

学术论文的可读性与影响力研究^{*}

——基于人文社会科学学科差异的视角

伍丹炜¹ 刘宇²

(1. 云南大学历史与档案学院, 昆明 650091; 2. 云南大学新闻学院, 昆明 650504)

摘要: 可读性是论文质量的重要特征之一, 影响着读者对研究内容的认知与理解。探索文本可读性与论文影响力之间的关系对学术知识的有效传播有着重要的现实意义。以2016—2020年由图书情报学、新闻传播学以及管理学3个学科代表性期刊发表的论文为样本, 研究摘要可读性与全文可读性对被引频次的影响, 并检验学科之间的差异。研究发现: 摘要易读而全文难读的论文更容易成为高被引论文, 学科作为调节变量对摘要可读性及被引频次的关系、全文可读性和被引频次的关系均有显著的影响。作者应在保证严谨精确表达的基础上, 尽可能提升论文的可读性, 以促进学科知识的交流与发展。

关键词: 文本可读性; 学术传播; 被引频次; 学科差异

中图分类号: G250 **DOI:** 10.3772/j.issn.1673-2286.2023.09.003

引文格式: 伍丹炜, 刘宇. 学术论文的可读性与影响力研究: 基于人文社会科学学科差异的视角[J]. 数字图书馆论坛, 2023 (9): 20-29.

学术论文作为传播研究成果的知识载体, 肩负着促进知识发展、引导社会思潮等重要使命。在学术研究与社会不断互动重构的背景下, 研究成果的传播范围已逐渐从相关领域的专业人士拓展至社会上对严肃文本有浓厚兴趣的业余人群。这无疑提高了对研究人员写作水平的要求, 即在撰写论文的过程中提高文本的可读性, 让读者更容易理解作者知识框架的构建过程、内容以及研究成果的社会价值^[1]。然而, 可读性作为评价论文质量的一个指标尚未受到学界的普遍支持与认可, 但它对知识传播的实质性影响不容忽视。本研究将借助荆溪昱等^[2]的可读性公式分别计算论文摘要及全文的可读性, 并利用被引频次作为测量论文传播效果的指标进行回归检验, 以探索学术文本可读性对论文知识传播效果的影响, 为提高学术交流的有效性提供新的视角及建议。

1 文本可读性相关研究

通常而言, 研究内容能满足读者的阅读需求是论文获得认可与引用的前提。高影响力的论文不仅要有契合时代发展的研究话题, 其文字表述还要通畅、所传达的内容信息要能使读者产生共鸣。为衡量阅读材料的难易程度, 学界已经产出大量有关可读性测评工具的研究。

1.1 可读性工具的研发

可读性又称为易读性, 是指文本材料所包含的影响读者有效使用的所有要素及其之间相互作用的总和^[3], 用于反映信息易于被发现、被接受和被理解的程度^[4]。20世纪20年代, 美国教育界为衡量教材难度、找到适合

收稿日期: 2023-07-05

^{*}本研究得到国家社会科学基金一般项目“中国人文社会科学领域的学术诚信观念与行为研究”(编号: 20BTQ019)资助。

当时中学生学习的教材,开始关注教材的可读性研究^[5]。在现实需求的推动下,针对英语文本的可读性公式逐渐走入大众视野。Vogel等^[6]引入桑代克词表,将学生的阅读成绩作为因变量,通过逐步回归的方法研制出明确文本难度的可读性公式,该公式包含4个预测变量,即词表中的不同词语数、词表中的介词数、词表之外的词数及75个样本句中的简单句数。随后,多元线性回归成为研制可读性公式的重要统计方法,大量科学、客观的可读性公式在Vogel等的研究基础上不断涌现,如Flesch Reading Ease、SMOG Grading、Flesch-Kincaid Formula等,它们被广泛应用于教材课本、国家政策等阅读材料的可读性测量。随着研究的深入,这些公式的研究范式逐渐趋于统一,主要通过测量词汇、句法、篇章等文本特征衡量文本可读性^[7]。20世纪90年代后,计算机的应用为文本可读性公式的深入研究提供了技术支持。Crossley等^[8]通过自然语言处理工具提取与文本理解和阅读速度相关的语言特征,开发出了预测文本可读性的新模型,并通过对比预测结果发现新模型比经典可读性公式解释更准确。此外随着大规模语料库的建设与深入研究,一系列自动化的在线分析工具如Coh-Metrix^[9]等逐渐产生,降低了人工计算带来的误差概率与成本。

英语可读性工具的研制带动了其他语言的相关研究。尽管汉语、日语这类语言的结构与英语差异较大,但是一些学者仍根据这些语言自有的特点展开了研究。例如,Yang^[10]以繁体文本为样本,用难词比、完整句子数及平均笔画数3个预测变量研制出了第一个汉语文本的可读性公式。荆溪昱等^[2]以台湾地区中小学12个年级的中文教材为研究样本,选用课文长度、平均句长、常用字占比、诗歌文体及文言文文体来拟合每个学期的教材难度,其提出的公式被认为是当前汉语可读性公式中最权威的可读性公式^[11]。由于汉语可读性的相关研究公式起步较晚,且汉语与英语的语言特点截然不同,当前汉语文本可读性研究仍以研制可读性公式为主^[6],例如:郭望皓^[12]通过问卷调查的方法,筛选出平均句长、汉字难度以及词汇难度3个对外汉语文本难度的影响因素,并通过CRITIC加权赋值法计算了各因素的权重系数,拟合出对外汉语文本的可读性公式;李萍^[13]则对荆溪昱等的公式进行创新,将专业术语作为衡量指标引入汉语可读性计算公式。直至2020年,针对汉语的可读性公式有将近30个^[14]。

1.2 可读性工具的应用

当前有关文本可读性的研究除可读性工具开发等理论研究外,还有一部分研究集中在可读性工具的应用上。初始阶段,可读性工具的应用主要集中在教育学领域对教材难易程度的预测方面,由于阅读在社会生活中的重要作用,可读性工具的应用逐渐扩展至商业、医疗等诸多领域^[15]。逯东等^[16]以2007—2016年A股上市公司为研究对象,发现年报可读性越低的上市公司面临的融资约束越强。柯青等^[17]构建具有高/低可读性的两组健康信息网页作为实验刺激材料,通过眼动实验检验了在线健康信息可读性对信息加工绩效的影响效应,发现可读性可以不同程度作用于总注视时间、注视点个数以及首次进入时间3个认知负荷因素,从而间接地影响信息加工绩效。

随着可读性工具在各领域的广泛应用,一些学者也开始关注文本可读性与论文影响力之间的关系。早期,学者主要通过外显的定性评价特征来评估论文的影响力。Hartley等^[18]研究了心理学领域文本可读性及其影响力之间的关系,通过比较经典论文与对照论文(同刊同期论文)的可读性发现,文本可读性正向影响着论文的重要性排序,但与其被引频次之间没有显著关系。Sawyer等^[19]以4种市场营销期刊上的论文为研究对象,通过对比获奖/未获奖论文的文本可读性来判断文本可读性与论文影响力的关系,发现获奖论文比未获奖论文更具可读性。随着引文分析技术的普及,被引频次作为衡量论文影响力的指标被广泛应用。Van Wesel等^[20]使用FRES可读性公式分别对社会学、普通内科和应用心理学论文的摘要、全文进行可读性测量,发现在社会学中全文可读性显著负向影响被引频次、在普通内科学中摘要可读性显著负向影响被引频次、在应用心理学中全文可读性显著负向影响被引频次。Lei等^[21]使用了FRE和SMOG两个指标测量了情报学领域2003—2012年期刊论文的摘要、全文可读性与标准化被引频次之间的关系,发现可读性与被引频次之间的相关关系不显著。陈练文等^[22]则利用SMOG、FKG、ARI及FOG 4个可读性公式分别测试了计算机和图书情报学学科的摘要及全文的可读性,发现高被引期刊的可读性得分高于低被引期刊,这一点在摘要上表现得更为明显。Jin等^[23]采用39个词汇和句法复杂性指标来衡量2012年和2018年发表在Science上的550篇论文摘要的可读性,并采用另类注意力评分来衡量读者的在线注意力,结

果表明摘要的可读性与网络关注显著相关,且这种关系受学科和发表时间的显著影响。

目前文本可读性研究在国内仍是一个新兴领域,其中对文本可读性公式的探索主要集中在计算机语言学学科^[24];而在可读性工具的应用方面,鲜有学者关注到文本可读性与学术论文影响力之间的关系。基于此,本文运用现有的可读性工具,展开对汉语文本可读性与论文影响力关系的研究,以填补国内可读性相关研究的空缺。

2 研究设计

2.1 数据来源

以中国知网(CNKI)期刊全文数据库为数据源,分别选择图书情报学、新闻传播学以及管理学3个学科的代表性期刊(《中国图书馆学报》《情报学报》《新闻与传播研究》《国际新闻界》《管理世界》《南开管理评论》)所收录的论文作为样本。李樵^[25]的研究显示,新闻传播学与管理学是与图书情报学知识交流最为紧密的两个相关学科。使用这3个学科的典型论文作为研究样本不仅可以降低以某一学科论文为样本带来的“以偏概全”风险,能够更为全面地展示国内社会科学界文本可读性的特点,也可以比较不同学科之间文本可读性与论文影响力之间关系的差异。于2023年3月4日在CNKI数据库抓取2016—2020年6种期刊发表的论文作为研究样本,并采集了样本论文的作者信息、期刊信息、摘要文本、全文文本及被引频次等信息。剔除会议报道等非研究性的记录,并分别计算样本摘要及正文的可读性,最终得到3 214个研究样本。

2.2 研究方法

2.2.1 因变量

使用被引频次作为测量论文影响力的变量。随着数据库在学术领域的广泛应用,读者对研究成果的关注度可以通过被引频次更加真实客观地反映出来,这在一定程度上解决了定性评价数据收集成本高、评价结果片面等问题。

现有研究发现被引频次并非呈正态分布^[26-27],而作为典型的计数变量,被引频次在本研究中还存在过离

散现象(均值为58.71,方差为102.153),因此负二项回归模型更适用于此次分析。使用SPSS 24软件的广义线性模型拟合负二项回归。

2.2.2 自变量

理想状态下,被引频次被视为一种测量研究内容有用性的工具变量,可用于描述科学知识的扩散轨迹^[28]。实际的引文动机却很复杂,受到许多社会因素的影响。Wang等^[29]在设计预测高被引论文工具时提出,高引用的特征可以归纳为质量机制和可见性机制,其中可见性机制是指通过影响论文可见性影响论文被引频次的因素集合,例如在知名期刊上发表或由知名人士撰写的论文更容易被读者关注,从而增加了这类论文被引用的机会。基于此,设置了“期刊是否为权威期刊”以及“第一作者的工作单位级别”,分别控制了期刊声望及作者声望两个社会因素对被引频次的干扰。此外,由于被引频次是一个累积的变量,新发表的论文与发表年龄较长的论文相比存在被引劣势,还设置了论文年龄(截至2023年的论文发表时间)来控制这一影响。

尽管越来越多的研究发现论文外部的社会因素是被引频次的重要影响因素,但是学界实际更关心论文本身的知识内容是否可以预测论文未来的学术影响力^[30-31]。当前已有学者通过提取论文内容特征的方式来评价论文的内容质量。例如,Stremersch等^[32]使用论文在期刊目次中的顺序、是否获得期刊年度优秀论文奖以及论文长度(页数)作为论文质量的测量指标,发现在市场营销领域这3个测量指标与论文的被引频次显著相关。然而上述研究仅从论文的某些具体特征出发判断论文的内容质量,忽略了论文主要由文字构成的事实。论文的摘要及全文作为作者传达知识的核心载体,其文本内容往往是读者判断论文引用价值的重要因素。如前所述,由于荆溪昱等的可读性公式在学界的认可度较高,使用该可读性公式分别计算了样本论文摘要及全文的可读性,以期探索论文摘要及全文的可读性对其被引频次的影响。可读性公式如式(1)所示。

$$R=8.761\ 056\ 04+0.002\ 724\ 38\times X_1+0.078\ 667\ 82\times X_2-8.943\ 110\ 10\times X_3+0.429\ 201\ 82\times X_4+3.236\ 771\ 41\times X_5 \quad (1)$$

式中: R 为可读性数值; X_1 为文本总字数; X_2 为平均句长; X_3 为常用字比; X_4 为诗歌文体; X_5 为文言文^[10]。由于研究样本是学术论文,不存在诗歌文体和文

言文体, X_4 、 X_5 取值均为0。此外公式中的常用字是指台湾地区小学的495个常用繁体字, 而研究样本均为大陆文本, 将相关内容替换为国家《通用规范汉字表》中的一级汉字表。最终计算得到的 R 值越大表示越不易理解, 即可读性越低。

研究样本来自于图书情报学、新闻传播学及管理学3个学科, 因此将学科分类这一变量设置为研究模型的调节变量。虽然3个学科均从属于社会科学, 但不同学科在知识传播途径、学者引用偏好、学科规模等方面存在诸多差异, 因此不同学科被引频次在数据级上可能存在差异, 难以直接用于比较^[33-34]; 此外, 不同学科之间的写作方式和话语风格也存在一定差别, 这也可能会影响文本内容的可读性。因此, 有必要考察不同学科之间文本可读性与被引频次之间的差异。

3 数据分析与发现

3.1 学术文本可读性的变化

2016—2020年论文摘要及全文的 R 值变化如图1所示。总体而言, 全文 R 值比摘要高(即全文比摘要更难阅读); 同时, 两者均表现出越来越难读的迹象, 尤其是全文阅读难度在2018—2020年明显提升。

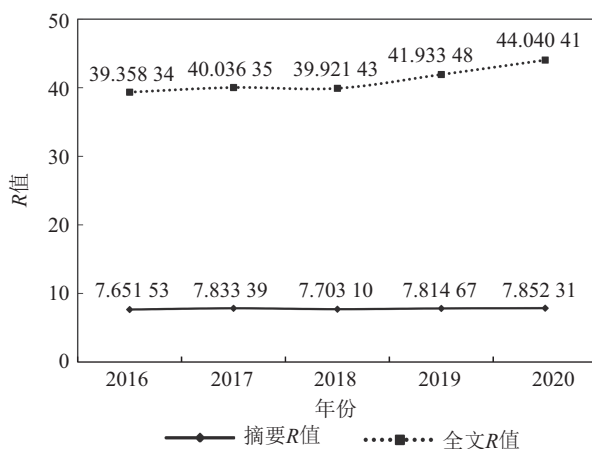


图1 2016—2020年样本论文摘要及全文 R 值变化趋势

不同学科论文摘要 R 值的变化趋势如图2所示。虽然图1显示2016—2020年论文摘要 R 值的变化趋势并不明显, 但3个学科论文摘要 R 值的变化趋势与均值变化趋势并不一致。新闻传播学论文摘要的 R 值主要呈现下降趋势, 表明其摘要随着时间的推移变得越来越容易阅读; 管理学论文摘要的 R 值呈现明显上升趋势; 图书情报学论文摘要的 R 值变化趋势比较平缓, 但是波动较大, 而且最高摘要 R 值均出现在图书情报学。整体上看, 图书情报学论文的摘要一直最难读, 新闻传播学论文的摘要一直最易读。

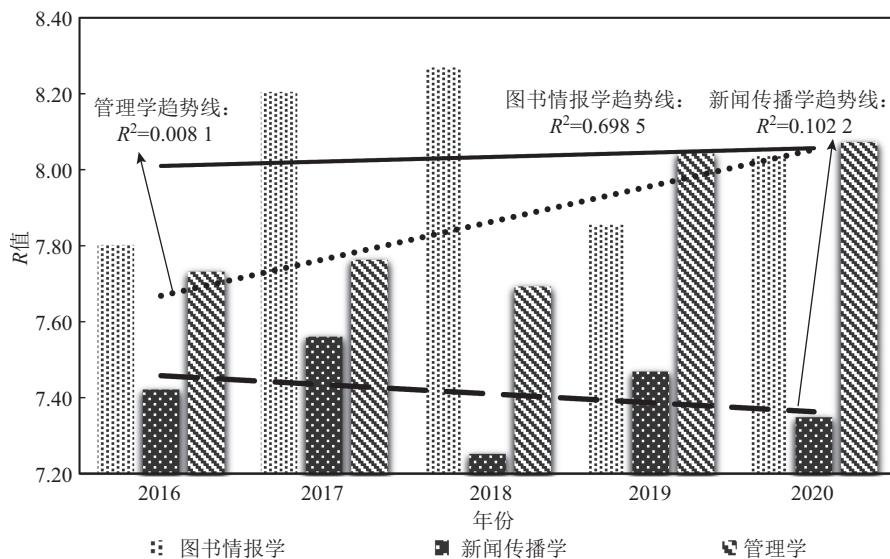


图2 2016—2020年3个学科样本论文摘要 R 值对比及变化趋势

不同学科论文全文 R 值的变化趋势如图3所示。3个学科论文全文的 R 值均有所提升: 图书情报学的提升速度最快, 管理学次之, 新闻传播学较为平缓。整体上

看, 管理学论文全文 R 值最高, 图书情报学论文全文 R 值最低, 也就是说管理学论文的全文最难读、图书情报学论文的全文最易读。

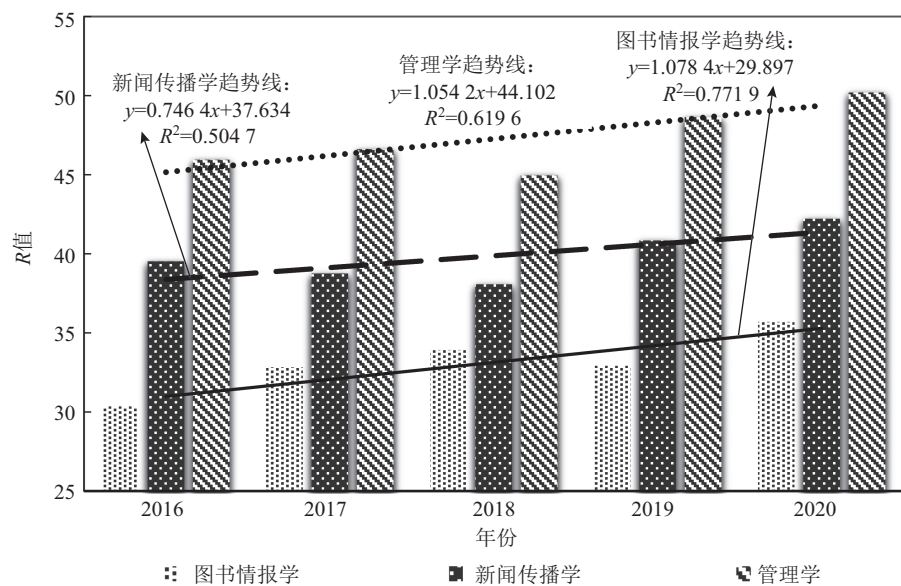


图3 2016—2020年3个学科样本论文全文R值对比及变化趋势

3.2 文本可读性与被引频次

根据文献综述分析,被引频次的影响因素复杂,且测量方式存在差异。为了更科学地展示研究变量的测量方式,根据变量的类型对其进行了描述性统计。

表1所示为分类变量的描述性统计信息。在控制变量中,非权威期刊的论文占比略多,这符合权威期刊版面紧俏、发文量有限的现实;第一作者的工作单位以高校为主,其中985高校的第一作者数量最多,占比约为52%,表明由于样本期刊均是各学科内的一流期刊,论文发表中存在“头部机构固化”现象^[35]。在调节变量中,管理学论文数量最多,约占40%,而图书情报学论文数量最少,约占28%,表明管理学的发文量更大。

表2所示为连续变量的描述性统计信息。因变量被引频次的极差较大,呈明显的离散分布现象。控制变量中,论文年龄呈轻微的正偏态分布,意味着每年发文量略微下降。预测变量中,摘要R值与全文R值都呈现出正偏态分布,说明其取值主要为较小的数值。

表1 分类变量信息

类 别	变 量	属 性	个案数	百分比/%
控制变量	权威期刊	否 (0)	1 706	53.1
		是 (1)	1 508	46.9
	单位级别	其他 (1)	333	10.4
		非985、211高校 (2)	594	18.5
		211 (3)	625	19.4
		985 (4)	1 662	51.7
调节变量	学科	图书情报学 (1)	889	27.7
		新闻传播学 (2)	994	30.9
		管理学 (3)	1 331	41.4

注:属性列括号内数字为后续模型分析中相应属性的代码。

表3所示为论文被引频次影响因素的负二项回归模型。模型一纳入了论文年龄、权威期刊以及单位级别3个变量,对可能影响被引频次的核心性社会因素进行控制;模型二增加了摘要R值以及全文R值两个预测变

表2 连续变量信息

类 别	变 量	个案数	最小值	最大值	平均值	标准偏差	偏 度	峰 度
因变量	被引频次	3 214	0	2 136.000 00	58.710 00	102.153 00	6.607	81.904
控制变量	论文年龄	3 214	3.000 00	7.000 00	4.970 00	1.414 00	0.036	-1.294
预测变量	摘要R值	3 214	3.014 66	29.740 19	7.771 42	2.649 48	2.108	7.641
	全文R值	3 214	10.523 79	141.114 24	41.091 46	12.435 54	1.100	4.300

表3 被引频次影响因素的负二项回归模型

参 数	模型一				模型二				模型三				模型四			
	β	瓦尔德卡方	P	$\exp \beta$	β	瓦尔德卡方	P	$\exp \beta$	β	瓦尔德卡方	P	$\exp \beta$	β	瓦尔德卡方	P	$\exp \beta$
截距	3.558	2 528.569	0	35.077	2.169	325.476	0	8.752	3.481	770.422	0	32.495	3.751	563.019	0	42.568
论文年龄	0.168	168.438	0	1.183	0.199	230.223	0	1.221	0.167	164.389	0	1.182	0.165	159.371	0	1.179
权威期刊(0)	-0.698	363.857	0	0.498	-0.465	146.191	0	0.628	-0.275	49.674	0	0.760	-0.284	51.260	0	0.753
权威期刊(1)				1.000				1.000				1.000				1.000
单位级别(1)	-0.388	38.998	0	0.678	-0.206	10.562	0	0.814	-0.156	6.082	0.014	0.856	-0.159	6.336	0.012	0.853
单位级别(2)	-0.027	0.309	0.578	0.973	-0.048	0.988	0.320	0.953	-0.203	17.247	0	0.816	-0.201	16.902	0	0.818
单位级别(3)	0.067	1.978	0.160	1.069	0.050	1.091	0.296	1.051	-0.084	3.074	0.080	0.919	-0.082	2.895	0.089	0.921
单位级别(4)				1.000				1.000				1.000				1.000
摘要R值					-0.019	7.098	0.008	0.981	-0.007	1.088	0.297	0.993	-0.022	3.383	0.066	0.978
全文R值					0.029	282.985	0	1.029	0.011	37.365	0	1.011	0.008	11.005	0.001	1.008
学科(1)									-1.204	559.726	0	0.300	-1.721	65.983	0	0.179
学科(2)									-1.097	589.929	0	0.334	-1.494	56.605	0	0.224
学科(3)												1.000				1.000
图书情报学摘要R值													0.036	4.693	0.030	1.037
新闻传播学摘要R值													0.007	0.165	0.684	1.007
管理学摘要R值																1.000
图书情报学全文R值													0.006	1.389	0.239	1.006
新闻传播学全文R值													0.008	4.097	0.043	1.008
管理学全文R值																1.000

注：显示 P 为0时，实际为 <0.001 。

量，用于考察文本可读性对被引频次的影响；模型三纳入了调节变量学科；模型四检验了学科对摘要R值、全文R值与被引频次关系的调节作用。Omnibus检验显示4个模型均显著有效，同时随着模型不断纳入新的变量，模型的皮尔逊卡方值也在不断减小，说明模型整体的解释力在不断上升并收敛。

模型一显示论文年龄、权威期刊以及单位级别均显著影响论文的被引频次。具体而言，发表时长越长的论文被引频次越高($\beta=0.168$, $P<0.001$)，论文年龄每增加一年论文的被引频次增加一个单位的优势比的概率会增加18.4%。相对于发表在权威期刊上论文，非权威期刊论文的被引频次增加一个单位的优势比的概率低49.2%。相对于在985高校任职的第一作者，在其他单

位任职的第一作者的论文被引频次增加一个单位的优势比的概率低21.4%。

模型二显示在控制论文年龄、权威期刊及单位级别3个变量的情况下，两个预测变量均显著影响论文的被引频次。具体而言，摘要R值越高论文的被引频次越低，摘要R值每增长一个单位，论文被引频次增长一个单位的优势比的概率就会减少1.9%，即摘要越容易阅读，论文被引用的可能性越高。而全文R值越高论文的被引频次越高，全文R值每增长一个单位，论文被引频次增长一个单位的优势比的概率会增加2.7%，即越难读的论文被引频次反而越高。

模型三显示在控制了论文年龄、权威期刊及单位级别和摘要R值、全文R值的情况下，学科对被引频次

的影响也十分显著。具体表现为,相对于管理学论文而言,图书情报学和新闻传播学论文的被引频次增加一个单位的优势比的概率分别低70.0%、66.5%。

模型四在模型三的基础上考察了学科对摘要 R 值、全文 R 值和被引频次之间关系的调节作用。结果表明,学科对摘要 R 值、全文 R 值与被引频次的关系都有一定的调节作用。具体而言,在摘要 R 值对被引频次的影响过程中,图书情报学与管理学之间存在显著的学科差异($\beta=0.036, P=0.030$),即图书情报学的摘要 R 值增加一个单位引发被引频次也增加的概率比管理学高3.7%。在全文 R 值对被引频次的影响过程中,新闻传播学与管理学之间存在显著的学科差异($\beta=0.008, P=0.043$),即新闻传播学的全文 R 值增加一个单位引发被引频次也增加的概率比管理学高0.8%。

值得注意的是,在纳入调节变量学科前,摘要 R 值对被引频次有显著影响(如模型二所示);当模型纳入学科变量后,摘要 R 值对被引频次的影响不再显著($\beta=-0.007, P=0.297$),这说明摘要 R 值对被引频次的影响被学科差异对被引频次的影响覆盖。在4个模型中,3个控制变量均显著影响着论文的被引频次,这一特征与已有研究结果相符,说明权威效应和时间效应对被引频次具有稳定的影响效果^[36-37]。

4 文本可读性对被引频次的影响

如前所述,虽然国外已有学者关注并研究论文可读性与被引频次的关系,但以汉语学术文本为对象的可读性研究还很缺乏;已有研究关于可读性与被引频次之间关系的实证结果并不一致,导致两者关系的研究结论仍存在争议。本研究的价值就在于以汉语文本为样本,揭示了学术文本可读性与被引频次之间的关系;同时,将学科作为调节变量,以探索不同学科之间文本可读性的变化规律以及文本可读性对被引频次的影响。

4.1 摘要可读性对被引频次的影响

根据前文分析可知,2016—2020年样本论文摘要的可读性呈轻微的下降趋势。加入学科变量后对这一趋势展开了进一步分析,发现不同学科的摘要可读性变化趋势也存在较大差异,主要表现为:图书情报学及管理学的摘要逐渐变难读,而新闻传播学摘要却逐渐易读;相对于其他两个学科而言,管理学摘要可读性每年

的变化幅度更为明显。这一发现与Gazni^[38]的研究结果相似,即有些学科的摘要可读性明显不如其他学科,而且摘要可读性在被调查的时间段里一般不会出现太大的变化。摘要是一篇论文的内容浓缩,是读者阅读全文的重要前提。不同学科话语模式的差异引发了摘要可读性的学科差异,在短期内同一学科话语模式相对稳定,则摘要可读性具有稳定性。管理学是一个学科范式比较混杂的学科,存在“管理科学与工程”和“工商管理”两大基本的话语模式。《管理世界》在2019年开始增设“管理科学与工程”专栏,这可能是管理学的摘要可读性与全文可读性在2019年、2020年发生突变的重要原因。

由模型二可知,在控制论文年龄、权威期刊及单位级别的条件下,摘要越容易阅读的论文被引频次越高。但在加入学科变量后,摘要 R 值的主效应不再显著,图书情报学及管理学的摘要可读性及被引频次之间的关系存在显著的学科差异。图4展示了学科对摘要 R 值与被引频次关系调节作用的简单斜率检验结果^[39]。由图4可知,随着摘要 R 值的提升,论文的被引频次逐渐降低,即图书情报学和管理学中,摘要越容易阅读的论文被引用的可能性越高。相对于管理学来说,图书情报学的斜率更大,即在图书情报学中摘要越易读,提升被引频次的效果越明显,因此相对管理学来说,图书情报学中的调节作用表现为增强作用。

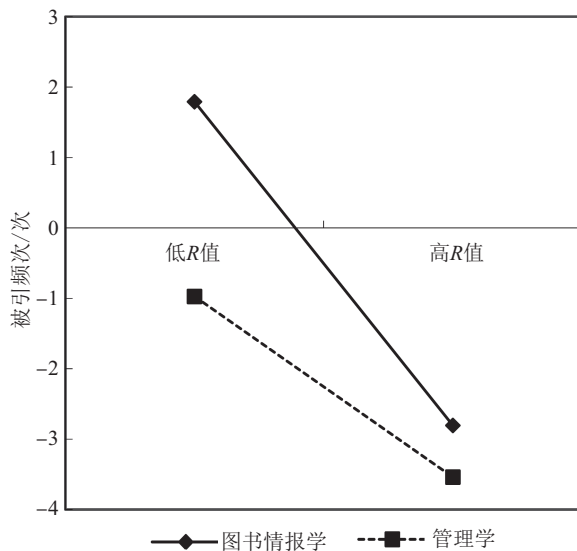


图4 学科对摘要 R 值与被引频次关系的调节作用

4.2 全文可读性对被引频次的影响

如前所述,与摘要相比,样本论文的全文 R 值的增

长趋势较为明显。加入学科变量对这一趋势进一步分析发现, 3个学科全文 R 值均表现为逐年增长趋势。这与国内社会科学研究日益规范化、定量研究方法日益普及有很大关系。

由模型二及模型四可知, 与摘要 R 值不同, 加入学科变量后全文 R 值对被引频次的主效应一直显著, 表明全文可读性对被引频次的影响效应比较稳定, 越难读的论文被引频次越高。新闻传播学与管理学论文的全文可读性及被引频次之间的关系存在显著的学科差异。图5展示了学科对全文 R 值与被引频次关系调节作用的简单斜率检验结果。如图5所示, 随着全文 R 值的提升, 论文的被引频次逐渐升高, 即新闻传播学和管理学论文都表现为全文越难阅读, 被引频次越高的特点; 而相对于管理学来说, 新闻传播学的斜率更小, 即在新闻传播学中全文低可读性对提升被引频次的效果没有管理学显著, 因此相对管理学来说, 新闻传播学中的调节作用表现为抑制作用。

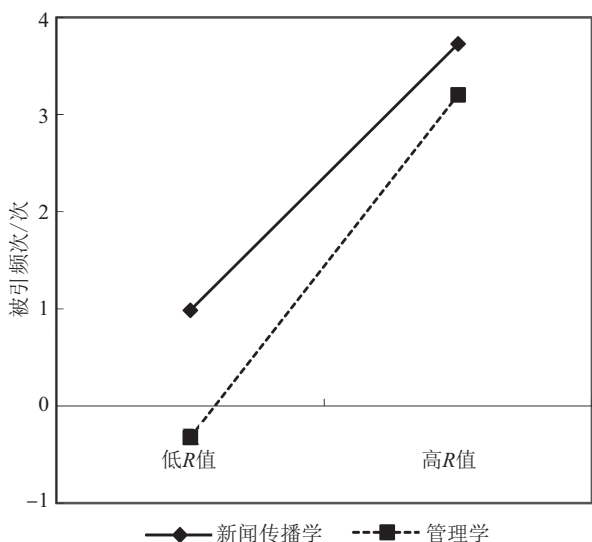


图5 学科对全文 R 值与被引频次关系的调节作用

5 结语

可读性是论文质量的一个重要特征, 本研究通过量化文本内容可读性的方式探索学术论文的文本可读性对论文影响力的影响, 对作者进行论文写作以及科研评价具有重要参考价值。研究使用荆溪昱等的可读性公式分别计算了样本论文摘要及全文的可读性, 并运用负二项回归模型检验了论文摘要及全文可读性与被引频次之间的关系。研究发现, 在控制论文年龄、权威期刊及单位级别的情况下, 摘要易读而全文难读的论文

被引频次较高。可读性与被引频次之间的关系受到学科的调节, 即相对于管理学来说, 若图书情报学论文摘要难读, 则论文被引频次更容易减少, 新闻传播学论文全文可读性对被引频次的影响不如管理学显著。

作为传播论文知识的两个核心部分, 摘要和全文扮演了两种不同的角色。摘要作为读者阅读论文的前提, 其易读性积极地影响着论文被引, 这意味着作者应该尽可能用简洁、直接的语言描述研究的核心内容, 以达到快速吸引读者注意的目的; 而全文为读者提供更详尽的知识信息, 影响着读者对知识结构严谨度、研究成果有用性等的感知和评价, 作者在撰写全文时应将研究步骤和方法尽可能详细地展示在读者眼前, 以满足读者的阅读需求, 获得更多的认可与支持。本研究在以下两个方面还存在一定缺陷。首先, 对文本可读性的量化依赖于可读性计算公式, 然而计算公式利用的可读性指标主要取决于文字内容的特征, 实际上文本可读性还受到记录载体^[40]、个人教育水平^[41]等因素的影响。可读性计算公式提供的可读性指标从客观层面揭示文本自身存在的难易度差异, 而忽略个体因素带来的影响。其次, 仅选择了社会科学中的图书情报学、新闻传播学以及管理学论文作为研究样本, 对于自然科学文本可读性的的发展及其对论文传播的影响需进一步的对比研究。

参考文献

- [1] 王楠, 罗珺文. 高校科研成果的非学术影响及其评估: 是什么, 为什么, 怎样做?[J]. 华东师范大学学报(教育科学版), 2020, 38(4): 62-71.
- [2] 荆溪昱, 赵世范, 翁凌志. 中文文章适读性线上分析系统之发展研究[C]//科技教育课程改革与发展学术研讨会论文集, 2007: 47-57.
- [3] EDGAR D, JEANNE S C. The concept of readability[J]. Elementary English, 1949, 26(1): 19-26.
- [4] 黄敏. 汉语特质与中文新闻易读性公式研究[J]. 新闻与传播研究, 2010, 18(4): 93-97, 112.
- [5] 王蕾. 文本可读性公式研究发展阶段及特点[J]. 语言教学与研究, 2022(2): 29-40.
- [6] VOGEL M, WASHBURN C. An objective method of determining grade placement of children's reading material[J]. The Elementary School Journal, 1928, 28(5): 373-381.
- [7] 吴思远, 蔡建永, 于东, 等. 文本可读性的自动分析研究综述[J]. 中文信息学报, 2018, 32(12): 1-10.

- [8] CROSSLEY S A, SKALICKY S, DASCALU M. Moving beyond classic readability formulas: new methods and new models[J]. *Journal of Research in Reading*, 2019, 42 (3/4) : 541-561.
- [9] GRAESSER A C, MCNAMARA D S, KULIKOWICH J M. Coh-matrix: providing multilevel analyses of text characteristics[J]. *Educational Researcher*, 2011, 40 (5) : 223-234.
- [10] YANG S. A readability formula for Chinese language[D]. Wisconsin: University of Wisconsin, 1970.
- [11] 秦琴, 柯青, 丁松云. 中文在线健康教育信息可读性计算及应用实证: 以食品安全领域为例[J]. *现代情报*, 2020, 40 (5) : 111-121.
- [12] 郭望皓. 对外汉语文本易读性公式研究[D]. 上海: 上海交通大学, 2010.
- [13] 李萍. 我国上市公司年报中“管理层讨论与分析”的可读性分析[J]. *时代金融*, 2018 (9) : 225-226.
- [14] 蔡建永. 汉语二语文本可读性公式研究[D]. 北京: 北京语言大学, 2020.
- [15] BAILIN A L, GRAFSTEIN A. The linguistic assumptions underlying readability formulae: a critique[J]. *Language & Communication*, 2001, 21 (3) : 285-301.
- [16] 逯东, 宋昕倍. 媒体报道、上市公司年报可读性与融资约束[J]. *管理科学学报*, 2021, 24 (12) : 45-61.
- [17] 柯青, 丁松云, 秦琴. 健康信息可读性对用户认知负荷和信息加工绩效影响眼动实验研究[J]. *数据分析与知识发现*, 2021, 5 (2) : 70-82.
- [18] HARTLEY J, SOTTO E, PENNEBAKER J. Style and substance in psychology: are influential articles more readable than less influential ones?[J]. *Social Studies of Science*, 2002, 32 (2) : 321-334.
- [19] SAWYER A G, LARAN J, XU J. The readability of marketing journals: are award-winning articles better written?[J]. *Journal of Marketing*, 2008, 72 (1) : 108-117.
- [20] VAN WESELE M, WYATT S, TEN HAAFF J. What a difference a colon makes: how superficial factors influence subsequent citation[J]. *Scientometrics*, 2014, 98 (3) : 1601-1615.
- [21] LEI L, YAN S. Readability and citations in information science: evidence from abstracts and articles of four journals (2003-2012) [J]. *Scientometrics*, 2016, 108 (3) : 1155-1169.
- [22] 陈练文, 李信, 赵超烨. 学术文本可读性和复杂度评价研究[J]. *数字图书馆论坛*, 2018 (5) : 62-66.
- [23] JIN T, DUAN H Q, LU X F, et al. Do research articles with more readable abstracts receive higher online attention? evidence from Science[J]. *Scientometrics*, 2021, 126 (10) : 8471-8490.
- [24] 陈洁. 国内可读性研究概述[J]. *黑龙江史志*, 2013 (9) : 212-213.
- [25] 李樵. 外部引用视角下的中国图书馆情报学知识影响力研究[J]. *中国图书馆学报*, 2019, 45 (6) : 65-83.
- [26] GARFIELD E. The history and meaning of the journal impact factor[J]. *The Journal of the American Medical Association*, 2006, 295 (1) : 90-93.
- [27] SEGLEN P O. The skewness of science[J]. *Journal of the American Society for Information Science*, 1992, 43 (9) : 628-638.
- [28] 刘玉仙, 武夷山. 关于引文本质的思考[J]. *科学学研究*, 2015, 33 (12) : 1779-1786, 1882.
- [29] WANG M Y, YU G, YU D R. Mining typical features for highly cited papers[J]. *Scientometrics*, 2011, 87 (3) : 695-706.
- [30] JABBOUR C J C, DE SOUSA JABBOUR A B L, DE OLIVEIRA J H C. The perception of Brazilian researchers concerning the factors that influence the citation of their articles: a study in the field of sustainability[J]. *Serials Review*, 2013, 39 (2) : 93-96.
- [31] GUALBERTO B C, IZABELA Z. Analysis of the relationship between the number of citations and the quality evaluated by experts in psychology journals[J]. *Psicothema*, 2010, 22 (2) : 270-276.
- [32] STREMER S, VERNIERS I, VERHOEF P C. The quest for citations: drivers of article impact[J]. *Journal of Marketing*, 2006, 71 (3) : 171-193.
- [33] RADICCHI F, CASTELLANO C. Testing the fairness of citation indicators for comparison across scientific domains: the case of fractional citation counts[J]. *Journal of Informetrics*, 2012, 6 (1) : 121-130.
- [34] 张垒. 高被引论文的特征因素及其对影响因子贡献研究[J]. *中国科技期刊研究*, 2015, 26 (8) : 880-885.
- [35] 甘琳, 李刚. CSSCI来源期刊发文机构固化现象的成因与影响: 以图情档期刊为例[J]. *图书馆论坛*, 2021, 41 (2) : 56-66.
- [36] PENG T Q, ZHU J J H. Where you publish matters most: a multilevel analysis of factors affecting citations of internet studies[J]. *Journal of the American Society for Information Science and Technology*, 2012, 63 (9) : 1789-1803.
- [37] 刘宇, 张永娟, 齐林峰, 等. 知识启迪与权威尊崇: 基于重复发表的引文动机研究[J]. *图书馆论坛*, 2018, 38 (4) : 49-57.

- [38] GAZNI A. Are the abstracts of high impact articles more readable? investigating the evidence from top research institutions in the world[J]. Journal of Information Science, 2011, 37 (3): 273-281.
- [39] Interpreting interaction effects[EB/OL]. [2023-03-28]. <http://www.jeremydawson.co.uk/slopes.htm>.
- [40] ŠTAJNER S, EVANS R, ORASAN C, et al. What can readability measures really tell us about text complexity[C]//Proceedings of Workshop on Natural Language Processing for Improving Textual Accessibility, 2012: 14-22.
- [41] REDISH J. Readability formulas have even more limitations than Klare discusses[J]. ACM Journal of Computer Documentation, 2000, 24 (3): 132-137.

作者简介

伍丹炜, 女, 硕士研究生, 研究方向: 学术传播与评价研究。

刘宇, 男, 博士, 教授, 通信作者, 研究方向: 学术传播与评价研究, E-mail: histly@hotmail.com。

Readability and Influence of Academic Papers: From the Perspective of Disciplinary Differences in Humanities and Social Sciences

WU DanWei¹ LIU Yu²

(1. College of History and Archives, Yunnan University, Kunming 650091, P. R. China;

2. School of Journalism, Yunnan University, Kunming 650504, P. R. China)

Abstract: Readability is one of the important characteristics of paper quality, which affects readers' cognition and understanding of research contents. Exploring the relationship between the readability of a paper and its influence is of great practical significance to the effective dissemination of academic knowledge. This study examines the impact of abstract readability and full-text readability on citation frequency based on papers published in representative journals of library and information science, journalism and communication, and management from 2016 to 2020, and tests the difference between the disciplines. It is found that papers with easy-to-read abstracts and hard-to-read full-text are more likely to be highly cited, and as the moderating variable, discipline has significant impact on the relationship between abstract readability and citation frequency and the relationship between full-text readability and citation frequency. Writers should strive to enhance the readability of the paper while ensuring rigorous and precise expression, so as to promote the exchange and development of subject knowledge.

Keywords: Text Readability; Academic Communication; Citation Frequency; Disciplinary Difference

(责任编辑: 王玮)