

基于Citespace的壮语研究知识图谱浅析

唐蒙

广西民族师范学院文学与传媒学院, 广西 崇左

收稿日期: 2023年5月16日; 录用日期: 2023年7月12日; 发布日期: 2023年7月21日

摘要

本文以CSSCI数据库为数据来源, 运用Citespace可视化软件从发文量、作者、机构、研究热点等角度对1998年至2020年间壮语研究文献进行可视化分析。研究发现, 壮语研究的发文量2017年以前呈曲折发展形式, 数量不稳定, 2017年之后呈稳定发展状态, 核心作者较少, 作者及研究机构没有形成紧密的合作关系, 研究热点主要集中在本体研究和语言接触研究。

关键词

壮语研究, 知识图谱, Citespace

Analysis of Knowledge Graph of Zhuang Language Research Based on Citespace

Meng Tang

School of Literature and Media, Guangxi Minzu Normal University, Chongzuo Guangxi

Received: May 16th, 2023; accepted: Jul. 12th, 2023; published: Jul. 21st, 2023

Abstract

This paper uses the CSSCI database as the data source, and uses Citespace visualization software to conduct a visual analysis of Zhuang language research literature from 1998 to 2020 from the perspectives of publication volume, authors, institutions, and research hotspots. The study found that before 2017, the volume of publications on Zhuang language research showed a zigzagging development pattern, and the number was unstable. After 2017, it showed a stable development state, with fewer core authors, no close cooperative relationship between authors and research institutions, and research hotspots mainly focused on ontology studies and language exposure studies.

Keywords

Zhuang Language Research, Knowledge Map, Citespace

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

壮族是我国少数民族中人口最多的一个民族，拥有自己的语言，壮语是其日常生活中的重要交际工具，是壮族文化得以传承的重要载体。随着国家对语言资源保护和研究队伍的不断壮大，利用科学的现代技术手段，多角度对少数民族语言文字进行科学保护，保持民族语言活力，铸牢中华民族共同体意识。壮语的研究内容也不断深化，在壮语语音、词汇、语法等领域均有丰硕的成果，但截至目前，学界还未有学者使用文献计量的方法将壮语研究的既有成果经过可视化统计分析。因此本文将借助 Citespace 软件系统将抽象的壮语研究变换为可视化的空间结构和知识图谱进行分析，把现有壮语研究成果的主题、热点及其现状，以相对客观、科学的角度显现出来，同时试图探寻壮语研究的演化路径和发展趋势。

2 研究方法与数据来源

2.1. 研究方法

本文利用 Citespace 5.8.R3 软件对我国壮语研究方面的相关文献进行了定量分析，并利用可视化方法展示了壮语研究的结构、规律和分布情况，在绘制知识图谱的基础上，将对近几十年来壮语研究的学者及其所在机构的影响力、合作网络、关键词的贡献网络和聚类结果进行分析，总结壮语的研究历史与现状、发展脉络及核心议题。CiteSpace 是一个在科学计量学、数据可视化背景下逐渐发展起来的一款引文可视化分析软件，主要用于分析科学分析的潜在知识，它是由美国德雷塞尔大学陈超美教授开发的，通过可视化的手段来呈现科学知识的结构、规律和分布情况，也称之为“科学知识图谱”[1]。

2.2. 数据来源

为保证研究的可靠性和权威性，保证学术含量。本文以中文社会科学引文索引(CSSCI)为数据源，在数据库中以“壮语”为题名或以“壮语”为关键词；发文年代选择从 1998 至 2020 年，文献类型为论文，检索得到相关研究的论文 175 篇，剔除其中 1 篇与壮语研究无关的文章，最终以 174 篇作为研究的有效数据源。

3. 可视化结果分析

3.1. 壮语研究发文趋势分析

年发文量的趋势变化对探讨壮语研究的热度及其发展趋势具有一定的价值。由图 1 可见。

在 1998~2020 年之间，我国对壮语研究的相关文章呈现出不断上升的趋势，在 1998 年到 2001 年期间，论文数量逐年上升，壮语研究慢慢受到重视；2001 到 2004 年发文量稳定，每年发文 3 篇，表明壮语研究暂时进入一个停滞瓶颈期；2005 年开始发文量回升，2006 年发文量 13 篇，接近历史最高点，表明壮语研究再次成为研究的热点。从 2007 年又逐步回落，降到 2011 年的 5 篇；2012 至 2015 年发文量处在

稳定阶段，年均发文量达到 10 篇以上；2016 年仅发表一篇，处于观察期内的最低点；2017 年年往后至今，年发文量稳定在 10 篇以上，成上升趋势。

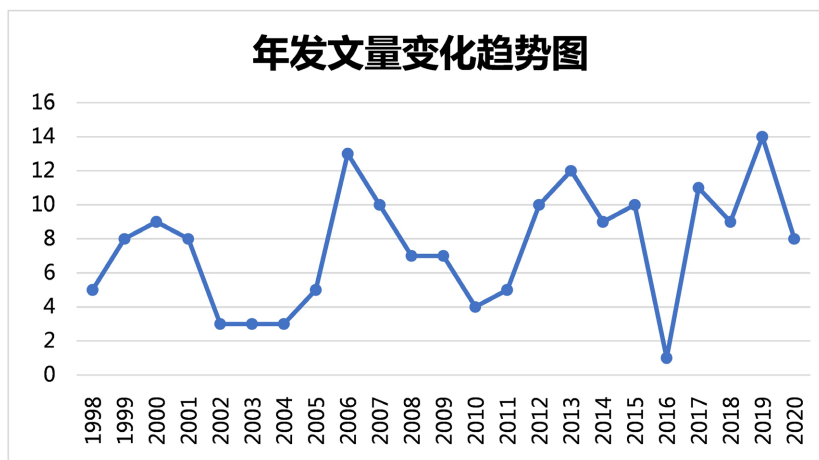


Figure 1. The trend chart of the annual publication volume of Zhuang language studies
图 1. 壮语研究年发文量变化趋势图

由以上的趋势图及分析可知，壮语研究的热度不稳定，但总体的研究还是呈现出上升的趋势，尤其是 2012 年之后发文数量显著增多，虽然 2016 年骤降，但是这期间的发文总量比 1998~2011 年有所增加。一方面是壮语研究团队的不断强大，对于壮语的研究更加深入；另一方面，受到国家语言政策的影响，语言资源保护意识不断加强，研究的手段和方法也更加多元化，发文量也随之增长。

3.2. 学者影响力和合作分析

借助 Citespace 软件的数据筛选，研究壮语的学者中，发文量前三的是覃凤余教授(18 篇)，覃晓航教授和吕嵩崧教授各 9 篇，韦景云教授(8 篇)。除发文量第一的覃凤余教授外，其他作者之间的差距较小。通过分析作者分布数据发现，壮语研究的 174 篇 CSSCI 文章共有作者 96 人，人均 1.81 篇，其中发文数量 5 篇及以上的作者人数为 9 人，发文数量 1 篇的作者人数是 60 人。由此可见，发表 1 篇论文的作者占据了作者总人数的 62.5%，高产作者不足 1% (表 1)。

Table 1. List of prolific authors (Papers ≥ 3)
表 1. 高产作者列表(文章 ≥ 3)

作者	发文量	作者	发文量
覃凤余	18	李心释	4
覃晓航	9	韦名应	4
吕嵩崧	9	李旭练	4
韦景云	8	林亦	4
何霜	6	梁敢	3
韦达	5	曾晓渝	3
蒙元耀	5	黎良军	3
班昭	5	李锦芳	3

Continued

蓝庆元	5	薄文泽	3
黄平文	4	黄静露	3
依常生	4		

学者合作网络图谱中节点的大小代表了作者发表论文的数量，节点间的连线表示学者间存在合作关系。在 CiteSpace 功能与参数设置区的 Node Type 选择为 Author，以 5 年为一个时间片段，运行得到的作者的合作网络如图 2 所示，图谱中存在合作关系密度很低，96 位学者间的密度仅为 0.0073，说明学者间的联系密不密切，相互之间的合作较少，学者们都偏向于独立发展，暂未形成系统性的合作关系。广西大学覃凤余教授位于图谱中心，其与黄阳、覃东生等有过合作，中央民族大学覃晓航教授、韦景云教授和百色学院吕嵩崧虽也是高产学者，但其作品均为独著居多。随着现代科学技术的发展，学者间的学术交流与合作对学科的发展起着不可忽视的重要作用，因此从事壮语研究学者，尤其是年轻学者之间需要扩大合作关系，拓展交流面，群体合作提升科研能力。

CiteSpace, v. 6.2.R4 (64-bit) Basic
 June 27, 2023 at 1:08:38 PM CST
 WOS: C:\Users\lenovo\Desktop\zhuangyuyan2023\data
 Timespan: 1998-2020 (Slice Length=5)
 Selection Criteria: g-index (k=25), LRF=3.0, L/N=10, LBY=5, e=1.0
 Network: N=100, E=38 (Density=0.0073)
 Largest CC: 9 (9%)
 Nodes Labeled: 1.0%
 Pruning: Pathfinder
 Modularity Q=0.9151
 Weighted Mean Silhouette S=1
 Harmonic Mean(Q, S)=0.9657

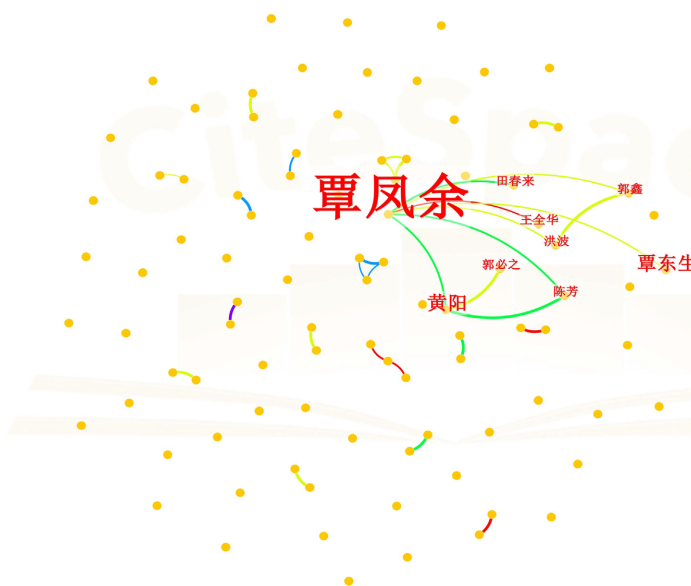


Figure 2. The cooperation network map Zhuang language research scholars
图 2. 壮语研究学者合作网络图谱

3.3. 发文机构及合作网络分析

在 CiteSpace5.8.R3 软件中选择 Institution，将 Thresholds 中的各调节键调整为 0，得到 N = 50, E = 11 的可视化图谱，图谱中的 N 表示的是 174 篇文章出自 50 个研究机构，E 代表研究机构之间的合作关系。对同一机构不同命名进行合并，如将广西民族学院合并到广西民族大学。合并后我们可以发现，壮语研究的主体机构主要是高等学校，专门研究机构只有中国社会科学院研究较多。中央民族大学、广西民族

大学和广西大学是发文量最高的主题机构，总发文量占 50%，各共发表论文 29 篇。其次是百色学院，共发表论文 9 篇，中国社会科学院民族学与人类学研究所发表论文 7 篇。其余的研究机构发文较少，这些学校部分开设有壮语的相关专业或课程(图 3)。

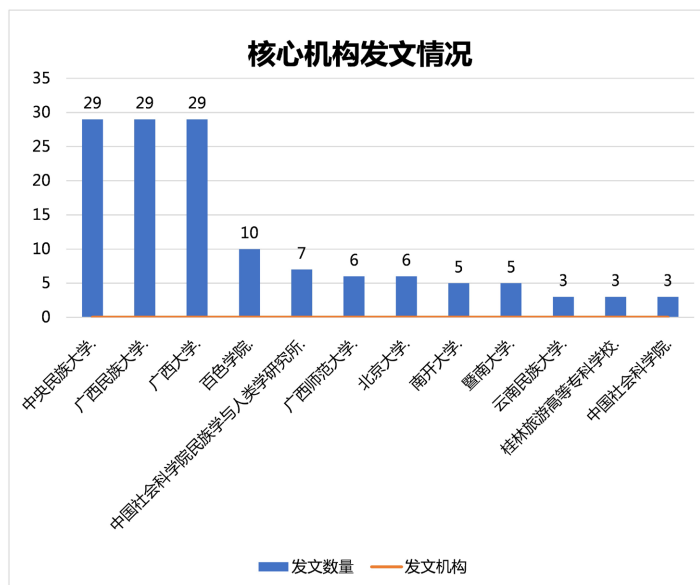


Figure 3. The core institutions
图 3. 核心机构发文情况

通过研究机构共现分布图可以发现(图 4)，机构合作网络中存在 50 个节点和 11 条连线，点多线少，说明机构之间的合作力度不足。11 条连线联结的合作网络有 5 个，第一个是以中央民族大学为中心，与中国社会科学院民族学与人类学研究所、桂林旅游高等专科学校、中国社会科学院研究生院、《三月三》杂志社均有过紧密合作；第二个是以广西大学为中心，与首都师范大学、上海师范大学、广西社会信息调查部有合作；第三个是以北京大学为中心，与广东技术师范学院和中国民族语文翻译局有过合作；第四个是广西民族大学和广西师范大学有过合作；第五个是鲁东大学和兴义民族师范学院合作。发文量较多的几个机构之间只有中央民族大学和中国社会科学院民族学与人类学研究所合作较为紧密，其他的核心发文机构之间尚未形成合作关系。

总体分析来看，进行壮语研究的各个机构几乎没有进行合作，结合前文对学者论文产出情况分析，我们可以看到：同一机构内的学者的合作产出成果较多，而跨机构的学者之间缺乏合作，产出的学术成果较少，这也就导致机构间的合作非常少，如广西大学的发文数量达到 29 篇，覃凤余教授就发表了 18 篇，其与校内的覃东生教授、黄阳教授等合作较多，但是与校外合作的非常少。由此可见，目前我国进行壮语研究的学者之间、机构之间还有很大的发展空间，各机构之间可以进行资源互补，共同推进壮语研究工作的发展。

3.4. 研究热点分析

研究热点从壮语研究中的关键词可以窥探出来，如果不同的文章选择了相同的关键词，关键词出现的频率代表研究议题的研究热度，分析研究热点有助于我们从宏观上把握壮语研究领域的现状，分析发展规律和趋势。



Figure 4. Co-occurrence distribution map of research institutions

图 4. 研究机构共现分布图

使用 citespace 5.8.R3 软件对关键词绘制可视化图谱如下图 5，观察壮语研究的关键词共现图谱可以发现：图谱中共存在 345 个节点，连线 696 条，网络密度为 0.0117，字体越大颜色越深表示该关键词出现的频次越高，除了壮语外，热度较大的语言接触、语法化、汉语、平话等，这说明目前壮语研究的热点主要为壮语与其他语言的接触，语言内部语法化过程。

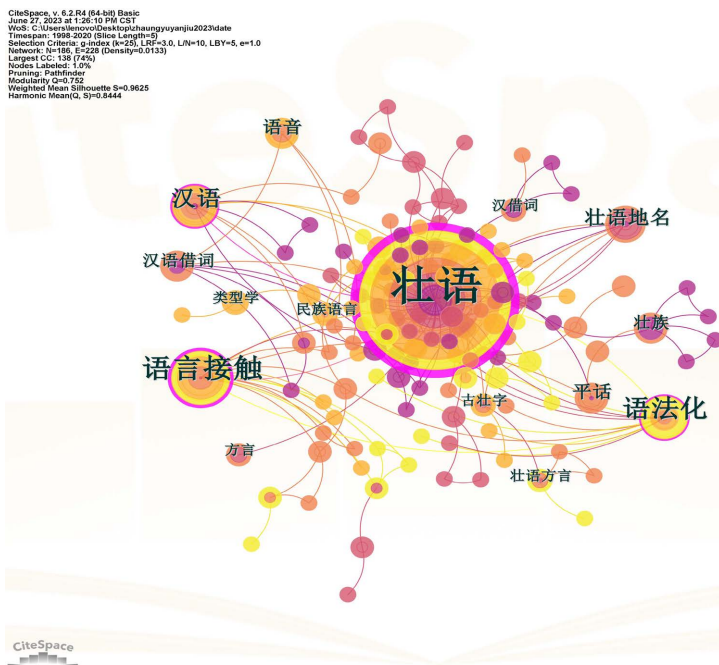


Figure 5. Keyword co-occurrence map

图 5. 关键词共现图谱

对关键词的分析除了共现图谱之外,还需要根据关键词进行聚类分析(图6),“聚类”是衡量一个学科研究状况及其热点主题的重要指标之一,从中可以分析诸多热点关键词之间存在的内在联系。借助运行 citespace 软件中关键词的聚类整合功能,我们最终得到 19 组聚类结果,聚类图谱中 Q 值为 0.708, S 值为 0.9456,这说明聚类结果符合要求,可信度较高。

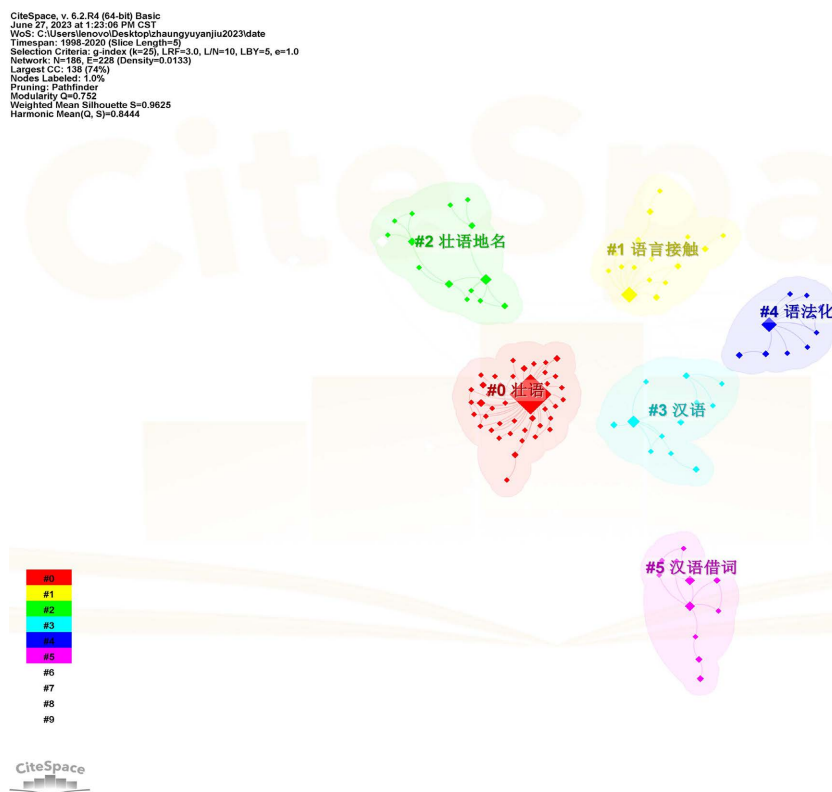


Figure 6. Keyword clustering map
 图 6. 关键词聚类图谱

壮语的关键词聚类图谱出现了#0 壮语、#1 都安壮语、#3 汉借词、#4 音变、#5 汉语借词、#6 文马土语、#7 壮语方言、#8 民族语言、#9 壮语地名、#10 壮语借词、#11 早期分布、#12 壮语语音、#13 中越边境、#14 音义、#15 变异、#16 回顾、#17 送气音、#18 方式助词等 19 组聚类。将聚类进行整合后发现 1998 年至 2020 年壮语研究的热点领域有两个。

3.4.1. 壮语方言研究

壮语方言研究一直是以来都是壮语研究的热点。对壮语方言研究较深的学者主要有覃凤余、李旭练、韦景云、班昭、何霜、韦达等。覃凤余对壮语语法进行了大量的描写研究,如量词定语标记、使役动词、分类词、方位词指示词等方面。李旭练主要研究都安壮语的形态变化和语法化,对都安壮语的语音屈折和形容词性相对比较句进行了细致的描写和研究[2]。韦景云主要从语言地理学的角度对壮语的音变和内部演化进行研究。班昭的研究主要是邕宁壮语,涉及动植物名称词的研究[3]。何霜从语法化和类型学的角度对忻城壮语的语法进行了细致的描写。韦达对壮语本体的研究角度比较特殊,主要从文化色彩方面映射描写壮语的词汇,以期探讨壮族的思维方式和文化心理特征,探寻壮族先民的早期分布和社会生活状况[4]。

3.4.2. 语言接触研究

语言演变主要有两个途径：一是语言内部自身的发展，如语音脱落、语流音变等；二是各种语言之间的相互接触产生的变化，如词汇的相互借贷、双语现象等。壮族作为中国人口最多的少数民族，历来与周边民族相处融洽，语言接触频繁，语言成分之间的相互借用和吸收，从语言接触的角度分析壮语，有利于我们探寻壮语的发展演变规律，了解壮族与其他民族的交流交融历史，有助于铸牢中华民族共同体意识。

壮语语言接触研究较多的关键词是借词和语言接触。其中研究语言接触最多的学者是吕嵩崧、覃远雄。吕嵩崧主要对德靖土语区的壮语与汉语方言接触借用进行研究，内容涉及德靖土语某些壮语特殊的语音和语法现象，比如汉语对靖西壮语构词法的影响及借用等，充分描写了德靖土语区的语言借用情况，以小做大，对某些语言现象进行细致入微的描写[5]。覃远雄则讨论平话、粤语中的某些语音、词汇受壮语影响的现象，如其提出“壮语从早期汉语吸收了‘与’，后来又带入了现代平话和粤语”[6]的观点，认为壮族在长期的民族接触中转用平话和粤语，因此在现代平话和粤语中有壮语的底层，最终发展成为现代平话和粤语的一部分。李心释主要对平话和壮语的接触产生的变异研究。

蒙凤姣探讨了壮语对瑶语布努土语的影响，从瑶族布努土语反身代词及动词“去”用法的不同之处进行分析[7]。这是在试验阶段唯一一篇专门从民族语之间的接触来探讨语言接触的文章。

3.5. 壮语研究热点分析

Citespace 软件的突变检测功能能够获取关键词的突变值，利用短期内多次重复出现的词显现出该领域研究热点的演变趋势。

Top 10 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	1998 - 2020
壮语	1998	1.45	1998	2000	
壮族	1998	1.39	1999	2002	
壮语地名	1998	1.93	2001	2007	
语法功能	1998	1.49	2006	2009	
平话	1998	2.01	2007	2012	
语音	1998	1.52	2009	2017	
民族接触	1998	1.32	2009	2009	
语言接触	1998	1.68	2010	2012	
类型学	1998	1.63	2013	2018	
语义演变	1998	1.39	2017	2020	

Figure 7. List of emerging words in Zhuang language studies

图 7. 壮语研究突现词列表

由图 7 可知，壮语研究的聚焦点在“壮语”、“壮族”、“壮语地名”、“语法功能”、“平话”、“语音”、“民族接触”、“语言接触”、“类型学”、“语义演变”等。每个热点在短期内都呈爆发趋势，随着时间的推移，研究热点在不断改变，结合前面的关键词频次我们可以看出，壮语的研究在随着语言学的热点的脚步向前发展，从最初的壮语语言本体研究到现在的壮语语言类型学研究，这些热点都紧随目前语言学的前沿理论。

4. 结论与启示

我们通过使用 citespace 可视化软件，对 CSSCI 数据库中的 1998~2020 年间的权威壮语研究论文进行量化统计分析发现：第一，从文献的发表情况看，壮语研究权威文章的发表在 2016 年之前处于不稳定状态，2017 年至今产量稳定，呈缓慢上升状态；第二，从作者和研究机构合作情况看，壮语研究的学者之

间合作关系不紧密,只有少量作者具有合作团队,且多为师生关系或同机构同事关系,部分高产出学者基本都是“单打独斗”,合作率较低,同时研究机构间的合作力度不够,只有中央民族大学进行相关的跨机构合作较多,其他的研究机构之间几乎没有合作关系;第三,从研究的关键词和热度分析来看,壮语研究的关注点主要集中在本体研究和语言接触研究,但是随着越来越多新理论的引进,壮语研究的关键词越来越丰富。

因此,在未来的壮语研究中,学者、机构之间应进行更多的合作与沟通;扩大壮语研究视角,在重视本体语言研究的同时也应注重研究壮语古籍文献,结合计算机科学、历史学、生态学等学科进行多元化研究,推动跨学科发展。

基金项目

广西民族师范学院 2019 年度民族学硕士点建设科研项目“崇左跨国婚姻家庭语言生活调查研究(2019MYB018)”;左江流域民族文化研究中心 2019 年度科研课题“左江流域壮族语言使用的代际差异研究(2019ZJYB006)”。

参考文献

- [1] 李杰. CiteSpace 中文指南[EB/OL]. <http://blog.sciencenet.cn/blog-496649-886962.html>, 2015-05-03.
- [2] 李旭练. 都安壮语的屈折形态[J]. 民族语文, 2008(2): 65-67.
- [3] 韦景云. 语言接触视域下壮语流音*r 的特征分布[J]. 中央民族大学学报(哲学社会科学版), 2015, 42(4): 136-141.
- [4] 韦达. 壮语词汇的文化色彩——壮族语言文化系列研究之二[J]. 中南民族大学学报(人文社科版), 2002, 22(3): 81-84
- [5] 吕嵩崧. 壮语 la(:)i1 的语义演变及其对广西汉语方言的影响[J]. 民族语文, 2020(3): 30-41.
- [6] 覃远雄. 平话、粤语与壮语“给”义的词[J]. 民族语文, 2007(5): 57-62.
- [7] 蒙凤姣. 壮语对瑶语布努土语的影响[J]. 民族语文, 2017(5): 95-96.