,则X不可能是丁。反常分布越多,可能性越低。

通过古矩阵分析,可以再次得出结论,《醒》的作者(X)不是蒲松龄,理由有四:(1)k中同时出现了极端对立分布和近极端对立分布,且散点分布的差异显著度极大。其分布特征更接近作者和体裁都不同且跨年幅度极大的情形(g—i)。(2)若X是蒲松龄,则作者相同体裁不同的k,散点分布的差异显著度不应反而高于作者和体裁都不同且跨年幅度极大的g,也不应出现极端对立分布,作者相同但体裁不同时,强程度标记一般不出现极端对立分布(a—c)。(3)若X是蒲松龄,则无法解释为什么作者、体裁都不同的l,其散点分布的差异显著度比作者关系格局不变但体裁更接近的d有了明显下降,且对立分布的极端程度降级。(4)就k和l的散点分布和极端对立分布的数据而言,蒲松龄的可能性甚至比丁耀亢更小。

还可以进一步得出结论,《醒》的作者(X)也不是丁耀亢,理由三:(1)若X是丁耀亢,则作者和体裁都相同的j,散点分布的差异显著度不应反而高于a(丁自己的小说和剧作)的情形,也不应出现近极端对立分布。据对金庸、古龙^①以及巴金^②多位作家的作品研究发现,同一作者不同时期相同体裁作品的语言风格差异不显著。(2)若X是丁耀亢,则无法解释为什么k散点分布的差异显著度比相同关系格局的g有了明显提升。(3)若X是丁耀亢,则无法解释为什么l散点分布的差异显著度比相同关系格局的a有了大幅提升。三方面时呈现出的反常分布,强化了对

《醒》作者不是丁耀亢的判断。

因此,蒲松龄和丁耀亢都不是《醒》的真实作者,而光从强程度标记分布的计量分析看,蒲松龄说比丁耀亢说更加离谱。由于可据观察的文本限制和工作量限制,一时还无法正面寻认出《醒》的真实作者,如果作者只有唯独这一部作品存世,则更可能永无答案,但本文已为《醒》的作者甄别拨开了迷雾,并提供了可资利用的研究范式。

结语

本文基于我们所构建的应用计量风格学范式,探究《醒》的作者归属问题,分别从对比分布、一般分布和散点分布三种统计路径,在《醒》《聊》间展开二元对比分析,反复证明《醒》《聊》的强程度词使用风格差异悬殊。然后展开多元对比分析,依次利用简单矩阵分析和扩充的矩阵分析,逐级扩大参证条件,观察作者、体裁、时间跨度等多因素对作品风格的影响力大小,据此获得的相对尺度来甄别作者身份。构建古今两个双体裁矩阵,古矩阵由《醒》《聊》和《醒》的另一个可能作者丁耀亢的小说和剧作构成,今矩阵由现当代两位京味作家老舍、苏叔阳的小说、话剧构成。双重验证了此前的结论和该方法的有效性。分析发现,作者、体裁、时间跨度都是影响强程度标记的对立分布和散点分布差异的重要因素,呈正相关。在各影响因素中,作者因素对强程度词使用风格的影响往往大于体裁因素对它的影响。根据对这些因素的影响分析结果,在更加充分的证据基础上判定,《醒》的作者不是蒲松龄,也不是丁耀亢。蒲松龄的可能性甚至比丁耀亢更小。多角度、多作者、跨时代的考察和重复验证,得出同一结论,这实际也自证了计量方法的科学性和一定程度的普适性。本文不仅排除了学界在《醒》的作者归属问题上的主流推测,而且在方法论上为解决文学史上作者归属疑案提供了一套客观可信的语言风格学的计量分析范式,甚至可为司法领域判断作品归属的语言指纹分析提供参考。利用可以反映语言指纹特征的计量风格学方法,提升了研究手段的科学性,这种研

究范式对完善语言风格学理论也有很好的学术价值。

注释:

①王守义:《〈醒世姻缘传〉的成书年代》,《光明日报》1961年5月28日。

②④徐复岭:《醒世姻缘传作者和语言考论》,济南:齐鲁书社,1993年,第21、32页。

③孙楷第:《一封考证〈醒世姻缘传〉信》,[清]西周生著,童万周校注《醒世姻缘传》,郑州:中州古籍出版社,1997年。

④林辰:《论明末清初小说的历史地位》,《社会科学辑刊》1982年第5期。

⑤③冯春田:《明清山东方言语法研究》,济南:山东教育出版社,2012年,第4、5页。

⑥《醒》作者身份之谜的论述,较具代表性的成果如:徐复岭:《醒世姻缘传作者和语言考论》,济南:齐鲁书社,1993年,第7页;罗敬之:《蒲松龄年谱》,台北:国立编译馆,2000年,第276页;段江丽:《〈醒世姻缘传〉研究》,长沙:岳麓书社,2003年,第13-25页。

⑦如:游国恩:《中国文学史》,人民文学出版社,1963年;谭正璧:《中国小说发达史》,光明书局,1935年;柳存仁:《中国小说史》,香港大众书局;北京大学中文系:《中国小说史稿》,人民文学出版社,1973年;《辞海》,上海辞书出版社,1979年。

⑧⑮胡适:《〈醒世姻缘传〉考证》,胡适《胡适古典文学研究论集》,上海:上海古籍出版社,1988年,第1027、1028页。

⑨孙楷第:《一封考证〈醒世姻缘传〉信》。[清]西周生著,童万周校注:《醒世姻缘传》,郑州:中州古籍出版社,1997年。

⑩路大荒:《蒲松龄年谱》,济南:齐鲁书社,1980年,第94-103页。

⑪⑩李焱、池静莲:《从差比句的比较看〈醒世姻缘传〉的作者》,《龙岩师专学报》2003年第4期。

⑫⑳㉑刘玥:《从处置式比较看〈醒世姻缘传〉非蒲松龄所作》,《蒲松龄研究》2010年第1期。

⑬㉒程志兵:《〈聊斋俚曲集〉与〈醒世姻缘传〉人称代词的差异——兼谈两书的作者问题》,《山东理工大学学报》(社会科学版)2009年第3期。

⑭㉓周沫、罗季:《基于支持向量机的〈醒世姻缘传〉作者归属研究》,《龙岩学院学报》2017年第4期。

⑯严云受:《〈醒世姻缘传〉作者问题补考》,《明清小说研究》1997年第3期。

⑰⑳张树铮、孙韵珩:《〈醒世姻缘传〉和〈聊斋俚曲集〉语法特点的差异》,《古汉语研究》2005年第4期。

㉑王洪秀:《〈醒世姻缘传〉与〈聊斋俚曲集〉限定性范围副词比较研究》,《黔南民族师范学院学报》2018年第5期。

㉒Bailey, R. W. & L. Dolezel. An Annotated Bibliography of Statistical Stylistics. Dept. of Slavic Languages and Literatures, U. of Michigan, Ann Arbor, Mich., 1968. 2.

㉓Calix, K., M. Connors, D. Levy, H. Manzar, G. McCabe & S. Westcott. Stylometry for E-Mail Author Identification and Authentication, Proceedings of CSIS Research Day, Pace University, 2008.

㉔Coulthard, M. Author Identification, Idiolect, and Linguistic Uniqueness, Applied Linguistics, Vol. 25(4), 2004.

㉕[英]约翰·奥尔森(J. Olsson):《司法语言学》(第2版),中译本,北京:中国人民公安大学出版社,2015年,第13页。

㉖㉗㉘刘颖、肖天久:《金庸与古龙小说计量风格学研究》,《清华大学学报》(哲学社会科学版)2014年第5期。

㉙施建军:《基于支持向量机技术的〈红楼梦〉作者研究》,《红楼梦学刊》2011年第5期。

㉚于涛、唐美华:《汉语小说中的词汇特征对比研究》,《重庆交通大学》(社会科学版)2017年第4期。

㉛周睿:《强程度范畴的语法语义研究》,南京大学博士论文,2021年,第220页。

㉜马清华:《文化语义学》,南昌:江西人民出版社,2000年,第110-113页。

㉝马清华:《汉语情态统辖结构的整合与变异》,《山西大学学报》(哲学社会科学版)2017年第1期。

㉞陈羽:《汉语程度标记的语体语法研究》,南京:南京大学博士毕业论文,2017年,第180页。

㉟抽样情况如下:《醒》的样本构成:[样本1]第3,39,91(部分)回;[样本2]第20,21,22(部分)回;[样本3]第34,35,36(部分)回;[样本4]第41,42,43(部分)回;[样本5]第61,62,63(部分)回。《聊》的样本构成:[样本1]《墙头记》第1回,《姑妇曲》第3回,《钗钏咒》第11回,《富贵神仙》第12回,《磨难曲》第5回,《增补幸云曲》第7回(部分);[样本2]《墙头记》第4回,《姑妇曲》第2回,《慈悲曲》第3回,《翻魔殃》第1回,《钗钏咒》第5回(部分);[样本3]《翻魔殃》第7回,《寒森曲》第7回,《蓬莱宴》第7回,《穷汉词》,《钗钏咒》第6,7回,《富贵神仙》第2回,《增补幸云曲》第25回(部分);[样本4]《增补幸云曲》

第10,11,12回,《磨难曲》第2,3,4回,《富贵神仙》第33,《攘妒咒》第19回(部分),总计20000字,[样本5]《磨难曲》第15,16,17回,《富贵神仙》第9回,《蓬莱宴》第4回,第《攘妒咒》3回(部分)。

③原始数据如下:《醒》X不过0,0,3,0,1(5个样本中的频次,下同);X得紧2,1,1,0,0;X死4,0,0,0,0;大6,8,5,1,3;过0,0,1,1,0;好0,0,3,0,0;好不1,0,0,2,2;好生1,0,0,1,1;狠0,0,1,1,0;狠命1,0,0,0,0;极8,7,10,5,4;极其1,0,1,1,0;绝1,0,0,2,0;老2,1,1,0,0;那样0,0,3,0,0;颇颇0,0,0,0,1;煞实1,0,0,0,0;甚6,4,4,2,4;甚是9,4,3,4,16;生1,1,0,0,0;十分0,1,3,2,0;太0,1,1,2,1;忒0,2,0,0,0;万分1,0,1,0,0;稀1,1,0,2,1;异常0,0,0,0,1;这般3,1,3,1,0;这等5,4,0,1,2;这样0,0,4,0,4;着实1,1,3,0,4;真0,0,0,0,1;至4,0,0,0,3;最5,0,3,0,2。《聊》X不过0,1,0,0,0;X得紧1,0,2,1,2;X非常0,0,1,0,2;X极了4,2,2,0,0;X难当1,1,2,0,0;X难堪:1,0,0,0,0;X死3,1,0,0,2;X之极:2,0,1,1,0;大16,9,16,3,5;分外1,0,1,0,0;怪1,0,0,0,2;过1,0,0,0,1;好9,7,9,16,13;好不2,2,3,2,1;极21,30,24,16,11;极其1,0,0,0,0;绝1,3,0,1,0;老大3,0,0,2,0;那么1,0,0,0,0;那样1,1,0,0,1;颇0,0,0,1,0;煞2,2,1,0,2;甚12,2,7,7,2;甚是1,0,3,0,1;太5,7,6,4,13;忒0,0,2,4,2;忒也7,3,2,2,5;稀0,0,0,0,1;异常0,0,1,0,2;这般1,0,0,0,0;这等0,0,0,1,1;这太0,1,2,0,0;这样7,4,9,0,3;着实3,1,2,4,1;真2,1,6,5,2;真么0,8,0,0,0;真是0,1,0,2,0;真真的0,1,0,1,0;真正1,0,0,0,1;至0,0,0,1,1;最0,0,0,1,1

④冯春田:《〈聊斋俚曲集〉语法研究》,开封:河南大学出版社,2003年,第51页。

④刘江涛、刘立佳:《SPSS数据统计与分析应用教程基础篇》,北京:清华大学出版社,2017年,第61-65页。

④Majtey, A. P., P. W. Lamberti & D. P. Prato. Jensen—Shannon Divergence as a Measure of Distinguishability between Mixed Quantum States, Physical Review A, Vol. 72(5), 2005

④Field, D. A. Laplacian Smoothing and Delaunay Triangulations, Communications in Applied Numerical Methods, Vol. 4 (6), 1988.

④陈军:《论戏剧与小说的文体区别及沟通》,《戏剧文学》2005年第12期。

④甘海岚、张丽娟:《京味文学散论》,北京:北京燕山出版社,1997年,第241-244页。

④抽样作品信息如下:老舍:《骆驼祥子》,海口:南海出版公司,2010年;老舍:《龙须沟》,北京:人民文学出版社,1979年;老舍:《茶馆》,成都:四川人民出版社,2017年;苏叔阳:《故土》,北京:人民文学出版社,1984年;苏叔阳:《左邻右舍》,北京:中国戏剧出版社,1981年;苏叔阳:《丹心谱》,北京:中国戏剧出版社,1980年。

④“丁耀亢说”如:张清吉:《〈醒世姻缘传〉作者补证》,《明清小说研究》1995年第1期;田璞:《〈醒世姻缘传〉的作者是丁耀亢》,《河南大学学报》(社科版)1982年第5期。

④“章丘人说”如:金性尧:《〈醒世姻缘传〉作者非蒲松龄说》,朱东润等主编:《中华文史论丛》,上海:上海古籍出版社,1980年第4辑;徐朔方:《论〈醒世姻缘传〉及其和〈金瓶梅〉的关系》,徐朔方:《古代戏曲小说研究》,杭州:浙江大学出版社,2008年,第449-464页。

⑤⑤童万周:《〈醒世姻缘传〉后记》,[清]西周生著,童万周校注:《醒世姻缘传》,郑州:中州古籍出版社,1997年,第948页。

⑤[民国]李葆恂:《旧学庵笔记》,义州李氏从刻本,1916年。

⑤王立鹏:《从〈醒世姻缘传〉的思想内容和结构看其作者》,《明清小说研究》1997年第3期。

⑤吴圣昔:《邓志谟乡里、字号、生年探考——〈邓志谟考论〉之一》,《明清小说研究》1992年第2期。

⑤罗训森:《中华罗氏通谱第1册》,北京:中国文史出版社,2007年,第592页。

⑤小说见[清]丁耀亢《丁耀亢全集》中卷(郑州:中州古籍出版社,1999年);杂剧见[清]丁耀亢《丁耀亢全集》上卷(郑州:中州古籍出版社,1999年)。

⑤邹宗良:《〈聊斋俚曲集〉前言》,[清]蒲松龄:《聊斋俚曲集》,北京:国际文化出版公司,1999年,第7-8页。

⑤对散点分布差异显著度大小的评估,应考虑时代因素。就强程度标记在不同体裁的散点分布而言,古代白话小说、剧曲的语言除有书面语和口语成分外,还有诗词所夹杂的不少文言成分,导致古代作品间的语体特征差异显大于现当代。因此古代作品可跟古代作品比(如g的散点分布差异显著度显大于a),现当代作品可跟现当代作品比(如i的散点分布差异显著度显大于b—c),不宜制定泛时即跨时代的统一的P值门槛。

⑤余韵:《巴金前后期小说的计量风格学研究》,华中师范大学硕士论文,2017年。