

“信息通信技术”在新媒体科学传播中的话语建构^{*}

王 晶

(中山大学, 广州 510275; 重庆邮电大学, 重庆 400065)

提 要: 近年来, 我国的信息通信技术(ICT) 产业高速发展, 如何应用新媒体中的科学传播话语建构“信息通信技术”这一概念成为本研究的课题。本文通过数据挖掘建立新媒体 ICT 科学传播语料库, 应用 T-LAB 软件展开类符/形符比、词频、语义韵和主题词分析, 考察“信息通信技术”在语料库中如何表现, 并对语料库做出整体的聚类分析, 以此反映当下 ICT 科学传播话语所涵盖的主要内容。

关键词: 信息通信技术; 新媒体; 科学传播; 话语; 语料库

中图分类号: H030

文献标识码: A

文章编号: 1000 - 0100(2019) 03 - 0045 - 6

DOI 编码: 10. 16263/j. cnki. 23 - 1071/h. 2019. 03. 008

The Discursive Construction of ICT in New Media Science Communication

Wang Jing

(Sun Yat-sen University, Guangzhou 510275, China; Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Since ICT industry is making fast progress in China, this paper focuses on how to construct ICT with science communication discourse. We build a corpus by data-mining articles popularizing ICT on the web and use T-LAB software to facilitate the discourse studies. On the one hand, we investigate how ICT is revealed through type-token ratio, frequency, semantic prosody and key word analysis; on the other hand, a clustering analysis is carried out to demonstrate the major categories of ICT science communication.

Key words: ICT; new media; science communication; discourse; corpus

由于生产、复制和传播的便捷性以及来源的多元性, 新媒体中的科学传播文本与传统媒介相比数量极其庞大。如此大数据的科学传播文本, 对于信息通信技术(ICT) 知识的传播和普及无疑很有益处。“单个文本是微不足道的, 传媒的影响力是通过累积而来, 是通过对某一事件的因果关系、事件主体反复刻意的描述以及对读者的反复教导等手段来完成的。”(Fairclough 1989: 54) 对于这样的研究对象, 大数据的语料库是开展话语分析的最佳辅助工具。本研究基于大数据挖掘和语料库, 运用大量真实的文本数据厘清 ICT 科学传播的话语现状, 并探究新媒体中的科学传播话语如何建构“信息通

信技术”这一科学概念。

1 新媒体话语对科学概念的建构作用

在信息社会, 新媒体在建构 ICT 的科学概念、培育公众的 ICT 科学素养方面起着举足轻重的作用。第九次中国公民科学素质调查报告^①显示, 公民利用互联网及移动互联网获取科技信息的比例达到 53.4%, 比 2010 年的 26.6% 提高一倍多, 已经超过报纸(38.5%) , 仅次于电视(93.4%) , 位居第二。互联网已成为具备科学素质公民获取科技信息的第一渠道。

Gerbner 指出, 传媒对读者的影响是长久深远

^{*} 本文系教育部人文社科项目“信息通信技术在虚拟公共领域中的科学传播话语体系研究”(17YJC740082) 的阶段性成果。

作者电子邮箱: 20324313@qq.com

的,这种影响起初是细微的,但随着时间的推移和文字或图像的反复,其影响力将呈几何式的增长 (Gerbner 1986)。传媒与读者的关系复杂且相互影响。Fairclough(1989),Conboy(2010)等都曾论述过传媒、话语与读者的关系,并且一致认为,传媒会根据其代表的价值观和权力关系用话语对现实进行重构,也会根据特定的读者群来选择特定的语言变体,从而帮助读者实现身份认同,并潜移默化地建构读者的思想。但读者并不总是被动地被构建,不总是信息的被动接收者,读者也能影响传媒的话语策略和文体方式。意义是在文本和读者的互动中创造出来的 (McIlvenny 1996) 或者说,传媒、信息与读者通过话语实现相互的建构。在新媒体的特定语境下,话语和读者在科学传播的过程中相互影响和掣肘,从而完成对科学概念的建构。

新媒体所传播的科学内容涵盖各门各类,非常庞杂,我们选择 ICT 领域作为研究对象的理由有二。首先,信息通信技术是第三次和第四次工业革命的技术基础。其次,根据 2016 移动互联网网民科普获取及传播行为研究报告^②,从 2016 年度网民对科普内容的关注度分布上看,信息科技成为移动端网民最关注的科普主题,用户关注份额为 24.8%。这也是笔者将 ICT 领域的科学传播文本作为研究对象的一个重要原因。

2 “信息通信技术”的概念考察

基于网站流量和 Alexa 排名,我们从中国的科学传播综合网站中选出 12 个最有影响力且最具代表性的网站,作为收集语料库文本的平台,并使用网络爬虫软件抓取出共 7 276 个 ICT 科学传播文本,这些文本的发布时间段是 2015 年 1 月 1 日至 2016 年 12 月 31 日之间。运用 T-LAB 软件基于以上文本建立新媒体 ICT 科学传播语料库,类符 (types) 123 184 个,形符 (tokens) 5 888 541 个,类符/形符比 (type-token ratio, TTR) 为 0.021,可见该语料库中的文本词汇密度 (lexical density) 并不高。其中出现频率仅一次的形符数量为 62 178 个。值得注意的是, T-LAB 的词汇密度计算不包括停用词。

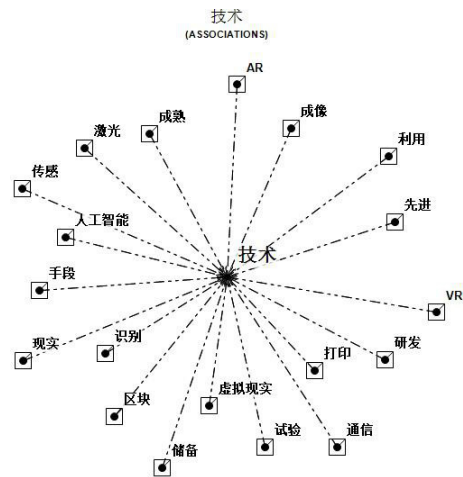
首先考察“信息通信技术”这一概念在语料库中的表现。“20 世纪 70 年代以后,语言学将词汇看成语言的核心”(陈功 梁茂成 2017: 18),我们将“信息”“通信”“技术”拆分并两两组合之后进行考察,得出各个词及词组的词频 (参见表)。“信息通信技术”一词(包括“信息和通信技术”“信息与通信技术”的形式)虽然是本语料库的主题,但在所有文本中出现的频数却很低,

只有 24 次,而其英文缩写 ICT 出现的频数反而较高,达到 270 次。这说明新媒体的科学传播更倾向于用便捷的英文缩写来代替。而且“信息通信”这种多词组合在科学传播文本中也并不受青睐,总共出现 54 次,因此,“信息通信技术”和“信息通信”可以不作为研究考察的重点。

表₁ 本语料库及参照语料库 CNC 的“信息通信技术”词频对比

词	词频数 (本语料库)	词频率 (本语料库)	词频数 (CNC)	词频率 (CNC)
信息	41251	0.3349	1637	0.1713
通信	12348	0.1002	188	0.0197
技术	72974	0.5924	5978	0.6256
信息技术	1711	-	-	-
通信技术	264	-	-	-
信息通信	54	-	-	-
信息通信技术	24	-	-	-
ICT	270	-	-	-
信息化	3123	-	-	-

在很多文本中,常常将“信息技术”与“通信技术”分开表述,而这二者在科学传播中的分布并不均衡,“信息技术”俨然成为关注的重点,出现的频数高达 1 711 次,而“通信技术”只有 264 次,这大概与“信息技术”和“通信技术”的学科侧重点有关。相较之下,信息技术与公众的生活更为接近,作为科学传播的内容更能引起读者的兴趣,这反映出当下新媒体科学传播中“读者中心”的发展倾向。尤其值得一提的是,“信息化”一词出现的频率非常高,甚至超过“信息技术”,达到 3 123 次,可见信息化是当下网络传播中的一个热点。



图₁ “技术”的搭配情况

此外,“信息”“通信”“技术”在语料库中的词频数都很高,但这 3 个词的词频率需要和参照语料

库进行对比,才能显示出如此高的词频率是本语料库特有还是普遍的现象。本研究所选择的参照语料库为国家语委现代汉语语料库,简称CNC^③。通过和CNC的对比可以看出,“信息”“通信”两词的词频率明显高于CNC,说明本语料库与信息通信领域存在紧密的关系。但“技术”的词频率略低于CNC。“技术”是比较宽泛的概念,可以和很多领域的词搭配使用,并不限于信息通信领域,因此其频率的高低不能反映本语料库的核心内容,这一点可从“技术”在本语料库中的搭配看出。显然,不能孤立地看语料库中的词,还须考察该词的搭配。早在20世纪50年代,英国语言学家Firth就提出“搭配”的概念,他认为搭配是词语之间的结伴关系。在语料库索引行中,以节点词为中心,左右的词数之和为跨距(span),跨距内每个位置上出现的词即节点词的搭配词。根据Baker等人的研究(2013:36),本研究将跨距设置为5。图1清晰地显示出主题词“技术”在本语料库中的主要搭配:在排名前19位的各个搭配词中,大部分(13个)都指向信息通信专业范畴的技术内容的搭配,如虚拟现实、人工智能、AR、成像等,两个形容词搭配(先进、成熟)具有明显的积极语义倾向,这可以显示当下在科学传播中对信息通信技术的态度是正面、褒扬的。剩下的4个搭配(利用、研发、试验、手段)都不具备

强烈的语义倾向,较为中性。可见,虽然“技术”一词在本语料库中的词频率略低于CNC,但其语义指向集中,且非常明确,与语料库的主题密切相关。

除分析“技术”一词的搭配情况外,进一步考查“信息”“通信”两个词的搭配情况能提供科学传播文本中对于信息通信技术这一主题更多的信息。科学传播中对于“信息”的表述更偏向与公众日常生活相关的内容,如“身份证”“办公室”等。即使“信息技术”一词在语料库中的频数达到1,711次,也未能排进“信息”一词的搭配热词前20位。20个搭配词中,动词占据半壁江山(提取、传递、交换、泄露、窃取、查询、发送、收集、获取、传输),分析这些动词的索引行可以看出,其中8个是中性词汇,另2个具有强烈的负面语义韵(semantic prosody)(泄露、窃取)。“泄露”“窃取”两个词频繁地与“风险”“安全”“担心”“摧毁”等词共现,体现出科学传播文本对于个人信息、数据信息安全事件的充分关注。与之相应,“安全”一词位居“信息”搭配热词的第19位,可以看出,在对“信息”相关的科学传播内容中,科学原理并不占据重要的地位,反而是与民生相关的信息安全、信息服务等内容成为传播的主要对象,这与“通信”一词的情况正好相反。

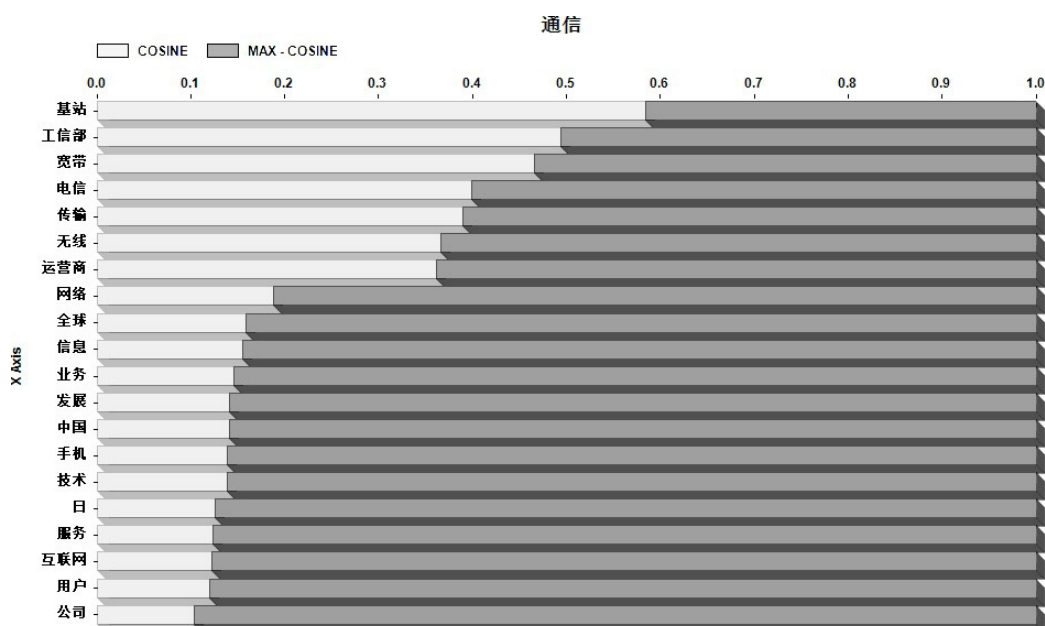


图2. “通信”的搭配词

在“通信”的搭配热词中(参见图),“信息”与“技术”都位于前20位,分别是第10位和第15位,这说明,当“通信”一词出现的时候,两个词的

共现频率也比较高。这20个搭配热词的分布充分体现出科学传播中关于“通信”的内容大致分布在3个话题中,首先是基础通信设施和运营商,

这包括“基站”“宽带”“电信”“运营商”“网络”“业务”“服务”“互联网”“公司”“手机”和“用户”共 11 个词,其次是与通信技术有关的内容,包括“传输”“无线”“信息”“技术”,最后是和通信的大环境相关的内容,如“工信部”“中国”“发展”等。这与“信息”主要搭配公众民生的选择大相径庭,“通信”一词的共现热词更多的是从宏观或者技术的角度来展开。以上的分析呈现出本语料库中单个文本分析无法显示的词语搭配倾向,虽然人们习惯将“信息通信技术”视为一个整体,但在实际的新媒体科学传播中,传播者对这一概念所包含的 3 个词俨然区别对待,这或许是文本生产者自己都没有意识到的。“词语像人类一样聚群”(梁茂成等 2010: 86)在语言运用的横向序列中,一些词总是与其他词共同使用,具有很高的共现频率;而在纵向聚合关系中,一些特定的词语群也总是会被触发,表达相同或相似的主题。通过 T-LAB 对以上 3 个词的索引及搭配热词的分析,揭示出隐藏在海量文本下的语用习惯,为我们对“信息通信技术”科学传播话语的认识开启新的视角。

3 ICT 科学传播语料库的整体聚类分析

虽然对“信息通信技术”这一概念在本语料库中的话语状况已有基本的认识,但本语料库的文本爬取工作是以笔者选取的 64 个 ICT 关键词为标准展开,这 64 个 ICT 关键词涵盖 ICT 领域的各个方面。要对本语料库的内容有一个整体的把握,仅仅考察“信息通信技术”概念显然不够,还需要对文本进行整体的聚类主题分析,以便研究者对庞杂的 ICT 科学传播所呈现的主题有了解。T-LAB 软件的基础文本主题分析功能(thematic analysis of elementary contexts)最终的聚类结果显示,ICT 科学传播文本共聚焦于 9 个主题,各主题的信息参见表。

从各主题的文本数量来看,排在前三位的主题分别是网络安全(15.12%)、市场(14.97%)、人工智能(14.57%)。这 3 个主题占据整个语料库近一半的文本,可见 ICT 的科学传播中侧重于以上 3 个方面的普及。根据《2016 年泰尔 ICT 深度观察》,“自 2012 年以来,全球网络与信息安全事件数量连年增加。截止到 2015 年 10 月,安全事件数量达到近 6000 万起,较 2014 年增加 38%,超过 2012 年安全事件数量的 2 倍”(中国信息通信研究院 2016: 175)。目前,网络安全形势不容乐观,攻防对抗不断升级,增加网络安全教育成为迫在眉睫的事情,因此这一主题成为 ICT 科学传播的首要关注也在情理之中。排在后

3 位的主题分别是数字出版(7.23%)、手机电脑及操作系统(7.85%)、运营商与监管(7.92%),这是 ICT 科学传播话语中相对较少关注的领域。

表₂ 语料库聚类主题信息

主题编号	主题名称	文本个数	百分比	前 5 个关键词
CLUSTERS 1	量子通信	1297	10.46%	量子 计算机 光子 传感器 性能
CLUSTERS 2	数字出版	896	7.23%	出版 互联网 产业 创新发展
CLUSTERS 3	互联网社交	1004	8.1%	红包 直播 大家 微博 粉丝
CLUSTERS 4	手机电脑及操作系统	973	7.85%	微软 苹果 谷歌 安卓 windows
CLUSTERS 5	人工智能	1806	14.57%	人工智能 人类 机器人 围棋 学习
CLUSTERS 6	运营商与监管	982	7.92%	诈骗 运营商 电信 电话 宽带
CLUSTERS 7	网络安全	1875	15.12%	黑客 FBI 密码 攻击 加密
CLUSTERS 8	市场	1856	14.97%	美元 季度 营收 增长 下滑
CLUSTERS 9	投资	1709	13.78%	投资 融资 乐视 资本 滴漏

除了文本在各个主题内的分布,我们还可以考察各个主题中文本的来源情况。本语料库中的文本爬取自 12 个科学传播网站,图₃是文本来源与文本所属主题的分布柱状图,该柱状图并未显示各来源的文本在数量上的差别,而是显示每个科学传播网站中的文本在各个主题的分布情况。虽然图中显示人工智能与量子通信所占的比重最大,但当把各个来源的文本数量考虑进来时,各个主题在语料库中所占比重就与此图所显示的不同。可以看出,主题分布最为单一的是科学松鼠会(SONGSHU-HUI),该网站的文本全部是人工智能的主题,网站的文本数量最少,只有 7 篇,这 7 篇中有 4 篇是名为“计算的极限”的系列文章,出自同一个作者,还有 2 篇是关于阿尔法狗,1 篇关于奥创,有限的文本与作者导致其主题的单性。值得注意的是,与科学松鼠会属于同一创始人的果壳网(GUOKR 86 篇)虽然有 5 个主题,但各主题间的分布极不均衡,人工智能主题依然占据该网站绝大多数的文本。通过科学传播文本主题的对比,从侧面反映出这两个网站的同源本质,即使一个为公益网站,一个为商业网站,但在科学传播的主题偏好上,仍然具有极大的相似度。

主题分布不均的还有新浪网(SINATECH)、微科普(WKEPU)与中国科普博览(KEPU)。新浪网(204 篇)与微科普(29 篇)的绝大部分文本分布在人工智能与量子通信主题上,中国科普博览除了这两个主题之外,还有相当一部分的文本是关于网络安全。在这 12 个科学传播网站中,主题最全面、分布最均衡的是腾讯科学(TECHQQ),不仅涵盖 9 大科学传播的主题,难能可贵的是这

9 个主题的柱状格几乎是均等的。作为网络科学传播事业的领头羊,腾讯科学在传播文本的数量(4,580 篇)上遥遥领先于其他的科学传播网站;在文本的质量上,从把握科学传播主题的均衡性来看,也对其余网站有压倒性优势。对文本主题

的分布情况分析进一步证明,腾讯科学网站在 ICT 传播中起着重要的作用。另外一个主题分布较为均衡的网站是科普中国(KEPUCHINA),该网站为本语料库只提供 24 个文本,在科学传播文本的生产数量上还明显不足。

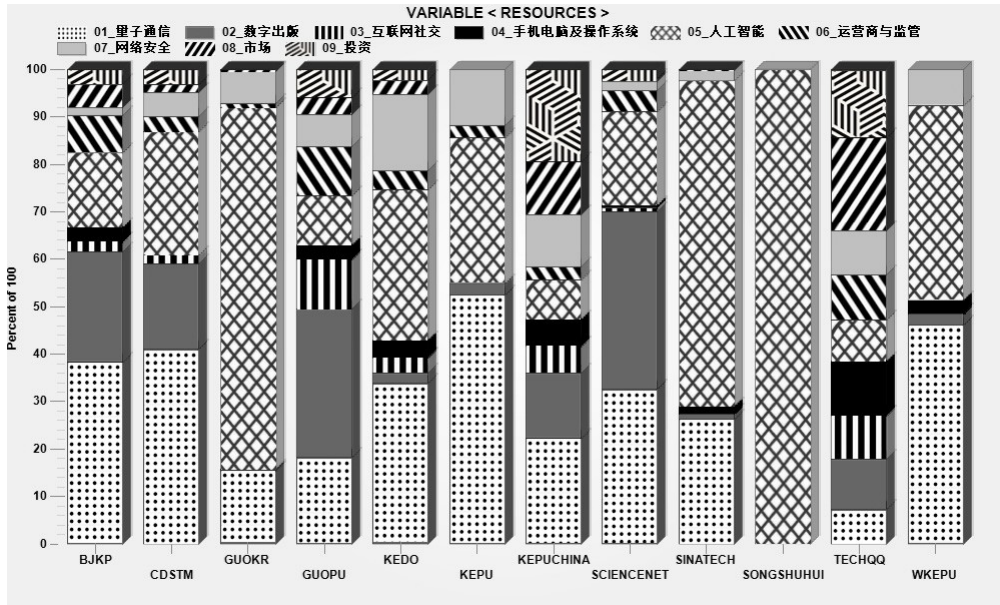


图3 各来源文本的主题分布情况

我们不仅可以通过聚类分析了解本语料库的主题及其文本分布,还可以查看各个主题之间的关系。通过关联分析(correspondence analysis),能

在二维图表(图4)中显示以上 9 个主题之间的亲疏远近关系,从而将话语中隐含的各个主题的关联用可视化的方式呈现出来。

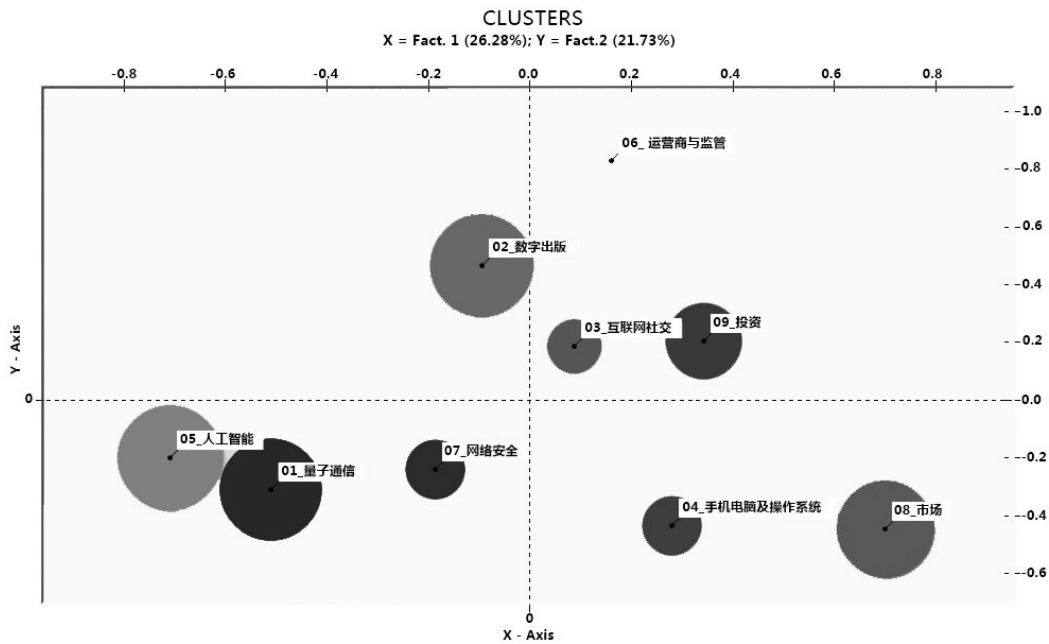


图4 9 个聚类主题的关系二维图

关联分析能发现在大量数据集中的关联性或相关性,从而描述语料库中某些属性同时出现的规律和模式。在图4中,以X轴和Y轴的0点为中心,9个聚类主题分布在X轴与Y轴的正负两极,各主题距离越近相关性就越高,反之则相关性越低。我们展开关联分析的目的是要找出具有强相关性的主题,即哪个主题出现时,另一个主题有可能共现。关联分析的结果显示,9个主题中关联系数最高的是互联网社交—投资(0.467),其次是手机电脑及操作系统—网络安全(0.414)、手机电脑及操作系统—市场(0.357)。呈弱相关的关联系数集中在运营商与监管这个主题上,分别是运营商与监管—人工智能(0.000)、运营商与监管—市场(0.024)、运营商与监管—量子通信(0.027)。以上对聚类主题的关联分析可帮助研究者探索ICT科学传播文本的各个主题在生产过程和传播过程中的潜在规律,如科学文本作者在创作互联网社交主题相关的文本时,最可能同时写到的内容与投资相关的内容;而在创作人工智能主题的科普文本时,基本不会提及运营商与监管方面的话题。

4 结束语

新媒体中的ICT科学传播话语是一个庞杂、多元、动态、碎片化且高速增长的研究对象,传统的话语分析方式难以全面地呈现其特征与话语体系。本研究尝试将语料库语言学、话语分析和数据挖掘等领域嫁接起来,在T-LAB软件的辅助下,探索新媒体话语对ICT科学知识的建构。本研究发现,“信息技术”和“信息通信”在真实的ICT科学传播中并不受青睐,“信息技术”和“信息化”是更为常用的词汇,而且“信息”“通信”“技术”的搭配和语义韵倾向差别较大。此外,语料库整体的聚类分析结果显示,“网络安全”“市场”“人工智能”这3个大类是ICT科学传播中最侧重的内容,“数字出版”“手机电脑及操作系统”“运营商与监管”方面则较少涉及,而各科学传播网站文本的主题分布并不均衡。从整体来看,网络ICT科学传播表现最突出的是腾讯科学,不仅文本数量遥遥领先,主题分布也最为均衡。Stubbs曾说,“重复的话语结构能证明某种评价性信息不仅仅是个人的判断,也不是乖僻的论断,而是在话语社区中具有广泛的共识。一个单词、短语或句式可以引发一种文化定型”(cultural stereotype)(Stubbs 2001: 215)。在新媒体中重复出现的话语(结构)会日积月累地、潜

移默化地影响它的读者。通过语料库的分析,能清晰地呈现其对文本对象,即ICT的传播定型和塑造,为当下的ICT科学传播话语做出描画和解析。

注释

- ①http://education.news.cn/2015-09/19/c_128247007.htm, 2017.
- ②<http://news.qq.com/cross/20170303/K23DV601.html#2>, 2017.
- ③<http://www.aihanyu.org/cncorpus/index.aspx>, 2017.

参考文献

- 陈功 梁茂成. 型式语法的产生、特点及其应用价值[J]. 外语学刊, 2017(1). || Chen, G., Liang, M.-C. The Origin, Features and Applications of Pattern Grammar[J]. *Foreign Language Research*, 2017(1).
- 梁茂成 李文中 许家金. 语料库应用教程[M]. 北京: 外语教学与研究出版社, 2010. || Liang, M.-C., Li, W.-Z., Xu, J.-J. *Using Corpora: A Practical Discourse* [M]. Beijing: Foreign Language Teaching and Research Press, 2010.
- 中国信息通信研究院. 2016年泰尔ICT深度观察[M]. 北京: 人民邮电出版社, 2016. || The China Academy of Information and Communications Technology. *CAICT Insight on ICT-2016* [M]. Beijing: Posts and Telecom Press, 2016.
- Baker, P., Gabrielatos, C., Mcenery, T. *Discourse Analysis and Media Attitudes: The Representation of Islam in the British Press* [M]. Cambridge: Cambridge University Press, 2013.
- Conboy, M. *The Language of Newspapers: From Socio-historical Perspectives* [M]. London: Continuum, 2010.
- Fairclough, N. *Language and Power* [M]. London: Longman, 1989.
- Gerbner, G., Gross, L., Morgan, M., Signorielli, N. Living with Television: The Dynamics of the Cultivation Process [A]. In: Bryant, J., Zillman, D. (Eds.), *Perspectives on Media Effects* [C]. Hillsdale: Lawrence Erlbaum Associates, 1986.
- McIlvenny, P. Heckling in Hyde Park: Verbal Audience Participation in Popular Public Discourse [J]. *Language in Society*, 1996(25).
- Stubbs, M. *Words and Phrases: Corpus Studies of Lexical Semantics* [M]. Oxford: Blackwell, 2001.

定稿日期: 2019-04-10

【责任编辑 王松鹤】