

基于《人民日报》新闻报道语义网络的广西壮族形象建构

胡中全¹ 樊 茜², 通讯作者

(1. 广西民族师范学院文学与传媒学院, 广西 崇左 532200;

2. 广西民族师范学院学生工作部(处) 通讯作者, 广西 崇左 532200)

摘要: 无论是传统媒体抑或是新媒体时代, 媒介形象是社会发展趋势的动态映射, 更是社会关系的情感彰显, 建构良好的媒介形象既可以准确把握社会的发展动态, 又能够正确处理社会各因素之间的关系。新闻报道所建构的广西壮族形象潜移默化地影响着大众对于广西壮族整体形象的认知和理解, 从而内化为其辨识和勾勒广西壮族形象的一套价值倾向和思维方式。本文以《人民日报》1958-2018年间涉及广西壮族的所有新闻报道为研究样本, 借助计算传播学的文本分析和语义网络分析, 再结合特定的历史背景和时代特色推导和凝练其所建构的广西壮族形象。

关键词: 人民日报; 语义网络; 广西壮族; 形象; 建构

广西壮族自治区是壮族最主要的聚居地区, 据最新统计结果显示, 广西壮族自治区的壮族人口现今已超过 1520 万人, 约占壮族总人口的 88%。广西壮族地区的政治、经济和文化等方面的发展, 在中华民族的发展中占据着重要的历史地位和作用。壮族的文化影响深远, 其有着悠久的历史渊源, 中国南方的古骆越人是壮族的先祖, 随着历史的发展和沉淀, 壮语、壮医、壮族音乐和壮族舞蹈等构成了一套完整的文化体系, 其蕴藏着深厚的文化内涵和历史底蕴。

广西壮族作为中国最大的少数民族群体, 一直受到党、国家和社会的普遍关注, 《人民日报》对于广西壮族的新闻报道也在不断地塑造其民族形象。自新中国成立以来, 中国社会经历了一系列的变革, 从“全面建设社会主义时期”到“改革开放”再到“新世纪深化改革”, 作为中国南大门的广西壮族地区自 1958 年自治区成立以来, 其不同历史时期呈现出了不同的发展态势, 在这段历史路径中也经历着社会的大变革和形象的转变。在过去的 60 年间, 《人民日报》对于广西壮族的新闻报道从一定程度上可以视为党报对其媒介形象的建构。

随着信息技术的不断发展, “大数据”时代的来临给我们的社会生产、文化教育带来了更多的可能, 在给予我们机遇的同时也带来了更大的挑战。“大数据”分析技术可以通过计算机语言获取大量的数据资源, 并且对这些数据资源进行处理、分析和整理, 然而这种数据资源在正常状态下并不能即时而全面地获得。现今大数据分析技术在学术研究中也被广泛运用, 越来越多的学者开始使用大数据技术进行数据的抓取、处理和分析, 其中作为人文社科领域的研究, 运用最多的研究手段则是数据挖掘、文本分析和数据可视化。通过大数据分析技术, 可以让研究者在海量而迅捷的社会信息中挖掘出信息的隐性价值。本文就是基于此背景, 尝试运用大数据研究方法再结合新闻传播学的相关理论, 分析新闻文本中所“隐匿”的社会价值。

本文尝试将语义网络分析引入媒介形象的研究中, 并以自治区成立以来 60 年的社会变革为背景, 将文本与实际情境相结合, 避免孤立的文本分析。语义网络分析是内容分析法的一种, 又称作关系内容分析。首先通过可视化数据采集软件 Octoparse 对《人民日报》1958-2018 年间有关广西壮族的新闻报道文本进行采集和存档, 通过对文本的筛选和过滤形成有效的新闻文本数据库; 随后将文本数据导入 ROST 中经过提取高频词、过滤无意义词、提取行特征、构建网络形成关键词共现矩阵; 接着将关键词共现矩

阵导入 Gephi 中生成可视化语义网络, 让各个有关因素的相互关系直观呈现; 最后结合各个历史时期的时代背景和新闻传播理论推导、总结和凝练广西壮族的媒介形象。

一、研究数据和方法

(一) 研究数据与预处理

本文所采集的新闻报道文本数据数量庞大, 时间跨度为 1958-2018 年, 且由于版权和平台限制, 所以本文的文本数据将通过第三方代理公司的数据库和中国知网数据库采集获得。《人民日报》数据库是大型的、纪实的、珍贵的资料库, 其具有相当的学术研究价值和珍藏价值, 所以其数据库的访问需要特定的权限和平台。本文所爬取的数据源并非来自“人民数据网 (data.people.com.cn)”中的《人民日报》图文数据库, 而是依托青苹果数据中心为南宁师范大学图书馆提供的《人民日报》数据库试用通道。青苹果数据中心的《人民日报》数据库虽经人民日报社授权, 但是其收集的《人民日报》数据库中的新闻文本只收录 2003 年, 故第一部分的数据采集为 1958-2003 年, 2003-2018 年后的相关文本数据则通过访问中国知网获得。

1. 文本数据采集

文本数据的采集分为两个部分, 第一个部分通过“青苹果数据中心”为南宁师范大学所开通的试用通道进行采集。主要操作步骤如下:

打开可视化软件 → 进入自定义采集界面 → 输入采集网址 → 搜索栏输入关键词“广西僮族” → 选定“主题标题和主题内容”选项 → 建立采集循环步骤 → 分别设置采集内容和标题 → 数据的导出和储存。在这里需要强调的是, 对于 1965 年之前的新闻文本需搜索“广西僮族”才可获得“广西壮族”的相关文本。

第二段数据的采集也根据以上采集步骤进行, 在这里需要注意的是, 由于防止采集数据过快, 导致数据库后台的反爬虫, 需要对流程进行一些设置, 执行前等待 10 秒, 点击后若网页无变化时自动重试, Ajax 加载数据中 Ajax 超时设置为 3 秒钟。数据采集的第一阶段为 1958-2003 年, 以“广西壮(僮)族”为主题的相关新闻数据 4724 条; 第二阶段为 2004-2018 年, 采集数据 2927 条, 两个阶段合计 7651 条。

2. 文本数据预处理

将 7651 条数据导入 Excel 后, 通过删除重复项和筛选等操作对数据进行预处理。首先由标题这一列进行筛选, 对重复标题的内容进行删除, 由于本文着重分析新闻文本, 所以图片新闻不在考量范围内, 也一并删除。在如此庞大的数据之中, 有一些数据其中虽涉及到广西, 但是其多出现于“政府工作报告”“行政区域的划分”“……名单”等新闻报道中, 并不是关于广西壮族地区的新闻报道, 所以在标题列, 通过输入关键词“关于……决定”“工作报告”“行政区域划分”“全国十大(佳)……”“全国……名单”“第……届关于全国……名单”“……当选”等字眼, 进行筛选删除。最后对新闻标题中涉及到其他省份的内容也一并删除, 特别是与广西壮族自治区临近的几个省份, 如广州、海南、四川、云南、贵州等, 这些报道中大多数只提及广西壮族地区, 并未着重报道。总体筛选后, 再通过人工阅读对文本内容进行评估, 删除不符合广西壮族主题的文本, 最终获得有效文本数据 6170 条。

3. 时间段划分

本文以广西壮族自治区成立以来社会主义发展的特定历史时期为基础,再结合广西壮族自治区在特定历史时期所表现的发展态势和重要作用,将整个历史时期划分为全面建设社会主义时期(1958-1965)、十年特殊时期(1966-1977)、改革开放时期(1978-1999)和新世纪深化改革时期(2000-2018)。

(二) 数据的处理和分析

1. 文本分词和关键词提取

为了弥补 ROST 分词功能的潜在缺陷,本文采用可视化分词软件 NLPIR 汉语分词系统进行预分词,其分词准确率高达 98%,主要通过词语的词频和权重值提取关键词,通过 NLPIR-Parser 的用户词典和提取新词的操作,对 ROST 的分词功能进行完善。出

于对特定时期历史背景和时代特点的考量,本文分词和关键词的提取主要以特定时期内的新闻文本作为单独模块进行分词处理,上文划分的第一个历史时期为全面建设社会主义时期,所以对 1958 年至 1966 年的新闻报道进行分析,共计 736 个样本。首先,在分词之前需要对分词系统的用户字典进行更新,以达到分词的准确率,在这里需要对所要分词的文本数据进行新词提取,如“僮族”“农业机械化配件”“榨糖”“备耕检查”“归国华侨”等,随后导入用户词典。其次,从广西壮族新闻报道的总库中筛选出 1958 年至 1966 年的所有文本数据,存入 TXT 文档,编码格式为 UTF-8。最后,导入到 NLPIR-Parser 中,进行分词和关键词的提取,并存入 TXT 文本,最终共获得词语 18617 个,其中名词为 n、动词为 v、新名词为 n_new,提取的部分关键词(如表 1)。

表 1 分词和关键词提取结果

Word (词语)	Part-Of-Speech (词性)	Weight (权重)	Frequency (词频)
生产	vn	396.79	1289
人民	n	278.36	3072
群众	n	214.48	1394
工作	vn	206.32	1582
领导人	n_new	168.34	596
民族	n	154.23	1065
僮族	n_new	135.85	280

通过 NLPIR-Parser 对关键词进行提取,选择文本权重在 20 以上的关键词词数共 966 个,对无意义词进行去除,共得到 754 个有意义关键词。同样的,在 ROST 中根据 NLPIR-Parser 中所获的新词对分词自定义词表进行更新,如“广西僮族”“人民”等,由于 ROST 可以通过用户自行定义高频词的选择数量,所以本文中为了与 NLPIR-Parser 做更好的对比,高频词的上限设定在 754,然后进行提取高频词和去除无意义词的处理,获得 641 个关键词。两组关键词做对比后得到 600 个权重和频词较高的关键词。在这里需要说明的是,由于之后三个时期的样本数量逐渐增多,在 NLPIR-Parser 筛选关键词下限的权重也会相应提高,随后三个时期均选择权重为 30 以上的关键词,平均基数为 830;相应的 ROST 中所选取的高频词也会相应增加,所选取的高频词上限也设定为 830。

二、生成词频共现矩阵和语义网络

在对 ROST 的分词自定义词表进行更新且经过多次调试之后,确定在第一个历史时期中,在 ROST 中所选择的高频词数为 600,为了更为全面地对文本内容进行可视化,所以构建网络和词频共现的词数均设置为 400。首先将新闻报道的文本数据导入 ROST 中,通过 ROST 高频词提取、过滤无意义词、提取行特征和构建网络等操作后生成词频共现矩阵列表。随后将共现矩阵导入 Gephi 中,共获得 728 条边和 108 个节点,对数据进行模块化、图密度、PageRank 等统计。最后在 Gephi 中进行过滤、外观、布局等操作以及细微的图像布局调整后得出可视化图像。其余三个时期都按照此操作路径获得可视化图像。

三、研究结果和分析

(一) 新闻报道数据整体分布



图 1 1958-2018 年广西壮族新闻报道数量趋势图

通过分布图我们可以看出，如图 1，从 1958 年起，《人民日报》对于广西壮族地区的相关报道呈现螺旋式的上升过程，1958 年、2008 年、2010 年、2012 年和 2018 年这 5 年，其报道篇幅都超过了 200 篇；1962 年、1967 年、1976 年这 3 年，其报道篇幅不到 50 篇，这里值得注意的是 1967 年广西壮族的新闻报道仅有 13 篇。这其中的变化有着明显的时代“印记”，1958 广西壮（僮）族自治区成立，广西壮族地区的社会发展受到了党和国家的格外关注，广西壮族地区的政治、经济、文化等方面也获得了翻天覆地的变化。在全面建设社会主义时期和“十年特殊时期”，其对于广西壮族地区的报道较少，主要原因有：第一，《人民日报》版面限制；第二，这两个时期整个社会局势和状态处于“非常”之中；第三，广西壮族地区地处偏远，经济发展水平落后。在改革开放初期，广西壮族地区作为中国对外开放的“南大门”，其受到了党和国家的高度重视，但是由于地处祖国南疆要地，广西依然处于战争前沿阵地，其社会生产和经济的发展受到了巨大的影响和阻碍，所以其相关的新闻报道也相对较少。新世纪深化改革开放时期，特别是自 2002 年《中国—东盟自由贸易区》开始筹建到正式启动以来，广西壮族地区面向东盟进一步深化对外开放，其开放的深度和广度不断拓展，其相关新闻报道也在不断上升。

(二) 各时期新闻报道的语义网络

1. 全面建设社会主义时期

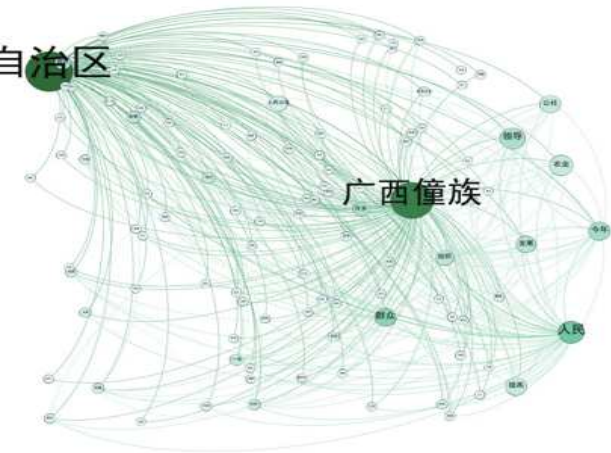


图 2 1958 年 -1966 年《人民日报》广西壮族相关报道语义网络

从图 2 的语义网络详细图我们可以看出，主要分为政治板块和经济板块，政治板块中出现“领导”“组织”“群众”等关键词，经济板块出现“社会主义”“农业”“发展”等关键词。《人民日报》对于广西壮族地区农业生产的新闻报道较多，其中主要有两个原因，一是农业社会主义改造完成了农民个体所有制到社会主义集体所有制的转变，激发了广大农民发展农业生产的活力。二是自广西壮族自治区成立以来，人民当家做主，农业、工业和畜牧业等得到持续发展。广西壮族地区作为经济欠发达地区，在这一时期主要以农业生产为主，见图 3。为了进一步分析广西壮族地区的经济发展状况，本文通过对“经济”“农业”“工业”等与经济相关的关键词进行搜索，通过重复项排查共获得 409 个样本，对样本数据进行可视化，见图 4，这一时期广西壮族地区的经济生产主要为农业生产。

2. 十年特殊时期

运用同样的步骤和方法，得到《人民日报》1967-1976 年广西壮族相关报道语义网络，如图 5 所示，这个时期主要模块：“主席”“群众”“干部”“路线”，可以看出这一时期《人民日报》

主要的新闻报道涉及政治方面。将图 5 放大后，我们可以看出，广西壮族在这一时期的农业仍有一定程度的报道。为了验证这一结论，再一次根据前文步骤进行可视化，见图 7，这一时期的农业生产受到强烈的政治运动冲击。



图 3 《人民日报》1958-1966 年广西壮族语义网详细图

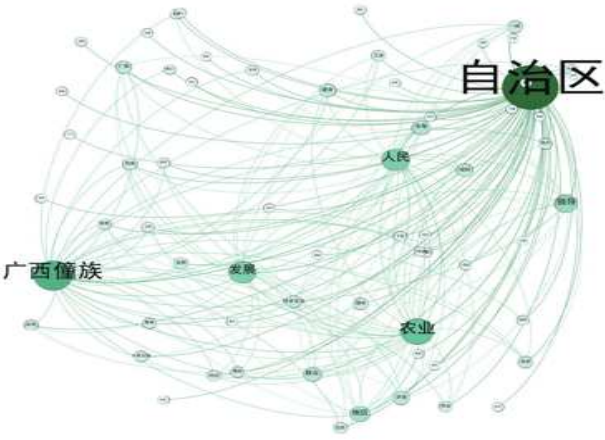


图 4 《人民日报》1958-1966 年广西壮族地区经济语义网络

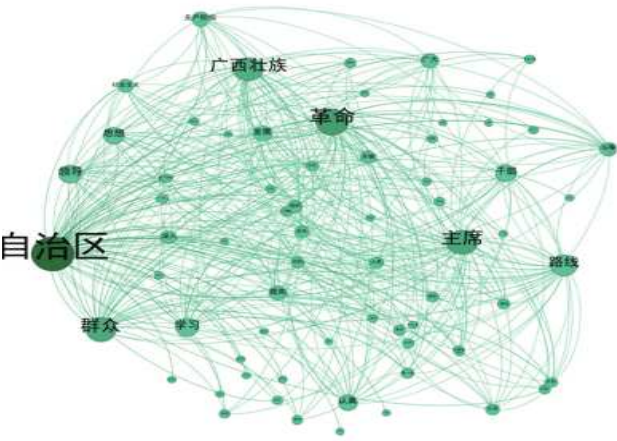


图 5 《人民日报》1967-1977 年广西壮族形象语义网络

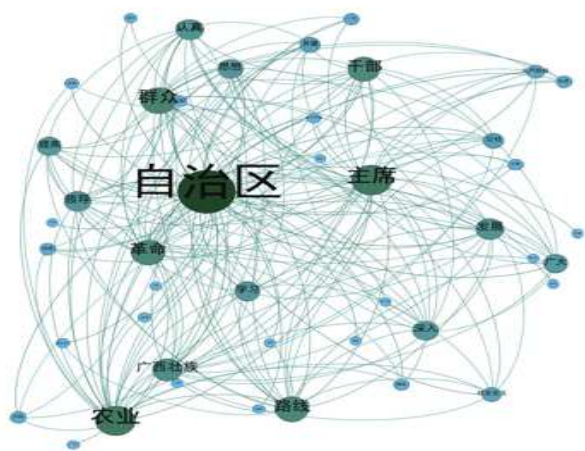


图6 《人民日报》1967—1977年广西壮族地区经济语义网络

3. 改革开放时期

十一届三中全会之后,伴随着家庭联产承包责任制和改革开放的确定和实施,广西壮族地区也迈进了一个新的发展阶段,见图7。由于特殊的历史原因和地理位置,在改革开放初期,广西壮族地区担负起了重大的历史使命,保卫祖国南疆。也正是这一历史原因,加之地处边陲山区先天条件不足,其经济发展和对外开放水平总体落后于其他省(市)。九十年代后,祖国南疆局势实现缓和与安定,特别是“南巡讲话”后,再次掀起改革开放的风潮,广西与东南亚国家的交流渐渐增多,广西壮族地区的经济发展水平和对外开放程度也慢慢提升,见图8。

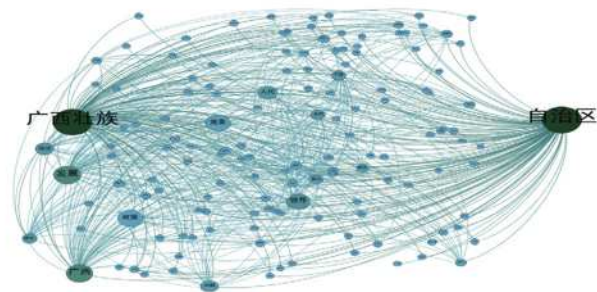


图7 《人民日报》1978—1999年广西壮族语义网络

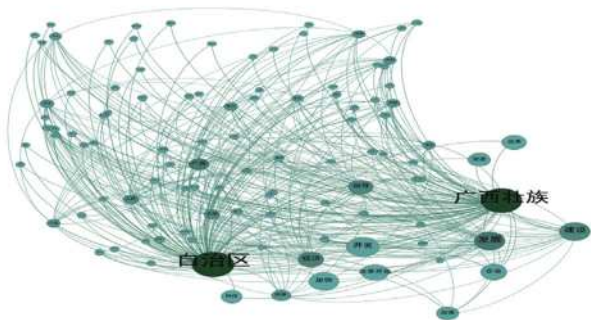


图8 《人民日报》1990—1999年广西壮族语义网络

4. 新世纪深化改革开放时期

进入新世纪后,《人民日报》对于广西壮族地区的新闻报道主要侧重于经济和政治方面,见图9。这一时期,广西的经济发展速度得到了很大的提升,特别是2004年以后,随着“中国—东盟

盟博览会”的成功举办以及“中国—东盟自由贸易区”的正式全面筹建,广西依托其得天独厚的区位优势国家的政策倾斜,迎来了新的历史机遇,见图10。然而作为西南欠发达的广西壮族地区,其农业人口众多,经济基础薄弱,人民生活水平相对较低,《人民日报》关于广西壮族地区的新闻报道在一定程度上侧重于“扶贫”等相关信息,见图11。总而言之,这一时期广西壮族地区地方经济的发展和深化对外开放齐头并进。



图9 《人民日报》2000—2018年广西壮族语义网络

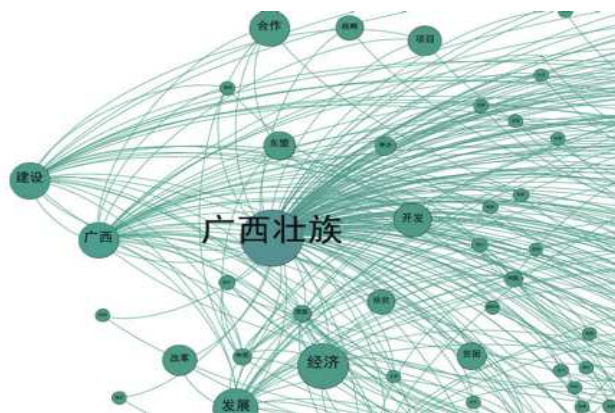


图10 《人民日报》2000—2018年广西壮族经济语义网络

四、广西壮族的媒介形象

本文以不同历史阶段为大背景,在数据分析和可视化呈现结果的基础上,再结合不同历史时期的社会发展特点,推导、总结和提炼出了五种媒介形象,分别为受助者、党和国家拥护者、西南边疆安全守护者、社会变革践行者和新时代开拓者。

这五种形象的建构具有纵向性,是一个发展、进步、承接的过程,是广西壮族形象在中国社会不断变革发展中的动态呈现,具有一定的继承性和时代性;但同时又暗含着横向性,“受助者”形象随着历史沿革变为“新时代开拓者”后,“受助者”形象并未消亡,而是被新闻报道所“隐匿”。总而言之,此文所提炼的广西壮族形象在蕴含时代性和历史性的同时又具有此消彼长的动态性,是社会不断变革发展的必然结果。

(一) 受助者

在广西壮族地区新中国成立前,壮族人民遭受到“三座大山”的多重压迫和剥削,其地区经济发展水平极其落后,政治民主更是无从谈起。新中国成立后,推翻了阶级压迫和剥削,广西壮族人民的身份发生了巨大转变,从被压迫者变为新社会的主人,无论是社会经济的发展还是民主政治的确立,其都受到了很大程度的帮扶。

改革开放以后,随着农业体制不断深化改革和各项富民政策的实施,广西壮族作为西南边陲的欠发达地区,通过合理利用其丰富的自然资源和旅游资源,大力发展粮食作物和经济作物的生产,在受惠于政府相关帮扶政策的同时,也受到了社会的各种援助。

新世纪之后,广西更是依靠着独特的区位优势 and 国家的深入对外开放政策,促成“中国—东盟自由贸易区”的全面启动,把广西建成面向东盟的南大通道,具有“一带一路”有机衔接的重要门户地位,是“海上丝绸之路”的重要节点。在地区的文化教育和基础设施的建设方面也极大地受惠于国家少数民族政策的倾斜,获得物力、财力和人力的专项投入,广西壮族地区人民的生活幸福感得到了极大的满足。

(二) 党和国家拥护者

广西壮族人民自从新中国成立以来,对国家的各项政策给予了极大的支持和肯定,无论是在改革开放前风雨飘摇的年代还是在改革开放后的和平建设时期,广西壮族人民对党和国家及其提出个各项政策和方针都给予最大的拥护。

广西壮族自治区正式成立,广西壮族地区以及其所在的省份拥有了自治权,《人民日报》报道了大量的庆典活动,在接下来的历史时期,报纸中会规律性的出现关于广西壮族自治区周年庆的报道,这些规律的庆典活动在表达对民族政策极力拥护的同时有体现了对党和国家的拥戴。在社会局势纷乱的时代,广西壮族人民积极参与农业生产,大力贯彻实施国家的政策和方针。进入新世纪后,随着经济发展、文化教育、社会民生等民族政策和措施的进一步倾斜,《人民日报》通过广西壮族鲜明的个人事迹和话语表达体现了对党和国家的拥护和爱戴。

(三) 西南边疆安全守护者

1979年改革开放初期,当全国各地都在播起战鼓进行改革开放,但就在这一关键时期,我国西南边疆地区却屡生事端——遭到外部侵扰和挑衅,边疆的安宁与人民的生命财产安全受到严重的威胁。广西壮族地区作为西南边境地区,积极组织民间力量英勇抗敌,保卫祖国南疆的和平与安全,为国家的改革发展稳定作出了贡献。

20世纪90年代,西南边疆地区逐渐恢复稳定,中国与东南亚各国的交往也日益密切,广西壮族地区由于其优越的区位优势,始终充当着周边外交的重要“使者”,与周边国家的合作、交往日益密切,更是接待各国贵宾和访问团的首选之地。

新世纪以后,随着“中国——东盟自由贸易区”全面启动和“一带一路”战略的实施,广西壮族地区抱着“居安思危”的态度不断守护着中国的西南边境,为中国社会的协调发展、民族团结和伟大复兴默默坚守。

(四) 社会变革践行者

解放之初,广西壮族人民对党和国家的政策方针贯彻到底,坚决实施;随后,依法行使自己民族区域自治的权利,自主管理本民族,处理本民族的事物,真正做到当家做主。社会主义建设“探索失误时期”和“十年特殊时期”,虽然广西壮族地区乃至整个中国都经历了前所未有的重创,但是由于广大工农干部群众的辛勤劳动,经济仍能保持较快的速度发展。这个时期壮族人民所经历的挫折是其价值和信仰的外在表现,而其践行者的形象并没有中断过。

改革开放之后,家庭联产承包责任制、社会主义市场经济、对外开放等一系列的经济发展政策相应出台,广西壮族人民以极大的热情谋求发展、因地制宜、吸引投资,因其丰富的资源优势 and 区位优势,农业、手工业、旅游业等产业得到了巨大发展。

进入新世纪后,面向东盟的广西壮族地区迎来的新的发展机遇,“一带一路”战略的提出,“中国—东盟自由贸易”的全面启动,广西壮族人民正不断践行着实现“中国梦”、实现“民族复兴”的伟大使命。

(五) 新时代开拓者

广西壮族地区作为中国的南大门,有着得天独厚的区位优势,

自21世纪以来,国家积极促进“中国—东盟博览会”的胜利举办,积极促进“中国—东盟自由贸易区”的全面启动。近年来,中央赋予广西“三大定位”,着力构建面向东盟的国际大通道、打造西南地区发展的新战略支点、形成“一带一路”有机链接重要门户。广西壮族地区作为面向东盟、面向世界的新开拓者,拥有巨大机遇和动力。广西壮族地区正在进行着巨大的社会变革,其将从西南边陲的“偏隅”之地转变成成为面向东盟的国际信息交流、贸易往来、合作洽谈和周边外交的战略重地。

五、结语

本文尝试运用计算传播学的研究路径,通过对新闻文本数据的采集、预处理、分析、构建语义网络和数据可视化,呈现《人民日报》对广西壮族新闻报道的整体态势,我们发现,由于国家政策、经济建设水平和文化发展水平的差异,其在不同的历史阶段对于广西壮族的新闻报道存在不同的报道倾向。

《人民日报》对于广西壮族媒介形象的建构起到了一定的作用,多见于政治和经济方面的报道,但是受众对于广西壮族的“刻板印象”并没有得到明显的改善。受众认知和“刻板印象”的改变是一个持久漫长的过程。总而言之,《人民日报》对广西壮族形象的建构总体上是比较成功的,也为媒介形象的建构提供了诸多有效的策略,但是在多次的调研和深度访谈中我们了解到,广西壮族由来已久的“地方偏远”“经济发展水平低”等方面的“刻板印象”并没有得到有效的改变,由此,我们需要进一步思考如何改进形象建构策略,注重少数民族文化的传播和软性新闻的报道能够在一定程度上改变大众对于广西壮族的“刻板印象”,从某种程度上能够促进广西壮族形象的建构。

参考文献:

- [1] 高焕静.《人民日报》(1950—2014)少数民族形象建构研究[D].浙江大学,2015.
 - [2] 祝建华,张丽华,黄显.计算传播学与传播研究范式转移[J].青年记者,2018(22):4-5.
 - [3] 张伦.计算传播学导论[M].北京:北京师范大学出版社,2018.
 - [4] 王扬,田野,李铁山,陈俊龙,彭东成,周义华.基于Gephi的航运招聘信息可视化分析[J].大数据,2018,4(03):81-91.
 - [5] 纪娇娇.基于语义网络分析的转基因议题框架量化研究[D].中国科学技术大学,2017.
 - [6] 中共广西区党委党史研究室.广西壮族自治区50年纪事[M].南宁:广西人民出版社,2008(77).
 - [7] 张媛.官方媒体中的少数民族形象建构——基于《人民日报》少数民族报道的分析(1979—2010)[J].国际新闻界,2013(08).
- 基金项目:2022年度广西高校中青年教师科研基础能力提升项目“计算传播学范式下的广西非遗文化传播路径研究”(2022KY0749);2021年度广西民族师范学院校级科研一般项目“广西民族影像与民族传播路径研究”(2021YB058)。

作者简介:胡中全(1990—),男,安徽六安人,硕士,广西民族师范学院文学与传媒学院教师,研究方向:计算传播学研究,少数民族文化传播研究。

通讯作者:樊茜(1993—),女,壮族,广西上林人,硕士,广西民族师范学院学生工作处,研究方向:少数民族社会工作研究。