

题名标点对我国人文社会科学学术 图书被引的影响研究

阮选敏¹, 吕冬晴¹, 成颖^{1,2}, 柯青¹

(1. 南京大学信息管理学院, 南京 210023; 2. 山东师范大学文学院, 济南 250014)

摘要 书名是学术图书的重要组成部分, 承担着概括主旨、吸引读者等要务。本文以CBKCI收录的1999—2009年出版的2844册学术图书为研究对象, 采用非参数检验、多元线性回归分析等统计方法探究书名中有无标点符号、标号与点号、标点组合、符号的数量以及标题句式等特征对被引的影响。研究结果发现, 相较于纯文字的书名, 在书名中合理地使用标点符号有助于提升图书的影响力; 书名仅包含点号或者同时包含点号和标号的图书被引较高; 未发现标点符号数量对图书被引的显著影响; 带有冒号的标题能够增加图书的被引; 顿号也可以让图书获得更高的引用; 书名中仅包含书名号时, 其被引较低; 复合式标题的图书比非复合式标题的图书具有更高的被引。

关键词 CBKCI; 标点符号; 图书被引; 线性回归

Factors Affecting Citations of Monographs in Chinese Humanities and Social Sciences: The Role of Punctuation Marks

Ruan Xuanmin¹, Lyu Dongqing¹, Cheng Ying^{1,2} and Ke Qing¹

(1. School of Information Management of Nanjing University, Nanjing 210023;
2. School of Chinese Language and Literature, Shandong Normal University, Jinan 250014)

Abstract: The title is an important part of monographs. It plays a significant role in summarizing the main content and attracting readers. This paper used monographs indexed in CBKCI and published between 1999 and 2009 to examine whether punctuation features such as the presence of punctuation as well as point and labelling marks, the combination and number of punctuation marks, and the structure of the title influence the number of times monographs are cited. Nonparametric tests and multiple linear regression analysis were used. The results showed the following. (1) Compared with plain text titles, correct use of punctuation in the title increased citations. Monographs whose title included labelling marks only or both labelling marks and point marks received more citations. The number of punctuation marks had no effect on citations. (2) The presence of a colon in the title increased citations of monographs. Titles with a slight pause mark were cited more than those without. Conversely, when the title included a punctuation mark used to enclose only the title of a book, the monograph had fewer citations. (3) Monographs with a compound title received more citations than those without.

Key words: CBKCI; punctuation marks; book citation; linear regression analysis

1 引言

标题是学术成果的点睛之笔, 以学术论文为

例, 标题是其结构上的重要组成部分^[1-2], 通常也是读者首次接触的部分^[3], 有时甚至是唯一可获取的部分^[4]。标题承担着反映论文内容、吸引读者^[5]等

收稿日期: 2018-12-20; 修回日期: 2019-03-07

基金项目: 国家社会科学基金项目“施引者引用意向与文献计量视角的学术论文被引影响因素研究”(17BTQ014)。

作者简介: 阮选敏, 女, 1994年生, 硕士研究生, 主要研究方向为信息检索、信息计量; 吕冬晴, 女, 1994年生, 硕士研究生, 主要研究方向为信息检索、信息计量; 成颖, 男, 1971年生, 教授, 博士生导师, 主要研究方向为用户信息行为、信息检索, E-mail: chengy@nju.edu.cn; 柯青, 女, 1979年生, 教授, 博士生导师, 主要研究方向为人机交互、用户信息行为。

重要任务。由于网络和各种文献数据库的出现,论文标题的“广告作用”日益凸显^[6],其质量在一定程度上能够影响论文的影响力^[7]。现有研究探讨了标题长度^[3]等特征对论文被引的价值,标题中的标点符号作为其中的一个显性特征受到了重点关注。目前,学界主要围绕标题中的非字母数字型符号(non-alphanumeric characters)^[8]及某些特定符号(如冒号、问号等)^[2,9]与论文被引间的关系等进行了深入研究。

学术图书作为重要的学术成果载体,通过检索未发现以学术图书为研究对象开展的涉及标题特征与被引间关系的研究。受到学术论文题名中标点符号对被引研究成果的启发,本研究推测学术图书的题名标点特征也会对其被引产生影响。潘文涛研究员^[10]根据2003年中国科技论文与引文数据库(CSTPCD)统计的结果显示,期刊论文占自然科学学术论文总被引的78.1%,而图书仅占15.5%;苏新宁教授^[11]根据中文社会科学引文索引(CSSCI)2000—2004年统的计结果显示,期刊论文占人文社会科学学术论文总被引的33.04%,而学术图书的比例则提高至50.45%。由此可见,学术图书在自然科学的学术交流中居于次要地位,而在人文社会科学中则明显重要得多。据此,本文拟先开展人文社会科学学术图书标题中的标点符号特征对被引的影响研究。

中文学术图书引文索引(Chinese Book Citation Index, CBKCI)数据库收录了1992年以来国内出版的中文人文社会科学21个学科的4000多册原创性学术专著,图书的遴选方法综合采用了文献计量与同行评议。鉴于CBKCI收录图书的高质量以及规范的图书标引质量,本文以CBKCI收录的1999—2009年出版的图书为研究样本,依据国家标准《GB/T 15834—2011 标点符号用法》(下文简称《用法》),抽取书名中的标点符号,探究其对图书被引的影响。考虑到CBKCI的引文尚未提供检索,本文的被引数据检索自CSSCI,具体研究问题如下:

(1) 书名中有无标点符号、标号与点号以及符号的数量对被引的影响。

(2) 书名中不同类型的标点符号(单个标点符号以及符号组合)对被引的影响。

(3) 书名中标点符号所表征的句式(复合式标题)对被引的影响。

2 相关研究

2.1 符号对被引的影响

(1) 标点有无。多项研究探讨了标题中的非字母数字型符号对论文被引的影响,部分研究发现了它的积极意义。例如, Gnewuch等^[8]基于经济学领域1980—2015年发表的312879篇论文,探讨标题中的非字母数字型符号等特征对论文被引的影响,结果发现它能够显著增加论文的被引。Fumani等^[12]通过对科学计量领域650篇论文标题的研究发现,多数高被引论文的标题中包含标点符号,其中被引排在前30位的论文中,有23篇的标题中包含标点符号;排在前10位的论文中,标题出现标点符号的论文更是多达9篇。Buter等^[13]的研究也证实了整体上标题包含符号的论文影响力更大。但是也有部分研究得出了相反的结果,比如, Nair等^[14]探索了管理科学论文标题的长度、有无非字母数字型符号、结构等特征与被引频次之间的关系,发现非字母数字型符号对论文的被引频次呈现显著负性影响。

(2) 冒号。冒号是文献标题中频繁使用的标点符号之一。一些研究发现,标题中包含冒号的论文被引更高,比如,林佳瑜^[15]的研究显示,标题有冒号论文的下载量和被引频次均显著高于标题无冒号的论文。Jacques等^[9]对全科与专科医学(General and Specialist Medical)期刊的研究发现,高被引论文中标题包含冒号的比例高于低被引组。部分研究得到了相反的结论, Paiva等^[4]发现标题有冒号和连接符的论文被引较少。Jamali等^[2]通过对6本PLoS期刊中2172篇论文的研究发现,标题包含冒号其下载量和被引频次均较少。此外,还有部分研究认为标题中有无冒号对被引无显著影响。魏瑞斌^[16]以JASIST(*Journal of the Association for Information Science and Technology*)和*Scientometrics*这2本期刊1997—2013年发表的5375篇论文为研究对象,研究结果表明,标题有无冒号的两组论文被引差异不显著。Subotic等^[17]研究了心理学论文的标题长度、趣味性、冒号等特征与被引的关系,结论表明,有无冒号与论文的被引以及下载量无关。Van Wesel等^[18]发现冒号与被引间的关系具有学科差异,冒号对全科与内科医学(General and Internal Medicine)领域论文的被引存在显著的负面影响,在应用物理领域表现为显著的正效应,在社会学领域中的影响不显著。

(3) 问号。Ball^[19]对Scopus数据库中1966—2005年医学、生命科学以及物理学3个学科发表的

近2000万篇文献进行了分析,发现每个学科标题中出现问号的比例都呈显著上升趋势,40年间的增长率分别介于50%~200%。标题中问号比例的显著增长,引起了学界的关注,一些研究探索了标题包含问号的文章与被引间的关系。Gnewuch等^[8]对经济学论文的研究证实,标题包含问号等特定符号可以增加论文的被引;魏瑞斌^[6]的工作也发现了相似的结论。Jamali等^[2]将问号视为问题型标题的标志,发现问句型标题提高了下载量,但被引不仅未见增加,反而有一定比例的减少。Paiva等^[4]选择PLOS期刊和12本BMC期刊在2008年9月发表的论文为研究对象,采用非参数检验等方法探究论文标题有无问号等5项特征对开放存取论文的浏览数(view)及被引的影响,结果表明,该特征对前者未见显著影响,对后者的影响则表现为负性,与Gnewuch等^[8]的研究结论相反。

2.2 句式对被引的影响

目前,复合式标题尚没有统一的定义。例如,Haggan^[20]认为复合式标题(compound title)是指两个名词短语并列在冒号、句号、破折号(dash)等标点的两侧;Rostami等^[3]将其称为组合式标题(combination title),指通过连字符(hyphen)或冒号分隔一个句子不同想法(ideas)的标题;甄长慧等^[21]在对中文论文的研究中,将包含破折号或冒号的标题视为复合式标题;刘永芳^[22]认为,“凡是被冒号、问号或连字符等隔开的词组、短语或句子即视为复合式标题”。从符号使用上看,中文标题中带破折号的约占复合式标题的73.5%,包含冒号的仅占26.5%^[21]。英文中的情况恰恰相反,刘永芳^[22]的统计结果显示,英文标题中带冒号、破折号以及问号的分别占98.2%、1.2%和0.6%。多项研究表明,复合式标题可以给论文带来更多的被引。Rostami等^[3]对Addictive Behavior期刊中302篇论文的标题进行了分析,发现通过冒号或连字符连接的复合式标题被引明显更高。甄长慧等^[21]选取CSSCI的期刊论文对中文复合式标题进行了研究,结果表明,复合式标题所占比例较高,其被引率和下载率均明显提高(数据来自CNKI)。

2.3 述评

标题是文献的重要组成部分,学界围绕标题特征与论文被引之间的关系开展了系列研究。标点符号作为标题的特征之一,也吸引了学界的关注。从

研究对象来看,相关研究多以英文论文为主,对中文标题的研究较少,尚未见学术图书标题与被引间关系的研究。中文标题中标点符号的种类以及作用与英文有所区别,英文标题的相关研究成果仅可供借鉴,故对中文标题特别是书名中标点符号特征对被引影响有探索之必要。另一方面,根据中文标点的分类,标点符号首先可以分成标号与点号两大类,二者在功能上有较大差别,不同的标点符号及其组合在功能和语义层面也存在差别。现有研究中标点符号特征多聚焦于标点的有无以及冒号、问号等少部分符号,更多标点特征与被引的关系有待进一步阐明。最后,中文标点符号本身具有特定的功能,蕴涵特定的语义,而当前的研究主要将标点符号作为“特殊字符”看待,对其功能及语义上的分析尚显不足,故需要突破现有的多从“形”的层面对标点符号与被引的关系研究,有必要探析其背后的深层逻辑并进而分析其对被引的影响。

3 数据与方法

3.1 数据

CBKCI共收录了21个学科共4364册学术图书,所选图书具有较高的原创性和学术水平。鉴于CBKCI的引文数据尚未提供检索,本研究的被引数据检索自CSSCI。CSSCI是国内人文社科领域权威的引文数据库,其引文数据比较规范准确,时间跨度为1998年至今,考虑其1998年初创之年的数据不够完整,故本文选择CSSCI 1999—2014年的引文数据作为图书被引的统计数据源。计量分析中需要为学术图书留出充足的被引时间窗,故本文选取1999—2009年出版的2844册图书为研究对象。

3.2 数据清洗

本研究获取图书被引信息的方法参考了苏新宁教授^[23]的做法,即采用图书的题名、作者、出版年和出版社4个字段进行匹配,若CSSCI提供的这4个字段与CBKCI完全一致,即认为是该图书的被引。此外,由于CSSCI数据库部分数据缺失、错误使得其与CBKCI图书的元数据不完全一致,故本文进一步采用模糊匹配方法查询图书的被引信息,具体做法是:

(1) 题名: CBKCI标注的图书题名包含CSSCI的被引文献题名即匹配成功。

(2) 作者: CBKCI标注的图书作者包含CSSCI的被引文献作者即匹配成功。

(3) 年代：CBKCI标注的图书出版年与CSSCI的被引文献年代一致即匹配成功。

(4) 出版社：CBKCI标注的图书出版社包含CSSCI的被引文献出版社或者反之即匹配成功。

若4个字段中的3个字段模糊匹配成功，即作为备选数据集；最后，对备选数据集进行人工筛选。经过上述操作后，本研究共保留被引记录111693条，图书的平均被引频次为39.27，年均被引频次为3.66，零被引图书74册。

3.3 自变量与因变量

依据《用法》，本文选取了6个书名标点符号的特征，分别是书名中标点符号的有无、数量、标号与点号、不同标点符号的组合、疑问式和/或复合式标题。在“标号与点号”变量中，依据《用法》的规定，不同点号表示停顿长短有区别，具体到本文的4种点号，停顿由长至短依次为问号>冒号>逗号>顿号，固进一步选取停顿的不同组合作为特征。参考Haggan^[20]的定义，本研究中的复合式标题是指由冒号、句号以及破折号连接而成的标题，并按照标题前后两部分的长度，进一步将其分为“长：短”（标题第一部分的字符数大于第二部分）、“短：长”（标题第一部分的字符数小于第二部分）以及平衡（两个部分的字符数相同）3类^[14]。疑问式标题指包含问号的标题，详见表1。

表1 自变量信息

变量名称	编码	变量取值
标点符号有无	X_1	分2类：有；无
标号与点号	X_2	分4类：无；仅点号（停顿的不同组合）；仅标号；标号+点号
标点符号组合	X_3	分39类：无；冒号；冒号+顿号等
疑问式	X_4	分2类：疑问式；非疑问式
复合式	X_5	分2类：非复合式；复合式（分3类：（长：短）；（短：长）；平衡）
标点符号数量	X_6	书名中包含的标点符号数量

本文首先将总被引频次作为因变量，考虑到时间因素和半衰期的影响，分别引入年均被引频次和五年被引频次作为因变量。此外，上述被引频次指标仅从“量”的角度进行了考虑而忽略了施引文献的“质”，故本文引入被引质量指标，从“质”的方面辅助前者完善测度，选取总被引质量和五年被引质量进行量化（表2），计算方式如下：首先，将CSSCI的来源期刊划分为3个等级，第一级是CSSCI评选出的人文社会学科的31本一流期刊；第二级是1998至2018年间持续入选CSSCI的来源期刊；

剩下的期刊归为第三级，依据期刊等级将施引文献划分为相应的3个级别，采用 n_i 表示第 i 级施引文献的数量。其次，参考里克特量表的反向编码方法，本文为一、二和三级施引文献分别赋予权重 $w_1=3$ ， $w_2=2$ ， $w_3=1$ 。最后，施引文献数量与对应的权重的乘积之和即为被引质量得分，总被引质量（ Q_1 ）及5年被引质量（ Q_5 ）的计算公式为

$$\text{一年被引质量: } Q = \sum_{i=1}^{i=3} w_i \times n_i \quad (1)$$

$$\text{总被引质量: } Q_1 = \sum_{j=\text{出版年}}^{j=2014} Q \quad (2)$$

$$\text{五年被引质量: } Q_5 = \sum_{j=\text{出版年}+1}^{j=\text{出版年}+5} Q \quad (3)$$

表2 因变量信息

因变量名称	编码	变量取值
总被引频次	Y_1	出版年至2014年期间的总被引频次
年均被引频次	Y_2	总被引频次/(2014—出版年)
总被引质量	Y_3	公式(2)
五年被引频次	Y_4	出版第二年至第六年间的被引频次
五年被引质量	Y_5	公式(3)

3.4 分析方法

1) 非参数检验

由于引文数据的偏态分布，本文使用非参数检验方法判断书名的标点符号特征与图书被引间的关系。具体而言，针对标点有无等二分类变量，使用Mann-Whitney U检验判断两组图书被引是否存在显著差异；针对标点组合等多分类变量，使用Kruskal Wallis检验判断多组图书被引是否存在显著差异；采用Spearman相关分析判断书名中的标点数量与图书被引间的关系。

2) 回归分析

为了排除其他变量对图书被引的影响，本文首先将图书相关因素、作者相关因素等10个变量作为控制变量（为节约篇幅，变量定义详见文献[24]），将书名的标点符号特征作为自变量、将对数变换后的 $Y_1 \sim Y_5$ 作为因变量，采取“进入”策略，分别构建多元线性回归方程；然后剔除回归系数不显著的自变量，再次构建多元线性回归方程，进一步判断书名的标点符号特征对图书被引的“净效应”。

4 分析结果

4.1 描述性统计

在所选的2844册图书中，标题中不含标点的图

书为1915册, 占有所有图书的67.33%; 包含标点的图书有929册, 占图书总数的32.67%; 无标点的图书约为有标点图书的两倍。依据《用法》的规定, 常用的标点符号有17种, 在名称包含标点的图书中, 标题中共出现冒号等10种标点符号(见表3)。

标题中包含冒号的图书有834册, 占标题包含标点图书总数的89.77%, 占全体图书的29.32%, 是书名中出现最多的标点符号。《用法》中将行文中冒号的使用方法概括为5种, 但是出现在标题中的用法与行文有所不同, 例如, 标题中鲜少出现“称谓语或称呼语之后”这种用法。使用冒号的标题形如A: B, 暗含着冒号前(A)后(B)两个部分之间存在丰富的语义关系^[25], 例如, 王培新所著的《乐浪文化: 以墓葬为中心的考古学研究》一书, 冒号前的“乐浪文化”是该书的研究主题, 冒号后的“以墓葬为中心的考古学研究”则是描述该书的研究视角, 是对冒号前研究主题的补充说明。

顿号在标题中的使用频率仅次于冒号。顿号在标题中的作用是表示标题内部并列词语之间的停顿, 与行文中的用法没有显著差异。间隔号在21册图书的书名中出现, 出现频次不高但是功能却较为多样。在所选图书中, 间隔号的用法主要涉及3

类: 用于人名、书名与卷名以及并列词语之间的分隔, 其中用于并列词之间的分隔是最主要的用法, 其他符号见表3。

标点符号的不同组合共计38种(连接号、间隔号的不同用法记为不同组合), 书名中出现的标点符号组合方式十分集中。单独使用冒号的标题有665个, 占有题名包含标点图书的71.58%, 远远超过其他标点组合的使用频次(图1); 冒号与顿号、冒号与引号、顿号单独使用的频次均超过了20, 是标题中使用较为频繁的组合。此外, 从组合的多样性来看, 冒号与其他标点的组合多达22种, 是组合形式最多的标点; 连接号(J)、顿号、引号与其他标点的组合形式分别是9种、8种和7种, 位列2~4位。

4.2 非参数检验

4.2.1 标点有无

在题名有无标点符号的2组图书中, 标点组的年均被引频次、五年被引频次、五年被引质量的中位数均显著高于无标点组($p<0.05$); 两组图书的总被引频次及总被引质量的差异并不显著($p>$

表3 题名标点描述型统计表

标点符号	数量	功能	举例	编号
无	1915	/	中国金融控股公司发展模式研究	A
冒号(:)	834	用于总说性或提示性词语之后, 提示下文; 总结上文; 在需要说明的词语之后, 表示注释和说明	乐浪文化: 以墓葬为中心的考古学研究	B
顿号(、)	105	表示标题内部并列词语之间的停顿	媒介批评: 立场、范畴、命题、方式	C
间隔号 (·)	16	并列词语之间的分隔	情感社会学: 理论·历史·现实	D
	2	书名与篇(章、卷)名之间的分界	中国画艺术专史·人物卷	E
	3	外国人名或少数民族人名内部的分界	经济学家的道德追问: 亚当·斯密伦理思想研究	F
引号(“”)	49	标示需要着重论述或强调、具有特殊含义而需要特别指出的成分	(1)“话语”视角的文学问题研究 (2)中国劳动力流动与“三农”问题	G
书名号 (《》)	34	标明文字、声音、图像等作品, 书、卷、篇、刊物、报纸、文件等出版物名称	周秦《尚书》学研究	H
括号(())	28	标示注释内容或补充说明	(1)俄国汉学史(迄于1917年) (2)中国刑事诉讼法(第二修正案)学者拟制稿及立法理由: 条文、释义与论证	I
连接号 (—、~、-)	51	标示相关项目(如时间、地域等)的起止	中国经济的转型与增长: 1978~2008年的经验研究	J
	2	在复合名词中起连接作用	资源—环境—经济复合系统诊断预警方法与应用	K
	2	外来语内部的分合	IRM—KM 范式与情报学发展研究	L
破折号 (——)	1	副标题前, 标示对主标题的注释、补充说明	西方政治哲学——从古希腊到当代	M
逗号(,)	5	标题内部一般性停顿, 停顿小于分号大于顿号	内外均衡, 一体循环: 循环经济的经济学思考	N
问号(?)	3	主要表示句子的疑问语气	(1)欧洲联盟对外政策一体化: 不可能的使命? (2)回乡, 还是进城? 中国农村外出劳动力回流研究	O

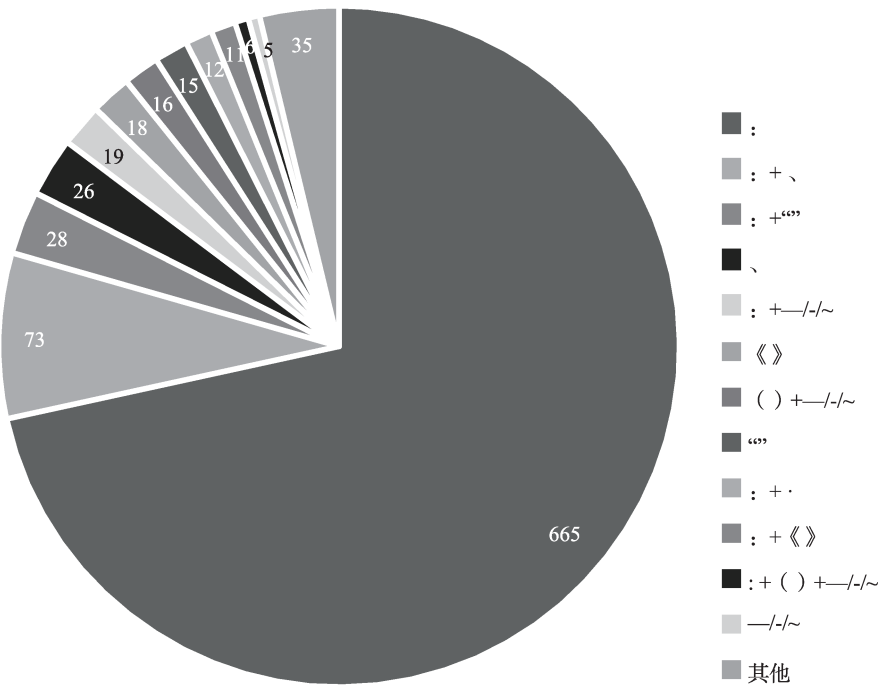


图1 标点组合比例

0.05) (表4)。可见, 书名中标点符号的使用在一定程度上提升了图书的影响力。每一个标点都有其独特的功能, 对图书被引的影响也可能存在差异, 故下文从不同标点功能的角度探究其与图书被引的联系。

4.2.2 点号与标号

《用法》规定, 10种标点符号(表3)中的问号、冒号、逗号以及顿号属于“点号”大类, 作用是点断, 主要表示停顿和语气; 其余6种标点符号

属于“标号”大类, 作用是标明, 主要标示某些成分(主要是词语)的特定性质和作用。依据这一分类体系, 将图书分为标题仅包含点号(27%)、仅包含标号(2.3%)、同时包含点号与标号(3.3%)以及不包含标点符号(67.3%)4组, 非参数检验结果表明, 4组图书的被引频次与质量存在显著差异($p<0.001$), $Y_1\sim Y_5$ 的中位数由高至低均依次为标题仅包含点号>无标点>=同时包含点号与标号>仅包含标号(表5)。

表4 “标点有无”非参数检验结果

标点有无	数目	Y_1	Y_2	Y_3	Y_4	Y_5
无	1915	27.00	2.46	44.00	12.00	20.00
有	929	27.00	2.75	47.00	14.00	23.00
Mann-Whitney 检验		874984.500	847136.500*	872656.000	836568.000*	838506.500*
总计	2844	27.00	2.50	44.00	12.00	21.00

注: *表示 $p<0.05$ 。

表5 “点号与标号”非参数检验结果

点号与标号	数量	Y_1	Y_2	Y_3	Y_4	Y_5
无	1915	27.00	2.46	44.00	12.00	20.00
点号	769	30.00	3.00	50.00	15.00	25.00
标号	65	10.00	1.27	14.00	6.00	10.00
点号+标号	95	22.00	2.25	38.00	11.00	20.00
Kruskal Wallis 检验		31.457***	31.876***	29.478***	37.412***	35.192***
总计	2844	27.00	2.50	44.00	12.00	21.00

注: ***表示 $p<0.001$ 。

鉴于书名包含点号图书的高被引特性,本文对点号做进一步的分析。非参数检验结果表明,书名包含不同停顿组合图书的总被引频次($p=0.453$)、年均被引频次($p=0.561$)、总被引质量($p=0.427$)、五年被引频次($p=0.465$)以及五年被引质量($p=0.462$)无显著差异。

4.2.3 句式

在所选图书中,书名是复合式标题的图书有836册,占29.4%,其中以冒号作为连接的有834册,破折号和问号各1册。由表6可见,书名是复合式标题的图书,其 $Y_1 \sim Y_5$ 均显著高于非复合式标题。复合式标题的(长:短)、(短:长)以及平衡3种类型分别有152册、63册和46册图书,3组图书的被引频次与质量没有显著差异($p>0.05$)。Nair等^[14]

的研究证实,(短:长型)、(长:短型)、平衡3种标题各占论文总数的44.8%、18.8%以及5.8%,发现仅在部分场景下平衡型标题可获得较多的被引。

4.2.4 标点组合

对不同组合的标点符号进行非参数检验。Kruskal Wallis 检验结果表明,标题中包含不同标点符号的图书,其 $Y_1 \sim Y_5$ 均存在显著差异(表7)。结合各组别的均值发现,标题中包含“冒号”、“冒号+顿号”2种方式的标题,其 $Y_1 \sim Y_5$ 的中位数均位居前列;与之相反地,标题中包含“书名号”及“其他”的图书,其 $Y_1 \sim Y_5$ 的中位数始终较低(表8)。此外,Spearman 相关分析的结果表明,书名中的标点数量与图书被引频次及质量之间无显著相关性($p>0.05$),与Buter等^[13]的研究结论一致。

表6 “标题句式”非参数检验结果

复合式标题	数量	Y_1	Y_2	Y_3	Y_4	Y_5
否	2008	26.00	2.42	43.00	12.00	20.00
是	836	29.00	2.87	49.00	14.00	24.50
Mann-Whitney U 检验		793083*	771270.50**	790755*	763257***	764335.5***
总计	2844	27.00	2.50	44.00	12.00	21.00

注:*表示 $p<0.05$,**表示 $p<0.01$,***表示 $p<0.001$ 。

表7 “标点组合”非参数检验结果

标点组合	标点符号	数量	Y_1	Y_2	Y_3	Y_4	Y_5
无	/	1915	27.00	2.46	44.00	12.00	20.00
B	:	665	30.00	3.00	51.00	15.00	25.00
B+C	:+、	73	29.00	3.20	51.00	15.00	24.00
B+G	:+“”	28	23.50	2.25	40.50	13.00	26.50
C	、	26	23.00	2.50	39.50	15.00	24.50
B+J	:+—/~/	19	23.00	2.29	40.00	8.00	16.00
H	《》	18	8.00	0.71	10.50	3.00	5.50
I+J	()+—/~/	16	22.50	2.00	41.00	10.00	18.50
G	“”	15	16.00	2.21	35.00	11.00	20.00
B+D	:+·	12	27.00	2.24	45.00	14.50	22.00
B+H	:+《》	11	27.00	2.54	51.00	10.00	18.00
其他	/	46	16.00	1.61	26.00	8.00	14.50
Kruskal Wallis 检验			39.249***	40.273***	38.062***	45.983***	44.385***
总计		2844	27.00	2.50	44.00	12.00	21.00

注:***表示 $p<0.001$ 。

4.3 回归分析

非参数检验的结果显示,书名中有无标点符号、不同标点的功能会对图书的被引产生影响,阮选敏等^[24]的研究还证实了图书被引受到书名长度等

多种非标点因素的影响。为了去除其他因素对标点符号特征的干扰,本文将阮选敏等^[24]研究中的10个变量作为控制变量,进一步探索标点特征对图书被引的影响。在初次构建回归方程后,发现自变量为 Y_1 及 Y_3 的方程中,回归系数不显著的自变量有:

表 8 “标点有无”与被引的回归分析结果

	Y_1	Y_2	Y_3	Y_4	Y_5
R^2	0.185	0.190	0.178	0.208	0.200
F 值	17.500***	17.722***	16.802***	19.650***	18.741***
有无标点	0.167***	0.177***	0.163***	0.169***	0.164***
控制变量					
书名长度	-0.228***	-0.213***	-0.221***	-0.189***	-0.184***
出版社是否百佳	0.124***	0.139***	0.118***	0.133***	0.126***
是否为丛书	0.062***	0.049**	0.068***	0.060***	0.067***
有无英文名称	—	0.041*	—	0.042*	0.038*
学科:哲学(基准组)					
教育学	0.045*	0.027	0.031	0.039	0.025
经济学	-0.023	0.010	-0.031	0.052*	0.033
历史学	-0.069**	-0.076***	-0.059**	-0.071**	-0.066**
社会学	0.097***	0.120***	0.092***	0.102***	0.097***
体育学	0.044*	0.046**	0.049**	0.067***	0.070***
图书馆情报与文献学	0.071***	0.082***	0.078***	0.110***	0.111***
艺术学	-0.143***	-0.133***	-0.150***	-0.154***	-0.164***
政治学	0.059*	0.069**	0.057*	0.083***	0.074**
中国语言学	-0.076**	-0.095***	-0.071**	-0.106***	-0.102***
宗教学	-0.103***	-0.109***	-0.094***	-0.11***	-0.106***
法学	0.115***	0.118***	0.111***	0.146***	0.137***
出版地:北京(基准组)					
华东	-0.012	-0.033	-0.010	-0.036*	-0.032
华北	-0.017	-0.024	-0.020	-0.034*	-0.035*
西北	-0.017	-0.024	-0.021	-0.048**	-0.050**
东北	-0.067***	-0.072***	-0.064***	-0.069***	-0.065***
上海	0.059***	0.044*	0.058**	0.044*	0.046**
国家和地区:中国(基准组)					
亚洲其他国家/地区	0.062***	0.072***	0.064***	0.060***	0.062***
其他西方国家	0.045**	0.052**	0.044*	0.041*	0.039*
美国	0.034*	0.040*	0.035*	0.023	0.022
项目类型:无(基准组)					
市厅级	-0.049**	-0.040*	-0.045*	-0.029	-0.030
省部级	-0.092***	-0.061***	-0.095***	-0.064***	-0.072***
国家级	-0.090***	-0.055**	-0.088***	-0.052**	-0.055**
(常量)	***	***	***	***	***

注：*表示 $p<0.05$ ，**表示 $p<0.01$ ，***表示 $p<0.001$ ；限于篇幅，未展示对 $Y_1\sim Y_5$ 均无显著影响的变量。

有无英文名称、作者人数及作者机构类型；自变量为 Y_2 、 Y_4 及 Y_5 的方程中，回归系数不显著的自变量有：作者人数及作者机构类型。故在对应方程中剔除上述系数不显著的自变量后得到最终的回归方程。

(1) 标点有无。各模型中变量的标准化系数如表 8 所示。回归分析的结果表明，有无标点对 $Y_1\sim Y_5$ 这 5 个因变量均存在显著的正性影响 ($p<0.001$)，

即书名中出现标点符号的图书被引显著高于书名为纯文字的图书。

(2) 点号与标号。将标号与点号转换为虚拟变量，以书名不包含标点符号的组为基准组，构建线性回归方程，发现书名仅包含点号或者书名同时包含点号和标号的图书，其被引更高；而书名仅包含标号的图书被引则低于基准组，但是差异并不显著（表 9）。

表9 “点号与标号”与被引的回归分析结果

	Y_1	Y_2	Y_3	Y_4	Y_5
调整后 R^2	0.194	0.198	0.187	0.217	0.208
F值	17.665***	17.748***	16.947***	19.745***	18.824***
点号+标号:无(基准组)					
点号	0.196***	0.204***	0.191***	0.197***	0.190***
标号	-0.033	-0.025	-0.032	-0.031	-0.030
标号和点号	0.101***	0.100***	0.103***	0.107***	0.108***
控制变量					
书名长度	-0.259***	-0.240***	-0.252***	-0.221***	-0.216***
出版社是否百佳	0.123***	0.138***	0.117***	0.132***	0.125***
是否为丛书	0.063***	0.050**	0.069***	0.061***	0.067***
有无英文名称	—	0.039*	—	0.040*	0.036*
学科:哲学(基准组)					
教育学	0.046*	0.028	0.032	0.040	0.026
经济学	-0.019	0.013	-0.027	0.056*	0.038
历史学	-0.066**	-0.072***	-0.055*	-0.068**	-0.063**
社会学	0.095***	0.117***	0.090***	0.100***	0.095***
体育学	0.044*	0.046**	0.049**	0.067***	0.070***
图书馆情报与文献学	0.077***	0.088***	0.084***	0.116***	0.117***
艺术学	-0.140***	-0.131***	-0.148***	-0.151***	-0.162***
政治学	0.067**	0.076***	0.064**	0.091***	0.082***
中国语言学	-0.069**	-0.089***	-0.064**	-0.099***	-0.095***
宗教学	-0.101***	-0.107***	-0.093***	-0.108***	-0.104***
法学	0.113***	0.117***	0.110***	0.145***	0.136***
出版地:北京(基准组)					
华东	-0.018	-0.039*	-0.016	-0.043*	-0.039*
华北	-0.018	-0.025	-0.021	-0.035*	-0.036*
西北	-0.018	-0.025	-0.022	-0.049**	-0.052**
东北	-0.064***	-0.070***	-0.061***	-0.066***	-0.062***
上海	0.053**	0.039*	0.052**	0.038*	0.041*
国家和地区:中国(基准组)					
亚洲其他国家/地区	0.061***	0.070***	0.062***	0.058***	0.061***
其他西方国家	0.043*	0.050**	0.041*	0.039*	0.037*
美国	0.033	0.040*	0.035*	0.023	0.022
项目类型:无(基准组)					
市厅级	-0.051**	-0.042*	-0.047**	-0.030	-0.032
省部级	-0.094***	-0.063***	-0.097***	-0.066***	-0.073***
国家级	-0.087***	-0.052**	-0.085***	-0.050**	-0.052**
(常量)	***	***	***	***	***

注: *表示 $p<0.05$, **表示 $p<0.01$, ***表示 $p<0.001$; 限于篇幅, 未展示对 $Y_1\sim Y_5$ 均无显著影响的变量。

(3) 复合式标题。将书名为复合式标题记为1, 非复合式标题记为0, 构建多元线性回归方程, 结果表明书名为复合式标题的图书, 其 $Y_1\sim Y_5$ 均显著高于非复合式标题的图书(表10)。

(4) 标点组合。将标点符号类别转换为虚拟变

量后代入线性回归方程, 将样本数小于10的所有组合合并为“其他”类别, 选取无标点符号为参照组, 回归分析结果如表11所示。由分析结果可知, 相较于书名为纯文字的图书, 题名包含冒号、冒号+顿号、冒号+引号、冒号+连接号对5个因变量

表 10 “复合式标题”与被引的回归分析结果

	Y_1	Y_2	Y_3	Y_4	Y_5
调整后 R^2	0.191	0.196	0.185	0.213	0.205
F 值	18.222***	18.291***	17.493***	20.259***	19.327***
复合式标题	0.206***	0.207***	0.202***	0.201***	0.196***
控制变量					
书名长度	-0.256***	-0.234***	-0.249***	-0.212***	-0.207***
出版社是否百佳	0.123***	0.138***	0.117***	0.132***	0.125***
是否为丛书	0.063***	0.049**	0.069***	0.061***	0.067***
有无英文名称	—	0.041*	—	0.042*	0.038*
学科:哲学(基准组)					
教育学	0.046*	0.027	0.032	0.04	0.026
经济学	-0.018	0.014	-0.026	0.056*	0.038
历史学	-0.064**	-0.072**	-0.054*	-0.067**	-0.061**
社会学	0.096***	0.118***	0.091***	0.101***	0.096***
体育学	0.046**	0.048**	0.051**	0.069***	0.072***
图书馆情报与文献学	0.075***	0.086***	0.082***	0.114***	0.115***
艺术学	-0.141***	-0.132***	-0.149***	-0.152***	-0.162***
政治学	0.065**	0.074**	0.062**	0.088***	0.079***
中国语言学	-0.070**	-0.091***	-0.065**	-0.101***	-0.097***
宗教学	-0.100***	-0.106***	-0.092***	-0.107***	-0.103***
法学	0.115***	0.118***	0.111***	0.146***	0.137***
出版地:北京(基准组)					
华东	-0.016	-0.036*	-0.014	-0.040*	-0.036*
华北	-0.018	-0.025	-0.020	-0.035*	-0.036*
西北	-0.018	-0.025	-0.022	-0.049**	-0.052**
东北	-0.065***	-0.070***	-0.062***	-0.067***	-0.062***
上海	0.055**	0.040*	0.054**	0.040*	0.042*
国家和地区:中国(基准组)					
亚洲其他国家/地区	0.061***	0.070***	0.062***	0.059***	0.061***
其他西方国家	0.043*	0.051**	0.042*	0.039*	0.038*
美国	0.033	0.040*	0.034*	0.023	0.022
项目类型:无(基准组)					
市厅级	-0.049**	-0.041*	-0.046**	-0.029	-0.031
省部级	-0.091***	-0.06***	-0.094***	-0.063***	-0.071***
国家级	-0.083***	-0.049**	-0.081***	-0.046**	-0.049**
(常量)	***	***	***	***	***

注：*表示 $p<0.05$ ，**表示 $p<0.01$ ，***表示 $p<0.001$ ；限于篇幅，未展示对 $Y_1\sim Y_5$ 均无显著影响的变量。

产生显著的积极影响；题名包含顿号对除总被引质量外的 4 个因变量均产生积极影响；题名包含书名号则对图书被引的影响呈负性；其他组合对 5 个因变量的影响不显著。

5 讨 论

5.1 标点有无

本研究发现，相较于纯文字标题，书名包含标

点符号的图书被引质量和被引频次均更高。部分有关论文标题的研究也得出了相似的结论，如 Gnewuch 等^[8]以及 Buter 等^[13]的工作。题名中的标点特征提高图书被引的内在机理可以从以下几方面进行解释：①可见性提高。Mann-Whitney U 检验的结果表明，包含标点符号的标题长度显著大于不包含标点的标题（表 12），该结果意味着前者往往包含了更多的关键词信息，增加了其在数据库检索结果中的“可见性”与“可查性”^[26]。②美化功能。语

表11 “标点组合”与被引的回归分析结果

	Y_1	Y_2	Y_3	Y_4	Y_5
调整后 R^2	0.193	0.199	0.187	0.216	0.208
F 值	14.917***	15.125***	14.369***	16.636***	15.920***
标点符号:无(基准组)					
B :	0.180***	0.186***	0.176***	0.177***	0.172***
B+C :+ 、	0.101***	0.109***	0.101***	0.098***	0.096***
B+G :+“”	0.062***	0.070***	0.065***	0.075***	0.077***
C 、	0.035*	0.043*	0.032	0.041*	0.037*
B+J :+ —/ - /~	0.055**	0.051**	0.057**	0.050**	0.052**
H 《》	-0.054**	-0.050**	-0.054**	-0.046**	-0.049**
I+J (+) + —/ - /~	0.029	0.023	0.031	0.019	0.022
G “”	-0.011	0.000	-0.010	0.002	0.001
B+D :+ ·	0.028	0.023	0.026	0.032	0.030
B+H :+《》	0.024	0.021	0.025	0.020	0.023
其他	0.024	0.027	0.024	0.016	0.017
控制变量					
书名长度	-0.259***	-0.241***	-0.254***	-0.215***	-0.212***
出版社是否百佳	0.125***	0.140***	0.119***	0.133***	0.127***
是否为丛书	0.065***	0.051**	0.071***	0.062***	0.068***
有无英文名称	/	0.041*	/	0.042*	0.039*
学科:哲学(基准组)					
经济学	-0.026	0.006	-0.034	0.049*	0.031
历史学	-0.069**	-0.076***	-0.060**	-0.071***	-0.066**
民族学	-0.035	-0.036*	-0.035	-0.038*	-0.037*
社会学	0.093***	0.115***	0.088***	0.098***	0.092***
体育学	0.043*	0.045*	0.048**	0.065***	0.069***
图书馆情报与文献学	0.073***	0.084***	0.080***	0.112***	0.113***
艺术学	-0.145***	-0.135***	-0.153***	-0.156***	-0.166***
政治学	0.057*	0.067**	0.055*	0.082***	0.073**
中国语言学	-0.069**	-0.089***	-0.064**	-0.100***	-0.096***
宗教学	-0.102***	-0.108***	-0.094***	-0.110***	-0.106***
法学	0.110***	0.113***	0.106***	0.141***	0.132***
出版地:北京(基准组)					
华东	-0.018	-0.039*	-0.016	-0.042*	-0.039*
华北	-0.018	-0.025	-0.021	-0.035*	-0.036*
西北	-0.017	-0.024	-0.021	-0.048**	-0.051**
东北	-0.063***	-0.069***	-0.060***	-0.065***	-0.061***
上海	0.054**	0.040*	0.053**	0.039*	0.042*
国家和地区:中国(基准组)					
亚洲其他国家/地区	0.061***	0.072***	0.063***	0.059***	0.062***
其他西方国家	0.043*	0.050**	0.042*	0.039*	0.037*
美国	0.034*	0.040*	0.035*	0.023	0.023
项目类型:无(基准组)					
市厅级	-0.050**	-0.041*	-0.046**	-0.030	-0.031
省部级	-0.091***	-0.060***	-0.093***	-0.064***	-0.071***
国家级	-0.086***	-0.051**	-0.083***	-0.049**	-0.051**
(常量)	***	***	***	***	***

注: *表示 $p < 0.05$, **表示 $p < 0.01$, ***表示 $p < 0.001$; 限于篇幅, 未展示对 $Y_1 \sim Y_5$ 均无显著影响的变量。

表12 标题长度与“标点有无”的非参数检验

有无标点	数量	书名长度
无	1915	8.00
有	929	18.00
Mann-Whitney U 检验		94198.000***
总计	2844	10.00

注：*表示 $p<0.05$ 。

言学者郭攀^[27]的研究表明，标点符号除了《用法》中阐明的“用于表示停顿、语气以及词语的性质和作用”这些共性的功能之外，还具有美化功能；陈望道^[28]的研究也指出，“标点可以神文字之用”。在已经到来的信息过载时代，每个人的信息承载量和记忆量都相当有限，欲使图书标题给读者留下深刻的印象，作者拟定标题时可以综合使用标点符号与文字以避免标题的同质化。包含标点符号的标题从形式层面来看，做到了惯习（convention）与惊喜（surprise）之间的平衡^[13]，因此在标题中“灵活地运用标点符号，往往可以使读者产生听觉或视觉上的联想，突出语意，增强节奏，使语言表达得更为生动、形象、鲜明”^[29]，为平凡的文字增加“惊喜”，吸引读者更多的关注。③语义增强。标题中的标点符号还在语义层面补充了“文字未能表达或难以表达的抽象思维性以及情绪性兼具的停顿、语气、语调等”，对“具象思维性以及情绪性内容进行了必要的添加性表达”，对“部分抽象思维性内容中认识深化的伴生性内容进行了标记”^[27]，而这些功能在未使用标点符号的纯文字标题中难以实现，即标点符号丰富了标题的语义。

上述3个原因可以解释Jamali等^[2]、林佳瑜^[15]研究中发现的该类论文可以获得更高下载率的结论，而谢娟等^[30]采用元分析证实了下载率与被引量间的正相关关系，即题名中标点的使用借助于下载、浏览等行为提高了论著被引的概率。当研究者详细地阅读论著之后的引用行为则更多地会受到研究主题、论文质量等因素的影响^[31]，此时标点符号的影响将显著下降，从而可以解释Nair等^[14]部分研究中发现的负相关或不相关结论。

5.2 标点组合

不同的标点符号及其组合对被引的影响存在较大差异。书名包含以下标点符号的图书被引较高：①仅包含点号。郭攀^[27]的研究表明，点号能够起到语义补充的作用；此外，具体的点号还有其独特的价值，比如，问号的使用容易抓住读者的注意力，

引起思考和阅读欲望；冒号、逗号及顿号3种句中点号的使用“不仅能使标题结构简洁、紧凑、表意明确，而且还具有较好的修辞效果”^[32]。②书名包含冒号或与其他部分符号的组合。以论文为对象的多项研究也证实了该结论，如Rostami等^[3]、林佳瑜^[15]以及甄长慧等^[21]的工作。然而本文的研究还发现，冒号与另一部分符合组合时，如“冒号+书名号”等，图书被引与标题为纯文字的图书无显著差异。故对于标题包含冒号的图书被引情况不能一概而论，需要综合考虑其他标点的功能及用法对被引的影响。③标题包含顿号。在标题中，顿号的作用主要是分隔标题中的并列词语，这些“并列词语在结构上都是平等的但是在意义关系上丰富多彩，有时会有逻辑或时间等顺序”^[25]。例如，《资本结构、治理结构和代理成本：理论、经验和启示》中的“理论”、“经验”、“启示”通过顿号和连词“和”连接的3个词语之间结构平等且具有内在的逻辑性。并列词语既能简洁、凝练、准确地概括图书主旨，整齐的格式又可以加强标题的语气，吸引读者的注意力，从而让图书的标题更具魅力。此外，在标题中，间隔号（D）的作用与顿号相似，比如，《敦煌道教文献研究：综述·目录·索引》中通过间隔号分隔并列词语“综述”、“目录”和“索引”，且三者之间具备写作顺序的内在关系。与顿号相比，间隔号更具标题特色，且间隔号由于采用的是圆点，居中，因此特别能够造成一种视觉上的“平衡美”^[33]，因此标题包含间隔号的两本图书被引也较高，但是由于样本较少，没有进行进一步的比较分析。

与之相反，本文研究发现书名中包含书名号时，图书的被引较低。Costello等^[34]对论文标题的研究发现，论文被引与其标题中出现的分类宽度及地理位置宽度正相关，标题范围较广的论文被引显著高于标题范围较窄的论文。与这一机理相似，从标点符号的功能来看，本研究中连接号的作用是标示图书内容的时间特征，书名号表示图书的研究对象为一本著作，即当出现上述标点符号时，往往表示对该书具体内容的展示，表明该图书研究的内容在主题分类体系中处于叶节点，该类型标题在一定程度上限制了图书的受众，即只有对标示的研究时间或研究对象感兴趣的读者才会进一步阅读进而引用。

5.3 复合式标题

20世纪60年代起，复合式标题在很多研究领

域的期刊中急剧增加^[6], Dillon^[35]、Lewison等^[5]工作也发现了相同现象。该现象吸引了文献计量学者的目光,甄长慧等^[21]的研究发现了复合式标题的被引和下载量相对较高。本研究也发现主要通过冒号表征的复合式标题的被引显著高于非复合式标题,语言学等学科的研究为该现象提供了多维度的解释。

从形式维度看,作者通常使用复合标题简洁地表达构成标题的两部分之间的相互关系^[36],两部分通常由(研究主题:研究方法)、(研究问题:对应的答案)、(较广的研究范围:具体的研究主题)^[14]等形式组成。复合式标题中出现最多的标点是冒号,冒号前是主题、冒号后是题旨的形式符合英语标题的基本格式,因而广受英美学者的欢迎^[22]。冒号具有提示引出的作用,可用在需要解释的词语后边,引出解释或说明,既引人注目,又节省词语,巧妙地概括出丰富的内容^[25]。从内容角度来看,复合式标题“既有主题,又有对主题的解释、说明、补充等,包含的信息量更大”^[37]。部分学者甚至认为冒号是学术性的体现,是研究不断深化的结果^[35]。最后,从吸引力的角度看,复合式标题具有明显的“视觉效果”^[38],标题中标点符号的停顿功能具有强调修辞的作用,如冒号^[22]。综上所述,标题的复合式结构让标题更加清晰、准确、信息量也更丰富,故容易获得更多的引用。

6 启示

学界对标题的功能和作用进行了广泛的讨论。Lewison等^[5]认为标题的主要意图是让读者被吸引和知情。Jamali等^[2]在研究中提到一个理想的标题应当简短、信息量丰富且有吸引力。可见,多数学者认同标题应当具备吸引力、简洁、信息量丰富等^[2,5]。但是正如 Kane^[39]所说,在实际操作中,平衡标题的这些特质十分困难;多数情况下,标题的吸引力与信息量具有“跷跷板”效应。故长期以来,标题的撰写都是一项重要且有难度的任务。本文探究了书名中标点符号特征对图书被引的影响,分析结果证实了合理应用特定标点符号可以让书名在吸引力、信息量、简洁性等方面有所改善,从而提升图书的影响力。为使图书有一个夺目的标题,参考本研究的结论,在书名设计上有以下启示:

(1) 善用复合式标题。形如 A: B, 以冒号、破折号、问号等标点符号连接的复合式标题是学术

界的“宠儿”,从信息量的角度,复合式标题能够巧妙地概括出丰富的内容^[25],让标题更加清晰、准确,信息量也更丰富^[37]。从简洁性的角度,复合式标题既可以节省词语,省略 A 和 B 之间的连接词^[25],构成一种形式上的直观简洁。从吸引力的角度,冒号的停顿功能具有强调修辞的作用^[22],具有明显的“视觉效果”^[38]。最后,从使用的灵活性上,利用复合标题可以简洁地表达构成标题的两部分之间的相互关系^[36],前后两个部分之间具有丰富的语义关系,包括(研究主题:研究方法)、(研究问题:对应的答案)、(较广的研究范围:具体的研究主题)^[14]等多种形式。复合式标题的流行正是得益于此类标题可以在一定程度上兼顾信息性、简洁性和吸引力,在应对研究内容十分复杂的学术图书时是一个很好的选择。但是,为了追求“流行”或者“高被引”而盲目使用复合式标题的行为是不可取的,这将不利于科学的进步^[40]。

(2) 巧用并列型标题。顿号在标题中表示并列词语之间的停顿,这些并列词语“在结构上都是平等的但是在意义关系上丰富多彩,有时会有逻辑或时间等顺序”^[25]。间隔号在标题中也用于并列词语之间的分隔,这种用法仅限在标题中使用。结合本文的研究可见,标题中包含被顿号分隔的并列词语的图书,具有较大的影响力。从信息量的角度来看,被分隔的并列词语通常是对图书内容的高度凝练,对反映图书主旨有重要的作用。从艺术造诣的角度来看,两项或多项结构平等词语并列,犹如行文中的排比句式,凸显标题的整齐连贯,增加标题的语气和气势,此外间隔号的使用会营造一种视觉上的“平衡美”^[33],从而达到吸引读者的目的。

(3) 慎用限定类标题。通过对低被引组的分析,在书名的撰写过程中,应当避免出现指代范围过于具体的词语。本文的研究发现,当书名中只包含书名号时,图书的被引较低。具体来说,书名号在图书中的作用是标示图书等出版物的名称,如《周秦〈尚书〉学研究》。当书名中出现出版物名称等过于具体的信息时,无意间将图书内容限制在了一个很狭小的范围内,无形之中限制了受众的范围,该现象与论文标题中出现特定国家、地区等地理名词时,被引较低的内在机理相似^[9]。所以在拟定书名时,对出版物名称等描述过于细致的词语需要慎重。若这些词语十分重要,不妨考虑采用复合式标题,在这些词语之前加上对图书主旨的描述。

7 结 论

本文以 CBKCI 收录的 1999—2009 年出版的 2844 册图书为研究样本,探讨书名中的标点符号特征对图书被引的影响,以期通过改进书名的质量提高其影响力。本文的主要结论如下:首先,相较于纯文字的书名,在书名中合理地使用标点符号有利于提升图书的影响力,书名仅包含点号或者同时包含点号和标号的图书被引较高,标点符号数量对图书被引没有显著影响。其次,本文发现带有冒号的标题,将书名分为具备一定语义关系的两部分,由于其清晰、准确、信息量丰富等特征能够增加图书的被引;书名中使用顿号分隔的并列词语,可以在形式上、语义上、逻辑关系上更好地概括图书主旨,兼备信息量与吸引力的特点,从而让图书获得更高的引用。与之相反,书名中仅包含书名号时,其被引较低,也即当书名中出现具体的概念时会降低图书的被引。第三,本研究证实了书名是复合式标题的图书比非复合式标题的图书被引更高,但是(长:短)、(短:长)以及平衡3种书名结构的图书被引不存在显著差异。最后,冒号是图书标题中使用最多的标点符号,其使用频率高且使用范围广,颇受图书作者的欢迎。

本文的局限主要体现在两个方面:一方面,由于样本量的限制,部分标点出现的频率较低,影响了研究的深入;另一方面,本文的引文数据仅来自 CSSCI 收录的论文,缺少图书之间的引用,待 CBKCI 正式使用之时可再进行全面探索性研究。

参 考 文 献

- [1] Sagi I, Yechiam E. Amusing titles in scientific journals and article citation[J]. *Journal of Information Science*, 2008, 34(5): 680-687.
- [2] Jamali H R, Nikzad M. Article title type and its relation with the number of downloads and citations[J]. *Scientometrics*, 2011, 88(2): 653-661.
- [3] Rostami F, Mohammadpoorasl A, Hajizadeh M. The effect of characteristics of title on citation rates of articles[J]. *Scientometrics*, 2014, 98(3): 2007-2010.
- [4] Paiva C E, Lima J P D N, Paiva B S R. Articles with short titles describing the results are cited more often[J]. *Clinics*, 2012, 67(5): 509-513.
- [5] Lewison G, Hartley J. What's in a title? Numbers of words and the presence of colons[J]. *Scientometrics*, 2005, 63(2): 341-356.
- [6] 姜亚军. 复合式论文标题之争: 评述与启示[J]. *外语教学*, 2010, 31(6): 29-32.
- [7] Wang Y, Bai Y. A corpus-based syntactic study of medical research article titles[J]. *System*, 2007, 35(3): 388-399.
- [8] Gnewuch M, Wohlrabe K. Title characteristics and citations in economics[J]. *Scientometrics*, 2017, 110(3): 1573-1578.
- [9] Jacques T S, Sebire N J. The impact of article titles on citation hits: An analysis of general and specialist medical journals[J]. *JRSM Short Reports*, 2010, 1(1): 2.
- [10] 潘云涛. 中国科技论文统计与分析[M]. 北京: 科学技术文献出版社, 2005: 104.
- [11] 苏新宁. 中国人文社会科学学术影响力报告(2000—2004)[M]. 北京: 中国社会科学出版社, 2007: 8.
- [12] Fumani M R F Q, Goltaji M, Parto P. The impact of title length and punctuation marks on article citations[J]. *Annals of Library and Information Studies*, 2015, 62(9): 126-132.
- [13] Buter R K, Van Raan A F J. Non-alphanumeric characters in titles of scientific publications: An analysis of their occurrence and correlation with citation impact[J]. *Journal of Informetrics*, 2011, 5(4): 608-617.
- [14] Nair L B, Gibbert M. What makes a 'good' title and (how) does it matter for citations? A review and general model of article title attributes in management science[J]. *Scientometrics*, 2016, 107(3): 1331-1359.
- [15] 林佳瑜. 论文标题与下载和引用的关系[J]. *大学图书馆学报*, 2012(4): 14-17.
- [16] 魏瑞斌. 论文标题特征与被引的关联性研究[J]. *情报学报*, 2017, 36(11): 1148-1156.
- [17] Subotic S, Mukherjee B. Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles[J]. *Journal of Information Science*, 2013, 40(1): 115-124.
- [18] Van Wesel M, Wyatt S, Ten Haaf J. What a difference a colon makes: How superficial factors influence subsequent citation[J]. *Scientometrics*, 2014, 98(3): 1601-1615.
- [19] Ball R. Scholarly communication in transition: The use of question marks in the titles of scientific articles in medicine, life sciences and physics 1966 - 2005[J]. *Scientometrics*, 2009, 79(3): 667-679.
- [20] Haggan M. Research paper titles in literature, linguistics and science: Dimensions of attraction[J]. *Journal of Pragmatics*, 2004, 36(2): 293-317.
- [21] 甄长慧, 曹凤龙, 郭书法, 等. CSSCI 期刊高、低被引论文标题特征对比研究[J]. *科技与出版*, 2014(10): 113-116.
- [22] 刘永芳. 基于语料库的英汉科技核心期刊论文复合式标题的结构特点[J]. *编辑之友*, 2012(2): 97-99.
- [23] 苏新宁. 我国人文社会科学图书被引概况分析——基于 CSSCI 数据库[J]. *东岳论丛*, 2009, 30(7): 5-13.
- [24] 阮选敏, 吕冬晴, 张培, 等. 我国人文社会科学学术图书被引影响因素研究[J/OL]. *图书馆论坛*, 2019: 1-13. [2018-11-29]. <http://www.cnki.net>

- //kns.cnki.net/kcms/detail/44.1306.G2.20180914.1817.002.html.
- [25] 尹世超. 标题语法[M]. 北京: 商务印书馆, 2001: 188, 192-194.
- [26] Shekhani H N, Shariff S, Bhulani N, et al. Bibliometric analysis of manuscript characteristics that influence citations: A comparison of six major radiology journals[J]. American Journal of Roentgenology, 2017, 209(6): 1191-1196.
- [27] 郭攀. 二十世纪以来汉语标点符号研究[M]. 武汉: 华中师范大学出版社, 2009: 194-195.
- [28] 陈望道. 陈望道语言学论文集[M]. 北京: 商务印书馆, 2009: 3
- [29] 兰宾汉. 标点符号运用艺术[M]. 北京: 中华书局, 2006: 18
- [30] 谢娟, 龚凯乐, 成颖, 等. 论文下载量与被引量相关关系的元分析[J]. 情报学报, 2017, 36(12): 1255-1269.
- [31] Tahamtan I, Afshar A S, Ahamdzadeh K. Factors affecting number of citations: a comprehensive review of the literature[J]. Scientometrics, 2016, 107(3): 1195-1225.
- [32] 朱五四. 英文标题中标点符号的用法[J]. 解放军外语学院学报, 1995(6): 21-24.
- [33] 陈佳璇. 间隔号用法的发展和统一: “专名化”与“均衡美”[J]. 语言文字应用, 2011(2): 22-26.
- [34] Costello M J, Beard K H, Primack R B, et al. Are killer bees good for coffee? The contribution of a paper's title and other factors to its future citations[J]. Biological Conservation, 2019, 229: A1-A5.
- [35] Dillon J T. The Emergence of the Colon: An Empirical Correlate of Scholarship[J]. American Psychologist, 1981, 36(8): 879-884.
- [36] Soler V. Writing titles in science: An exploratory study[J]. English for Specific Purposes, 2007, 26(1): 90-102.
- [37] 张树铮. 标题中冒号的新用法及其规范问题[J]. 语言文字应用, 2003(4): 105-111.
- [38] Perrie Y. Colonic irrigation or irritation: It is all in the title[J]. CRS Newsletter, 2008, 25(2): 2.
- [39] Kane S T. The Oxford essential guide to writing[M]. Berkley Books, 2000: 58.
- [40] Ramos-Mrosovsky C. Escaping the colon: A paper title pitfall[N]. The Daily Princetonian, 2003-05-05.

(责任编辑 魏瑞斌)