

用计量风格学方法考察《水浒传》的作者争议问题 ——以罗贯中《平妖传》为参照

宋丽

清华大学人文学院
中国语言文学系, 北京 100084
song-l19@mails.tsinghua.edu.cn

刘颖

清华大学人文学院
中国语言文学系, 北京 100084
yingliu@mail.tsinghua.edu.cn

摘要

《水浒传》是独著还是合著, 施耐庵和罗贯中是何关系一直存在争议。本文将其作者争议粗略归纳为施耐庵作、罗贯中作、施作罗续、罗作他续、施作罗改五种情况, 以罗贯中的《平妖传》为参照, 用假设检验、文本聚类、文本分类、波动风格计量等方法, 结合对文本内容的分析, 考察《水浒传》的写作风格, 试图为其作者身份认定提供参考。结果显示, 只有罗作他续的可能性大, 即前70回为罗贯中所作, 后由他人续写, 其他四种情况可能性都较小。

1 引言

《水浒传》是我国古典长篇小说四大名著之一, 讲的是北宋宣和年间, 一众好汉聚义梁山泊反抗欺压, 后被朝廷招安, 为国征战, 损兵折将, 最终消亡的故事。然而《水浒传》的作者究竟是谁众说纷纭, 施耐庵与罗贯中是否为同一人也尚无定论。温庆新(2018)根据多项研究成果指出明代人关于《水浒传》的作者主要有三种代表性意见, 分别为施耐庵作、罗贯中作、施罗合作。现代文学界也围绕《水浒传》的作者身份展开了大量研究。刘冬(1992)、洪东流(2008)等认为施耐庵就是兴化人施彦端; 王晓家(1998)、顾文若&焦中栋(1999)等认为施耐庵是罗贯中为避文祸而取的托名; 罗尔纲(1984)从赞词、叙事、对待人民大众的态度等三方面对《水浒传》百回本和罗贯中《三遂平妖传》(即《平妖传》二十回本)进行对勘, 得出《水浒传》前70回为罗贯中一人所著, 后30回为他人续加的结论; 吕乃岩(2008)通过将《水浒传》后半部分与罗贯中的其他作品相对照, 得出《水浒传》前半部分为施耐庵所作, 后半部分为罗贯中续加的结论; 杨林(1999)、欧阳健(2003)、李永祜(2011)等认为《水浒传》是施耐庵作底本, 罗贯中改写而成的。这些研究大多是从文献学、历史学的角度展开的, 或多或少会存在主观性和疏漏, 本文则利用统计学原理和计算机技术抽取文本特征, 采用计量风格学的多种方法对《水浒传》的写作风格进行考察, 试图为《水浒传》的作者身份认定提供一些参考。

由于《平妖传》是罗贯中所作的另一部著名小说⁰(或为冯梦龙增补, 下文将对其版本问题加以说明), 可以用于考察罗贯中与《水浒传》之间的关系, 而施耐庵除《水浒传》外并无其他传世之作, 故本文以罗贯中的《平妖传》为参照, 将《水浒传》的作者身份争议粗略地划分为以下五种情况:

- (1) **施耐庵作:** 施耐庵与罗贯中并非同一人, 《水浒传》为施耐庵一人独作, 即《水浒传》和《平妖传》分别为两人所作, 《水浒传》全书的写作风格应保持一致, 且与《平妖传》不同。
- (2) **罗贯中作:** 施耐庵是罗贯中的托名, 《水浒传》为罗贯中一人独作, 即《水浒传》和《平妖传》为同一人所作, 《水浒传》全书的写作风格应保持一致, 且由于两部小说题材不同, 故《水浒传》的写作风格会与《平妖传》有所差异。
- (3) **施作罗续:** 施耐庵与罗贯中并非同一人, 《水浒传》前半部分为施耐庵所作, 后半部分为罗贯中续写, 即《水浒传》后半部分的作者与《平妖传》为同一人, 写作风格应较为相近, 且与《水浒传》前半部分有所差别。

©2020 中国计算语言学大会

根据《Creative Commons Attribution 4.0 International License》许可出版

⁰由于罗贯中的其他作品目前还没有高质量的人工标注文本, 且现有的自动分析工具遵循的分词和词性标注标准与本文使用的台湾中央研究院近代汉语标记语料库并不完全相符, 无法通过自动分析直接进行量化对比, 故本文仅选取《平妖传》为参照文本。另外, 由于施耐庵的真实身份、生卒年代尚无定论, 故本文选取参照作品时忽略时代因素。对于其他相关作品的分析将在后续研究中进一步展开。

- (4) **罗作他续**: 施耐庵是罗贯中的托名,《水浒传》前半部分为罗贯中所作,后半部分为不知名的他人续写,即《水浒传》前半部分的作者与《平妖传》为同一人,写作风格应较为相近,且与《水浒传》后半部分有所差别。
- (5) **施作罗改**: 施耐庵与罗贯中并非同一人,《水浒传》底本由施耐庵所作,后罗贯中对全书进行了修改润色,但修改比重不详。也就是说《平妖传》的作者对《水浒传》进行了修改,使得《水浒传》的写作风格与《平妖传》有相似之处,但具体对哪些部分进行了修改,以及修改的篇幅则有待考察。

《水浒传》自传世以来出现了多种版本,据《水浒书录》(马蹄疾,1986)可知,明刊本有25种,清刊本有47种。这些版本可大体分为繁本和简本两个系统,且文学界对繁本更为重视。繁本系统主要包括七十回本、百回本和百二十回本三种规模,每种规模又包括若干个版本。从大体上说,这三种规模的版本是内容依次增加的关系,七十回本的内容与百二十回本中的前70回基本一致,写到分封一百单八将为止;百回本在七十回本的基础上增加了接受招安、抗击辽国军队和剿灭方腊义军的故事;百二十回本又在百回本征方腊的情节之前增加了剿灭王庆和田虎所率领的起义队伍的故事,百二十回本中较为通行的是袁无涯本。《平妖传》有二十回本和四十回本两个版本系统,且它们分别有多种刻本,作者问题也存在一定争议:孙楷第(1958)等认为二十回本是罗贯中的原著,而四十回本为冯梦龙增补(这种观点较为流行);欧阳健(1985)等认为四十回本是罗贯中原著,而二十回本是其删节本;等等。尽管如此,罗贯中是《平妖传》的主要作者这一观点几乎得到公认,也就是说无论哪个版本,《平妖传》中必然会体现出罗贯中的写作风格。由于受时代所限,近代汉语小说在版本方面大多都存在有待考证的问题,难以避免,且版本的考证问题并非本文的重点,故本文仅以内容最为完整且最为通行的《水浒传》繁本百二十回本中的袁无涯本、《平妖传》四十回本中的嘉会堂本为研究对象,运用计量语言学的研究手段,对《水浒传》的写作风格进行考察,从而推测以上五种情况的可能性大小。

由于《水浒传》作者身份争议较大的部分主要体现在前70回与后50回的差异,所以本文以70回结尾为界,将《水浒传》分为两部分。为方便阐述,下文分别用“SHZa”“SHZb”和“PYZ”代指三个待考察总体,SHZa中包含70个样本,即《水浒传》的前70回;SHZb中包含50个样本,即《水浒传》的后50回;PYZ中包含40个样本,即《平妖传》的全40回。另外,使用 $Diff(X, Y)$ 表示两个总体之间的差异程度,如 $Diff(SHZa, SHZb)$ 表示SHZa和SHZb之间的差异程度。上述五种作者身份的情况可简单总结为表1。

序号	作者争议	说明	量化表示
1	施耐庵作	施耐庵一人独作120回	$Diff(SHZa, SHZb)$ 最小,且 $Diff(SHZ, PYZ)$ 较大
2	罗贯中作	罗贯中一人独作120回	$Diff(SHZa, SHZb)$ 最小, $Diff(SHZ, PYZ)$ 也较小
3	施作罗续	施作前70回,罗作后50回	$Diff(SHZb, PYZ)$ 最小
4	罗作他续	罗作前70回,不知名者作后50回	$Diff(SHZa, PYZ)$ 最小
5	施作罗改	施作底本,罗改写,改写程度不详	/

Table 1: 《水浒传》作者的主要争议情况归纳

2 语料概况及特征提取

两部小说的电子文本均来源于台湾中央研究院近代汉语标记语料库¹,语料均已进行了人工分词和词性标注,质量较高,且均为繁体中文(该语料库中使用的词类标记共计59个,排除标题、外文标记、被解释的字、记录者姓名,以及两部小说中均未出现的引述句这5项,本文考察的词类共计54项,其中虚词13项,实词41项)。由于古代汉语使用句读进行断句,数字化古籍中的标点符号都是由古籍整理人员根据现代汉语标点规范人为增添的,因此下文中提及的所有统计数据 and 文本特征均不考虑标点。两部小说的基本统计数据见表2。

¹<http://lingcorpus.iis.sinica.edu.tw/cgi-bin/kiwi/pkiwi/kiwi.sh?ukey=-1316487165&qtype=0>。

关于文本特征，一方面，本文提取了作者身份识别任务中较为常用，且已在多项实验中被验证有效的能够体现作者写作风格的特征，包括语法单元的长度、虚词、词类等。另一方面，虽然文本中实词的相关特征通常用于考察文本内容和主题，但也有不少学者从用词情况、对人物形象的塑造等角度来分析《水浒传》的作者(何心, 1954; 杨林, 1999; 李永祜, 2011)，可见考察《水浒传》的作者身份不能完全不考虑实词。为避免与内容高度相关的特征对作者身份判定造成影响，本文只补充提取在三个总体中均出现的高频实词Top500的词频作为特征。提取的特征共计1808个，见表3。

书名	《水浒传》	《平妖传》
章回数	120	40
总字数	726833	189356
总词数	546969	146134
总字型数	4217	3459
总词型数	19603	10922
总整句数	44225	9634
总小句数	114625	29034
平均每回词数	4558.08	3653.35

* 由于该语料库采用Big5字符集，有少量字符无法显示，且文中所有诗词均被删除，故上表中关于字数、类符的统计结果有些许误差。

Table 2: 《水浒传》和《平妖传》语料的基本统计数据 (不含标点)

特征	数量
词长、整句长、小句长的均值	3
词长、整句长、小句长的离散度	3
词汇丰富度 (类符形符比)	1
每个虚词的出现次数 (每千词)	247
在三个总体中均出现的高频实词Top500的出现次数 (每千词)	500
不同词类的数量 (每千词)	54
二元词类标记串 ² Top500的出现次数 (每一千个标记串)	500
三元词类标记串Top500出现次数的 (每一千个标记串)	500
总计	1808

* 虽标点符号均为后人所加，但断句情况在一定程度上也能体现作者的写作风格，所以本文也提取了整句和小句的长度及离散度作为特征。

Table 3: 提取的特征及其数量

3 “施耐庵作”&“罗贯中作”&“施作罗续”的可能性分析

首先使用假设检验和聚类分析的方法对“施耐庵作”“罗贯中作”和“施作罗续”这三种情况的可能性进行考察。

3.1 假设检验

假设检验³可用于考察不同总体之间是否存在显著差异。由于本文考察的三个总体的样本量不大且不相等，为避免数据分布和方差的差异对假设检验结果造成影响，本文先针对每个特征，分别对三个总体做Shapiro-Wilk检验 (适用于样本量小于2000时的正态性检验) (Shapiro & Wilk, 1965)和Levene方差齐性检验(Levene, 2004)，抽取出了在三个总体中均服从正态分布 (显著性水平 $\alpha = 0.05^4$ ，下同) 且满足方差齐性的特征56个 (包括高频实词“去”、虚词“了”、词类VC (动作及物动词)、平均小句长等)。然后，分别针对这56特征，对SHZa、SHZb和PYZ这三个总体做单因素方差分析 (可用于考察多个正态总体的均值之间差异是否显著，仅涉及一个变量)，结果显示三个总体在47个方面存在显著差异。为进一步确认显著差异存在于哪几组数据之间，还利用LSD检验法(Fisher, 1935)进行了多重比较，得到了每两个总体中有显著差异的特征的数量，见表4 (最小数量加粗，下同)。

针对余下特征，对三个总体做Kruskal-Wallis H检验⁵ (通过秩和均值来考察多个总体差异是否显著，不依赖总体的分布，仅涉及一个变量。以下简称K-W检验) (Kruskal & Wallis, 1952)，结果显示三个总体在1227个方面存在显著差异，进一步做多重比较(Gibbons & Chakraborti, 2011)，每两个总体中有显著差异的特征数量见表4。

²“词类标记串”是指已进行词性标注的文本中相邻的n个词类标记形成的串，如句子“莊客/Na 報知/VE 史進/Nb”中三个词的词类标记分别为“Na (普通名词)”“VE (动作句宾动词)”“Nb (专有名称)”，则“Na VE”和“VE Nb”为二元词类标记串。“Na VE Nb”为三元词类标记串。

³假设检验的具体方法可见《概率论与数理统计》(茆诗松&周纪芑, 2000)。

⁴统计学上一般把概率 ≤ 0.05 的事件称为小概率事件。

⁵有学者将Kruskal-Wallis H检验称作单因素方差分析的非参数方法。

被比较对象	差异显著的特征数量		
	LSD检验 (共47个特征)	K-W多重检验 (共1227个特征)	合计
SHZa vs. SHZb	33	622	655
SHZa vs. PYZ	29	419	448
SHZb vs. PYZ	41	673	714

Table 4: 每两个总体中差异显著的特征数量对比

观察表4可知，在被实施多重比较的特征中，不论是进行LSD检验的，还是进行K-W多重检验的，《水浒传》前70回与《平妖传》呈现显著差异的特征数量都是最少的，也就是说与其他两组相比，它们在较少的方面存在差异。众所周知，这两部小说的题材不同（《水浒传》为英雄传奇小说，《平妖传》为神魔小说），但多重比较结果却显示出自同一部小说的《水浒传》前70回和后50回之间的差异程度较大，与《水浒传》后50回和《平妖传》的差异特征数量接近，反而是分别出自不同题材的两部小说的《水浒传》前70回和《平妖传》差异程度较小。

3.2 文本聚类

接下来采用两种较为常用的聚类方法——K均值聚类和层次聚类⁶来对全部160个样本进行聚类分析，考察两部小说中哪些章回的写作风格相似度高，哪些章回的写作风格差异大，写作风格相似度越高的章回越有可能出自同一位作者。为消除量纲的影响，在预处理时对所有特征均进行了Z-score标准化处理⁷。（本文使用的聚类和分类方法均在预处理时对特征进行了标准化处理，下文不再赘述。）

K均值聚类以 k 为输入参数，把 n 个对象的集合分为 k 个簇。本文指定 $k = 2$ （即分为两类），文本之间的相似度使用欧氏距离计算。分类结果见表5（数字加粗表示被分到相应类别的章回数较多）。由于样本数据处于一个高维空间中，K均值聚类的结果难以用平面图表示，所以采用主成分分析法（Principal Component Analysis, 简称PCA）（Jolliffe, 1986）进行维数约减，提取了最重要的两个主成分（保留的信息量分别为0.11和0.05），将结果映射到二维平面，作散点图1。从图中可见，160个样本明显被分为了两类，也就是说仅0.16的信息量即可得到近似线性二分的结果。

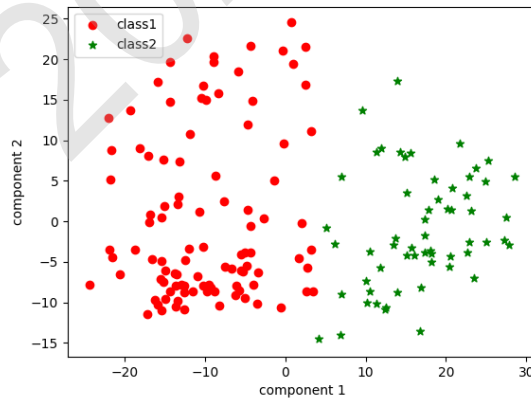


Figure 1: 两部小说各章回的K均值聚类结果（主成分Top2）

层次聚类将数据对象组成一棵聚类树，无需指定类别数量。本文用欧氏距离计算文本之间的相似度，用离差平方和计算类与类之间的相异度。聚类结果见图2，其中横坐标为各章回编号，纵坐标为文本之间的相似度。以150为界作一条横线，可将160个样本分为两类，类别内部的欧氏距离均小于150，而两个类别之间的欧氏距离超过了200。各章回的具体分类情况见表5。

观察K均值聚类和层次聚类的结果，发现《水浒传》前70回和《平妖传》中的大部分文本都聚在了同一类，而《水浒传》后50回中的大部分文本则聚在了另一类，这说明三个总体中，《水浒传》前70回和《平妖传》之间的差异程度最小，与第3.1节中假设检验的结果相符。

综合假设检验和聚类分析的结果可知，三个总体之间差异程度的关系应为： $Diff(SHZa, PYZ) < Diff(SHZb, PYZ) < Diff(SHZa, SHZb)$ 。然而根据表1可知，若《水浒传》为“施耐庵作”或“罗贯中作”，则 $Diff(SHZa, SHZb)$ 应最小；而若是“施作罗续”，则 $Diff(SHZb, PYZ)$ 应最小，据此可推断，“施耐庵作”“罗贯中作”和“施作罗续”这三种情况的可能性小。

⁶聚类分析的具体方法可见《数据挖掘：概念与技术》（Han & Kamber著，范明&孟小峰译，2007）。

⁷公式为 $x^* = \frac{x - \bar{x}}{\sigma}$ 。

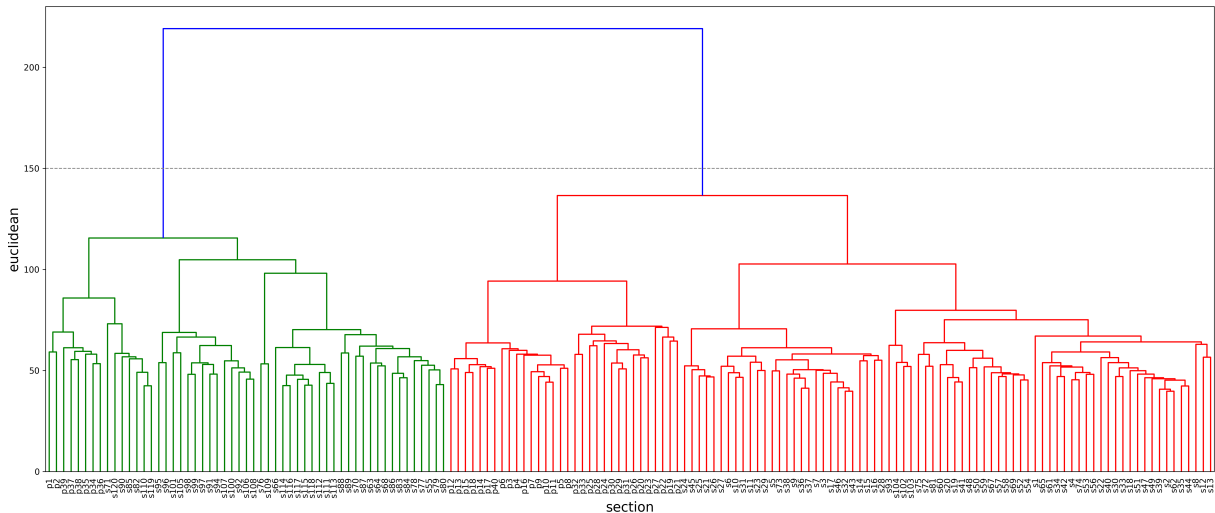


Figure 2: 两部小说各章回的层次聚类结果

文本	K均值聚类				层次聚类			
	类别一		类别二		类别一		类别二	
	章回编号	数量	章回编号	数量	章回编号	数量	章回编号	数量
SHZa	1-47; 49; 51-53; 56; 61-62; 65	55	48; 50; 54-55; 57-60; 63-64; 66-70	15	1-54; 56-62; 65; 67; 69	64	55; 63-64; 66; 68; 70	6
SHZb	72-74; 81; 93; 102-104	8	71; 75-80; 82-92; 94-101; 105-120	42	72; 74-75; 81; 93; 102-104	8	71; 73; 76-80; 82-92; 94-101; 105-120	42
PYZ	1-33; 35; 37-40	38	34; 36	2	3-33; 40	32	1-2; 34-39	8

Table 5: 三个总体中各章回的聚类结果详情

4 “罗作他续”的可能性分析

由第3节可知，SHZa、SHZb、PYZ这三个总体之间差异程度最小的应为SHZa和PYZ，这一结果与“罗作他续”相匹配。为进一步验证这一情况，本文利用文本分类方法支持向量机（Support Vector Machine，简称SVM）(Cortes & Vapnik, 1995)，从作者身份归属（Authorship Attribution，即判断待分类文本的作者是已知作者中的哪一个）和作者身份验证（Authorship Verification，即判断待分类文本的作者与已知作者是否为同一人）两个角度对上述三个总体加以考察。

4.1 作者身份归属

从作者身份归属的角度出发，假定《水浒传》前70回为罗贯中所作，后50回为另一不知名作者所作，考察《平妖传》中的各章回被分类到《水浒传》前70回和后50回的可能性大小。也就是说，假定SHZa和SHZb中的样本被分为不同的两类，用SVM考察PYZ中的各样本被分到这两类的可能性。由于SHZa和SHZb中的样本数量不均衡（70 vs. 50），故在实际操作时，随机剔除SHZa中的20个样本进行训练，重复10次，以最终输出的概率均值为预测结果，作堆积柱状图3。图中横坐标为PYZ中各章回的编号，纵坐标为概率值，每一章回被分到SHZa和SHZb的概率值之和都为1。

虽然总体看来，PYZ中的文本被分类到SHZa和SHZb的几率大致相当，但若除去开头和结尾的几个章回，则更多地被分到SHZa中。研究《平妖传》的多数学者认为其二十回本为罗贯中所作，而四十回本是冯梦龙增补而成。谭红(2013)从多方面对这两个版本进行了细致的对比分析，综合其分析结果可知，与二十回本相比，人物方面，四十回本中新增的人物（如九

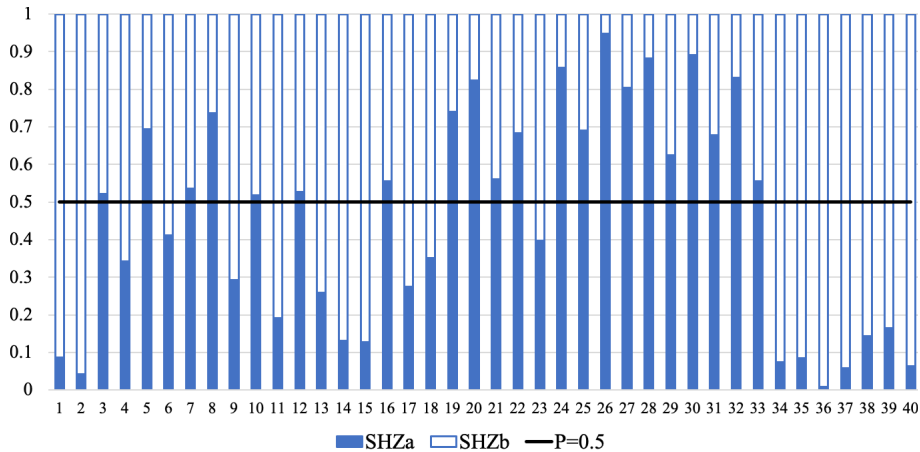


Figure 3: 《平妖传》中各章回被分类到《水浒传》前70回和后50回的可能性

天玄女、白猿神)和对原有人物(如王则、蛋子和尚)的改写大多出现于开篇第1回和结尾第37-40回;情节方面,四十回本新增和大幅扩充的内容主要出现在第1-18, 34-40回。也就是说,中间的章回,尤其是第19-33回基本保留了罗贯中二十回本的原貌。这一结论与图3的结果高度匹配,可见《平妖传》中保留的罗贯中的写作风格与《水浒传》前70回很相近,所以《平妖传》与《水浒传》前70回的作者很可能都是罗贯中,这说明“施耐庵”可能只是罗贯中的一个托名,而《水浒传》前后写作风格差异大,后半部分可能是由他人续写。这一结果支持了“罗作他续”的说法。

4.2 作者身份验证

Koppel & Schler (2004)指出,由于主题、体裁、写作目的、写作年代等因素的变化,甚至出于掩饰身份的目的,同一个作者所作的不同作品在少量特征方面会有明显不同。也就是说,这些特征对作者身份的验证造成了困难。为了解决这种困难,他们提出了“揭露(Unmasking)算法”,其思路为将作者身份已知的文本和作者身份未知的文本拆分为多个组块,并假定它们分属不同的两类,不断删除对区分这两类文本最有用的特征,测量交叉验证时分类正确率下降的速度,若下降速度快,则表示两个文本难以区分,即作者为同一人,反之则作者并非同一人。在验证小说*The House of Seven Gables*的作者身份时,揭露算法在三位候选作者中明确选中了正确的作者Hawthorne⁸。Bevendorff et al. (2019)通过改变生成组块的方式,打破对组块长度的要求,将揭露算法引入了对短文本的作者身份验证任务中,也取得了与目前高性能算法相当的效果。本文参考揭露算法的思路,以小说章回为文本的自然组块,进行了如下实验:

迭代50次{

步骤①: 对总体A和B,假定它们的作者并非同一人(即分属两个不同类别),利用LinearSVM模型训练分类器,做10折交叉验证,计算平均正确率

步骤②: 去掉步骤①训练出的分类器中权重最高的5个特征和权重最低的5个特征

步骤③: 回到步骤①

}

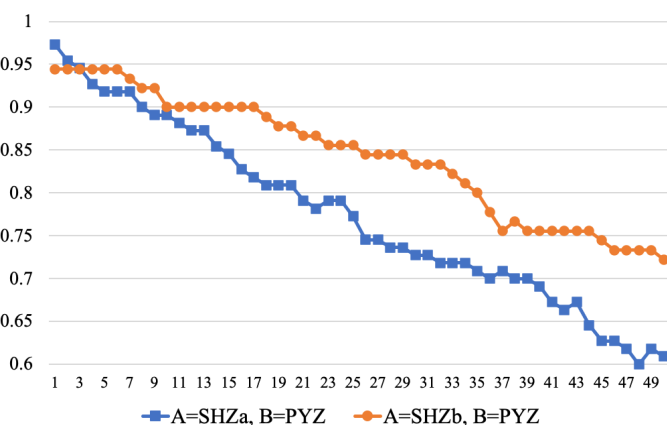


Figure 4: 揭露算法的分类正确率变化情况

⁸将*The House of Seven Gables*与其他作者的作品比较时,正确率下降缓慢,而与Hawthorne的作品比较时,正确率下降迅速。

分别设A=SHZa, B=PYZ; A=SHZb, B=PYZ⁹, 各进行一次实验, 将50次的分类正确率作折线图4, 横坐标为迭代次数, 纵坐标为交叉验证后的平均分类正确率。从图中可以很直观地看出, SHZa vs. PYZ的分类正确率的下降速度明显比SHZb vs. PYZ快, 也就是说, 《水浒传》前70回和《平妖传》的作者为同一人的可能性大, 而《水浒传》后50回和《平妖传》的作者为同一人的可能性较小。这一结果也支持“罗作他续”的说法。

4.3 对诗词的补充分析

由于本文所考察的文本数据中的诗词均被剔除, 也就是说上述统计数据并未能覆盖小说中诗词部分的信息, 所以需对两部小说中的诗句、赞词等内容进行补充分析。罗尔纲(1984)曾对《水浒传》百回本和《平妖传》二十回本中的赞词进行对勘, 发现《平妖传》二十回本中

水浒传		平妖传	
诗词	章回	诗词	章回
祥云迷凤阁, 瑞气罩龙楼。……隐隐净鞭三下响, 层层文武两班齐。	1	祥云迷凤阁, 瑞气罩龙楼。……隐隐净鞭三下响, 层层文武两班齐。	35
祥云笼凤阙, 瑞霭罩龙楼。	72		
香焚宝鼎, 花插金瓶。……琥珀杯中, 满泛着瑶池玉液。……鳞鳞脍切银丝, 细细茶烹玉蕊。	2	香焚宝鼎, 花插金瓶。……琥珀杯中, 满泛瑶池玉液。……玻璃碗, 供熊掌驼峰。鳞鳞脍切银丝, 细细茶烹玉蕊。	25
玻璃碗内, 供献上熊掌驼蹄; 琥珀杯中, 满斟下瑶池玉液。	119		
桂花离海峤, 云叶散天衢。……冰轮展出三千里, 玉兔平吞四百州。	2	桂华离海峤, 云叶散天衢。……冰轮碾破三千里, 玉魄横吞万里秋。	18
门迎黄道, 山接青龙。	9	门迎黄道, 山接青龙	9
银河耿耿, 玉漏迢迢。……贪淫妓女心如铁, 仗义英雄气似虹。	21	银河耿耿, 玉漏迢迢。……妖邪贼侣心如, 忠义英雄气似虹。	38
尽道丰年瑞, 丰年瑞若何? 长安有贫者, 宜瑞不宜多。	24	纷纷柳絮, 片片鹅毛。……万户银装, 多少行人肠断。正是: 尽道丰年瑞, 丰年瑞若何; 长安有贫者, 宜瑞不宜多。	18
纷纷柳絮, 片片鹅毛。……万户银装, 多少幽人成佳句。正是: 尽道丰年好, 丰年瑞若何? 边关多荷载, 宜瑞不宜多。	93		
平生正直, 禀性贤明。……慷慨文章欺李杜, 贤良方正胜龚黄。	27	平生正直, 禀性贤明。……果然是慷慨文章欺李杜, 贤良方正胜龚黄。	29
身穿缟素, 腰系孝裙。……恰似嫦娥离月殿, 浑如织女下瑶池。	32	身穿缟素, 腰系麻裙。……恰似嫦娥离月殿, 浑如织女下瑶池。	31
犯由牌高贴, 人言此去几时回。……长休饭钵内难吞, 永别酒口中怎咽。	40	两声破鼓响, 一棒碎锣鸣。……犯由牌高贴, 人言此去几时回。……长休饭, 喉里难吞。永别酒, 口中怎咽。……刀剑林中刽子手, 犹如追命鬼。	40
两声破鼓响, 一棒碎锣鸣。……刀剑林中, 掌法吏犹如追命鬼。	62		
骊山顶上, 多应褒姒戏诸侯; 赤壁坡前, 有若周瑜施妙计。	41	骊山顶上, 料应褒姒逞英雄。扬子江头, 不若周郎施妙计。	18
金钉朱户, 碧瓦雕檐。……若非天上神仙府, 定是人间帝王家。	42	金钉朱户, 碧瓦雕檐。……若非天上神仙府, 定是人间帝王家。	25
		金钉朱户, 碧瓦盈檐。	28
正天仙容描不就, 威严形像画难成。	42	苍形古貌, 鹤发童颜。眼昏似秋月笼烟, 眉白如晓霜映日。……正大仙客描不就, 威严形像画难成。	25
苍然古貌, 鹤发酡颜。眼昏似秋月笼烟, 眉白如晓霜映日。……	53		
一群白鹤听经, 数个青衣碾药。	53	仙童击鼓, 一群白鹤听经; 玉女鸣钟, 数个青猿煨药;	28
凤落荒坡, 尽脱浑身羽翼。……吕虔亡所佩之刀, 雷焕失丰城之剑。	56	凤落荒坡, 脱尽浑身锦羽。……吕虔亡腰下之刀, 雷焕失匣中之剑;	38
六尺以上身材, 二十四五年纪。……着一双土黄皮油膀脚靴。	61	六尺以下身材, 二十二三年纪。……着一对土黄色多耳皮鞋,	24
唇若涂朱, 睛如点漆, 面似堆琼。	61	面如傅粉, 体似凝脂, 唇若涂朱, 目如点漆。	35
人人要建封侯绩, 个个思成荡寇功。	91	人人欲建封侯绩, 个个思成荡寇功。	35
凤眼浓眉如画, 微须白面红颜。……七尺身材壮健。善会偷香窃玉,	101	凤眼浓眉如画, 黄须白面高颧。……六尺身材壮健。善会开弓发弩,	31
神器从来不可干, 僭王称号谁能安?	113	神器从来不可干, 僭王称制谁能安。	40

Table 6: 两部小说中的相似诗词及其所在章回

⁹由于该实验主要关注的是删除特征后分类正确率的变化速度, 所以并未对两类样本数量的不均衡做额外处理。

有13处赞词被直接或经改写后插入了《水浒传》百回本的前70回中（插入15处）。这些赞词均被相应地保留在了《水浒传》百二十回本和《平妖传》四十回本中。此外，本文还通过字符串比对¹⁰，在两部小说中找到了另外9处或长或短的相似诗词，例如《水浒传》第41回描写众好汉火烧黄文炳家的场景时用了一句“**骊山顶上，多应褒姒戏诸侯；赤壁坡前，有若周瑜施妙计。**”《平妖传》第18回描写胡员外家解库起火的场景时用了一句“**骊山顶上，料应褒姒逞英雄。扬子江头，不若周郎施妙计。**”这两句话的句式、用词、内容都很相似，只改动了个别字词（相同的字词加粗）。

表6罗列了两部小说中24处相似的诗词以及它们各自的章回信息（受篇幅所限，只摘录部分相似内容）。其中仅有6处出自《水浒传》后50回，其他全都出自《水浒传》前70回。可见，《水浒传》前70回和《平妖传》中的诗词有较高的相似度，这进一步佐证了SHZa和PYZ很可能是出自同一人之手。当然，不可否认的是，诗词的相似度高也有可能是因为不同作者挪用并改写，这有待进一步考证。

5 “施作罗改”的波动风格计量分析

商韬&陈年希(1999)指出，宋元以来的许多话本小说和章回小说往往不是个人的独立创作，而是时代积累型的集体创作。所以仅凭上述分析，不能直接断定“施耐庵”只是罗贯中的一个托名，《水浒传》是罗作他续。或许《水浒传》也并非一人独著或有人续写，而是一部集体创作的，无法区分各章回分别由谁所写的作品。接下来，本文假定《水浒传》的确符合第五种情况“施作罗改”，引入波动风格计量的方法对二位作者的风格比重加以分析。

波动风格计量（Rolling Stylometry）是由Eder提出的一种考察合作型文本的文体特征或写作风格的方法，它将有监督的机器学习分类方法与序列分析相结合。Eder (2016)指出线性序列中元素的顺序与元素本身同样重要，文本中连续的部分可以体现风格特征的线性发展。这一思想最早被应用于语言学是1913年Markov (2006)¹¹提出的马尔科夫链（Markov chains），只不过马尔科夫链的滑动窗口只包含连续的几个字符，而波动风格计量方法的滑动窗口则包含连续的上百或上千个词，其目的在于从文本中生成一连串含有重叠内容，前后相互关联的虚拟子样本，从而可以使用有监督的机器学习分类方法测试它们在整个文本中的风格一致性。与传统的方法相比，波动风格计量分析方法打破了文本中预设的界线，如章回、段落等，转而对一系列连续的，长度相等且包含重叠成分的样本进行分类，因此可以用于考察合作型文本中不同作者各自的贡献份额的变化情况。图5是该方法抽取子样本的示意图，其中 k 表示每个子样本包含的词数， d 表示连续两个子样本中重叠的内容长度。Plecháč (2019)利用Shakespeare和Fletcher的各4部剧作，以500个高频节奏类型和500个高频词为特征，测试了这种方法的有效性¹²，结果显示，在Shakespeare的剧作中，Fletcher的贡献程度都非常低，反之亦然，若根据概率值大小判断作者身份，正确率则高达0.9977，可见波动风格计量分析方法对作者贡献程度的度量具有较高的有效性。

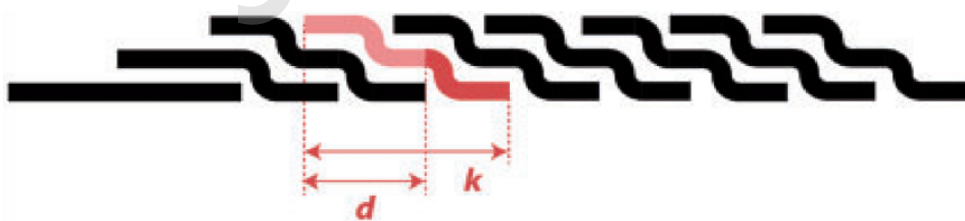


Figure 5: 波动风格计量方法抽取子样本的示意图（截取自Eder (2016)）

本文考察的《水浒传》的作者身份的第五种情况是“施作罗改”，其蕴涵为施耐庵与罗贯中并非同一人。由于除《水浒传》以外，施耐庵并没有其他传世之作，所以无法选用其他作品作为施耐庵的训练集。综合持“施作罗改”观点的多位学者的分析，他们虽然对施耐庵和罗贯中对《水浒传》的具体贡献各持己见，但几乎都认为施耐庵的底本主要是以宋江的故事为主线，罗贯中在前期保留了与宋江有关的主要情节，即约为后人总结的“宋十回”（第33-42回）。从

¹⁰诗词比对选用的是简体版文本。

¹¹该篇论文成文于1913年，2006年由*Science in Context*期刊出版。

¹²与Eder (2016)不同的是，Plecháč (2019)设置的滑动窗口以剧作的台词（line）为单位，而非单个词。

第3.2节聚类分析的结果可知，这些章回也的确被聚在了同一类。所以，本文提取“宋十回”覆盖的97个样本作为施耐庵的训练样本，以《平妖传》的284个样本作为罗贯中的训练样本，用波动风格计量的方法对《水浒传》全书进行测试。

抽取子样本时，本文设定 $k=5000$ ， $d=4500$ ，即对训练样本和测试样本，均每隔500词截取一个子样本，每个样本规模为5000词，也就是说每两个连续样本中有4500词的内容重叠。文本分类方法同样选取SVM。因滑动窗口中有句子被拆分，故排除了与句长相关的特征。另外，由于训练样本数量不均衡，故在实际操作时，随机选取《平妖传》中的97个样本参与训练，重复50次，以最终输出的概率均值为预测结果。图6显示了测试结果，其中横坐标为1085个测试子样本的编号（根据文本内容排序），纵坐标为SVM的分类可能性（概率值），由于是二分类问题，故每一个样本被分到两类的可能性之和均为1（为便于作图，分类为施耐庵的可能性用正数表示，分类为罗贯中的可能性用负数表示）。黑色曲线为分到两类的可能性的均值，若落在深色区域，则表示该部分内容主要为施耐庵所写；若落在浅色区域，则表示该部分内容主要为罗贯中所写。有底色的部分为抽取的施耐庵的训练样本，所以测试结果显示被分为施耐庵所作的概率接近1。点线（……）为《水浒传》每10回的分界线，虚点线（- · - · - ·）为前70回和后50回的分界线。

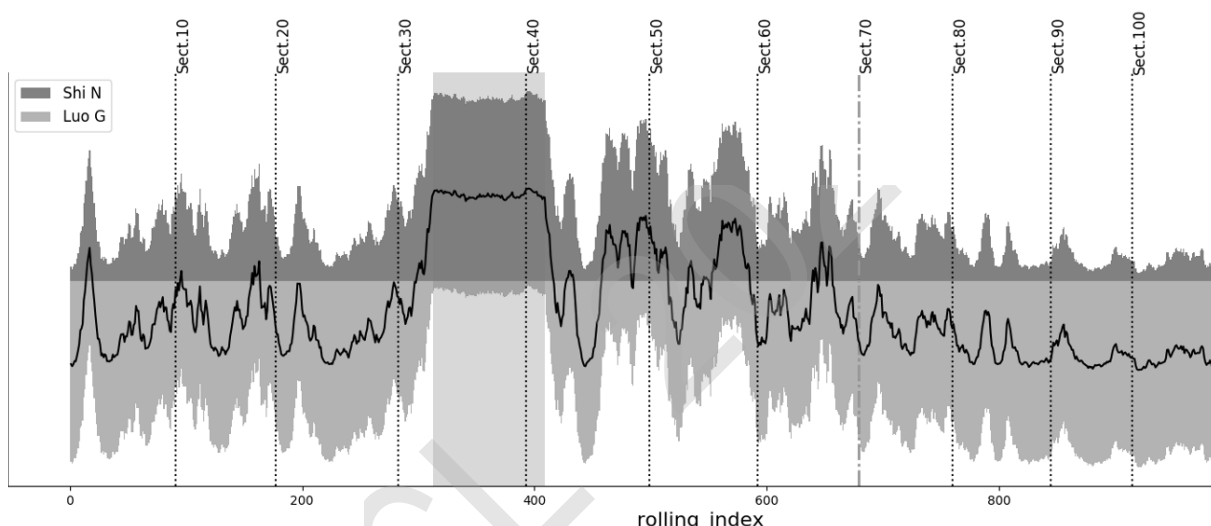


Figure 6: 《水浒传》作者身份识别的波动风格计量测试结果

从图6中可见，在假定《水浒传》是“施作罗改”的前提下，波动风格计量方法的测试结果的确实显示《水浒传》全书中，二者的风格占比存在明显的波动。但是，除了本就被选为施耐庵训练样本的部分，其余部分的黑色曲线大多落在了罗贯中的区域，约第1-32，43-46，52-56，60-110，117-120回，也就是这些部分都主要由罗贯中所写。罗贯中对《水浒传》的贡献度很高，且显然远超施耐庵，由此可以推测，“施作罗改”的可能性较小，即使真的存在所谓施耐庵的文本，在经过罗贯中的改写后，也已经基本不复原貌。

由于事先已做了“施作罗改”的假定，且没有施耐庵的其他作品作参照，所以同样还需进一步结合文献学和历史学的知识对该测试结果加以考证。下面结合波动风格计量的测试结果与《水浒传》的文本内容，就相关研究成果中的部分观点加以比对分析：

(1) 侯会(1986)、杨林(1999)认为《水浒传》开篇关于鲁智深、林冲等人的故事（即第1-13回）和描写武松的“武十回”（从第23回武松打虎开始，到第32回武行者醉打孔亮结束）这些与宋江的故事毫无联系的内容均为罗贯中增补。波动风格计量的测试结果显示，前32回都应主要出自罗贯中之手，所以该观点仅与测试结果部分相符。

(2) 李永祜(2011)认为杨志押运生辰纲（第16回）是施耐庵重点描写的内容；宋江私放晁盖、杀掉阎婆惜后逃亡避难（分别第18、21-22回）等情节均出自施耐庵之手；第72回柴进潜入宫禁，在睿思殿看到素面屏风的场景是施耐庵根据自己的真实经历模仿的；而宋江率军队征讨辽国（第83-89回）的故事是罗贯中设计出的鲁智深经历的重要关节。波动风格计量的测试结果显示，第16、18、21、22、72、83-89回均主要为罗贯中所作，所以这几个观点中只有征辽情

节由罗贯中设计这一个观点与测试结果相符。

(3) 孟繁仁(1990)、宣啸东(1991)等认为“征田虎”部分的人物许贯忠是罗贯中在修改施耐庵创作的底本时创作出的自己的虚像,也就是说关于许贯忠的内容均是由罗贯中主笔撰写。原著中许贯忠于第90回作为燕青巧遇的故交出场,带他到家中做客,提醒他功成名就之时要及时抽身,送他画作,着墨较多,第91、94回也对许贯忠有所提及。波动风格计量的测试结果显示,罗贯中对第90-94回的改动的确较大,所以该观点与测试结果相符。

(4) 杨林(1999)认为宋江等人平方腊有功,升官受赏是施耐庵底本的结局,而他们功成名就之后被害死的情节并非出自施耐庵之手,而是罗贯中在元、明朝代更迭之际看到乱世的残酷现实后所增益的内容。原著中征方腊的情节从第110回开始,到第119回以宋军平定方腊叛乱,梁山好汉得到封赏,宋江衣锦还乡结束,而宋江、卢俊义等人被蔡京、高俅等奸臣迫害致死的情节出现在第120回。波动风格计量的测试结果显示,罗贯中对第120回的改动的确较大,对第111-116回的改动较小,所以该观点与测试结果部分相符。

总的来说,波动风格计量分析的结果仅与“施作罗改”的少数观点相符,其他观点诸如施耐庵集撰了前人所作的水浒故事,罗贯中删去了施耐庵底本的一些内容等等,仅仅基于本文的测试结果无法得到验证,有待进一步考察。

6 结论及未来展望

本文以《水浒传》的作者身份为研究对象,根据前人研究将《水浒传》的作者争议粗略地归纳为施耐庵作、罗贯中作、施作罗续、罗作他续、施作罗改等五种情况,以罗贯中的另一部小说《平妖传》为参考,利用统计学原理和计算机技术对文本特征进行抽取,采用计量风格学的多种方法对《水浒传》的写作风格进行考察,试图为《水浒传》的作者身份认定提供参考。

首先通过假设检验、K均值聚类和层次聚类发现,在《水浒传》前70回、《水浒传》后50回、《平妖传》这三个总体中,《水浒传》前70回和《平妖传》的写作风格之间差异最小,从而得出结论:《水浒传》为施耐庵作、罗贯中作、施作罗续的可能性小,而罗作他续的可能性大。然后,从作者身份归属和作者身份验证两个角度出发,利用文本分类方法SVM,结合对两部小说中相似诗词的补充分析,进一步验证了罗作他续可能性大这一结论,这或许可以为“施耐庵是罗贯中的化名”这一观点提供佐证。此外,还采用波动风格计量的方法对施作罗改的情况加以考察,测试结果显示,“施作罗改”的可能性也较小,即使的确存在施耐庵的底本,罗贯中的改写也已使其基本不复原貌,通过与原著内容的比对可知测试结果仅与施作罗改的少数观点相符。综合而言,《水浒传》作者争议的五种情况中只有罗作他续可能性大,其他四种情况可能性都较小。

由于施耐庵与罗贯中究竟是什么关系这一问题尚无定论,而本文在采用有监督的机器学习分类方法时对施耐庵的身份做了假定,而且本文尚未考察“罗作他续”中“他”的身份,所以未来还需进一步结合文献学和历史学的知识对分类结果的可靠性加以考证,并探究罗作他续中“他”的身份。此外,未来还需进一步结合《水浒传》和《平妖传》的文本内容对本文的计量结果加以分析。

致谢

感谢审稿人的中肯建议。本研究得到了教育部社会科学研究一般项目“语体特征的自动提取和研究”(17YJAZH056),国家社会科学基金重大项目“基于大数据技术的古代文学经典文本分析与研究”(18ZDA238),以及清华大学人文社科振兴项目“基于大数据技术的明清小说分析和研究”(2019THZWJC38)的资助。

参考文献

- Bevendorff J, Stein B, Hagen M, et al. 2019. Generalizing Unmasking for Short Texts. *North American Chapter of the Association for Computational Linguistics*, 654-659.
- Cortes C, Vapnik V. 1995. Support-Vector Networks. *Machine learning*, 20(3):273-297.
- Eder M. 2016. Rolling Stylometry. *Digital Scholarship in the Humanities*, 31(3):457-469.
- Fisher R A. 1935. *The Design of Experiments*. Oliver and Boyd, Edinburgh and London.

- Gibbons J D, Chakraborti S. 2011. *Nonparametric Statistical Inference*. Taylor and Francis Group, Florida.
- Jolliffe I T. 1986. *Principal Component Analysis*. Springer Verlag, New York.
- Koppel M, Schler J. 2004. Authorship Verification as a One-class Classification Problem. *International Conference on Machine Learning*.
- Kruskal W, Wallis W A. 1952. Use of Ranks in One-criterion Analysis of Variance. *Journal of the American Statistical Association*, 47(260):583–621.
- Levene H. 1960. Robust Tests for Equality of Variances. *Contributions to probability and statistics*, 278–292.
- Markov A A. 2006. An Example of Statistical Investigation of the Text “Eugene Onegin” Concerning the Connection of Samples in Chains. *Science in Context*, 19(4):591–600.
- Plecháč P. 2006. Relative contributions of Shakespeare and Fletcher in Henry VIII: An Analysis Based on Most Frequent Words and Most Frequent Rhythmic Patterns. *arXiv: Computation and Language*.
- Shapiro S S, Wilk M B. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3–4):591–611.
- (加)Han J, Kamber M著. 范明, 孟小峰译. 2008. 数据挖掘: 概念与技术. 机械工业出版社, 北京.
- 顾文若, 焦中栋. 1999. “施耐庵”为罗贯中之托名. 晋阳学刊, (1):107–108.
- 何心. 1954. 水浒研究. 上海文艺联合出版社, 上海.
- 洪东流. 2008. 水浒解密考. 学林出版社, 北京.
- 侯会. 1986. 《水浒》源流管窥. 文学遗产, (4):64–71.
- 李永祜. 2011. 施耐庵和罗贯中对《水浒传》成书的贡献. 菏泽学院学报, 33(4):24–37.
- 刘冬. 1992. 施耐庵探考. 南京出版社, 南京.
- 罗尔纲. 1984. 从罗贯中《三遂平妖传》看《水浒传》著者和原本问题. 学术月刊, (10):22–32.
- 吕乃岩. 2008. 试说罗贯中续《水浒》. 北京大学学报(哲学社会科学版), (2):68–76.
- 马蹄疾. 1986. 水浒书录. 上海古籍出版社, 上海.
- 茆诗松, 周纪芑. 2000. 概率论与数理统计. 中国统计出版社, 北京.
- 孟繁仁. 1990. “许贯忠”是罗贯中的虚象. 晋阳学刊, (4):21–29.
- 欧阳健. 1985. 《三遂平妖传》原本考辨. 中华文史论丛, (3):149–165.
- 欧阳健. 2003. 罗贯中研究三题. 东南大学学报(哲学社会科学版), 5(5):104–111.
- 商韬, 陈年希. 1986. 用《三遂平妖传》不能说明《水浒传》的著者和原本问题——与罗尔纲先生商榷. 学术月刊, (2):55–59.
- 孙楷第. 1958. 日本东京所见小说书目. 人民文学出版社, 北京.
- 谭红. 2013. 《三遂平妖传》世德堂本与嘉会堂本比较研究. 山东师范大学, 山东.
- 王晓家. 1998. 《水浒传》作者考论. 陕西人民出版社, 西安.
- 温庆新. 2018. 文献传播学视野下的《水浒传》作者研究. 中国文化研究, (2):123–132.
- 宣啸东. 1991. 许贯忠之原型即罗贯中辩. 晋阳学刊, (3):66–69.
- 杨林. 1999. 罗贯中散论. 海南师范大学学报(社会科学版), (3):15–22.