

大数据与 AI 技术赋能广西壮语活态传承的路径研究

蒋佳祁月, 孙大正

南宁理工学院, 广西 南宁 530100

DOI: 10.61369/TACS.2025090046

摘要：本研究以广西壮族自治区壮语为样本，聚焦少数民族语言面临的代际传承断层、技术适配不足、文化传播瓶颈的危机。当前，壮语 15 岁以下青少年使用率较十年前下降，公共领域使用频次减少，诸多词汇随长者离世而消失。为此，研究提出大数据与 AI 技术赋能的活态传承新路径，构建“文献研究+问卷调查+田野调查”技术框架。研究团队深入多个壮族古村落，借助智能设备收录语料，建成壮语动态语料库，并开发 AI 学习平台、文化传播工具等，推动语言资源数字化保存与创新传播。实证结果显示，该模式实施一年后，预计语言保护效率提升，青少年学习参与度得到大幅提高，试点学校口语达标率迅速增长，为多民族地区语言可持续发展提供了范式参考。

关键词：广西壮语；传承发展；大数据；语言保护

Research on the Path of Empowering Live Transmission in Guangxi Zhuang Language by Big Data and AI Technology

Jiang Jiaqiye, Sun Dazheng

Nanning Institute of Technology, Nanning, Guangxi 530100

Abstract： Taking Zhuang language in Guangxi Zhuang Autonomous Region as a sample, this study focuses on the crisis of intergenerational transmission fault, insufficient technical adaptation and cultural communication bottleneck faced by minority languages. At present, the usage rate of young people under the age of 15 in Zhuang language has decreased compared with that of ten years ago, and the frequency of use in the public domain has decreased, and many words have disappeared with the death of the elderly. To this end, the research puts forward a new path of big data and AI technology empowerment in live transmission, and constructs a technical framework of "literature research+questionnaire survey+field investigation". The research team went deep into several ancient villages of Zhuang nationality, collected language materials with smart devices, built a dynamic corpus of Zhuang language, and developed AI learning platforms and cultural communication tools to promote the digital preservation and innovative communication of language resources. The empirical results show that one year after the implementation of this model, it is expected that the efficiency of language protection will be improved, the participation of young people in learning will be greatly improved, and the rate of reaching the standard of oral English in pilot schools will increase rapidly, which provides a paradigm reference for the sustainable development of languages in multi-ethnic areas.

Keywords： Guangxi Zhuang language; inheritance and development; big data; language protection

少数民族语言是民族文化的“活态基因库”，承载着特定族群的历史记忆、思维范式与文化认同，对维护全球文化多样性具有不可替代的价值^[1]。广西作为壮族聚居核心区，壮语不仅是区域交际工具，更承载壮族歌谣、仪式祝词等非遗传承使命，是壮族文化得以延续的核心载体^[2]。然而，全球化与城镇化冲击下，壮语传承陷入危机：代际使用断层、公共领域空间收缩、独特方言词汇面临消逝风险。传统保护模式（人工记录、线下教学）受时空限制，且主流数字技术对壮语适配不足，进一步加剧传承难题。有声民族语言是作为民族文化的重要载体，它承载的不仅是民族的表层有形的文化形式，还蕴含着当地居民更深层次的地域文化内涵，为民族地区民族语言文化传承指明了方向。基于此，本研究探索“大数据+AI”赋能新路径，推动壮语从“静态保护”转向“动态活化”，不仅为广西壮语保护提供实践方案，更可为其他濒危少数民族语言的传承提供理论参考，助力多民族语言文化生态的维护与发展^[3]。

一、壮语传承困境的多维解析

（一）代际断层危机

民族文化的传承有其内生机理^[4]，以广西南宁市武鸣区（壮语

使用核心区）为调研样本，数据显示壮语使用呈现显著的代际割裂。年长群体长期以壮语作为日常交流与文化表达的主要语言，但与之相反的青年群体多使用通用语言，壮语掌握程度与使用意愿不足。

这一断层的成因可归结为三重因素：一是城镇化推动壮族群众融入多元环境，公共服务与社交以普通话为主，压缩壮语空间；二是部分学校未开设壮语课程，且存在“方言影响普通话学习”的认知偏差；三是年轻父母优先用普通话亲子交流，弱化家庭传承场景。若此趋势持续，壮语可能退变为“文化记忆符号”，导致文化基因断裂与民族认同弱化，动摇壮语活态传承的核心基础。

（二）技术适配不足

当代数字技术的快速发展为语言传承提供了新可能，但壮语在与主流数字技术的适配中面临诸多障碍：AI 翻译、语音识别工具常出现识别偏差与语义误读，降低使用便利性；更深层的技术瓶颈在于标准化语料库的缺失^[9]。壮语存在北部、南部两大方言区，方言内部又有多个次方言，不同区域在发音、词汇、语法结构上差异显著。这种语言多样性导致统一标准化语料库建设难度极大，进而导致 AI 工具的识别精度、语义理解能力及应用范围均受限，无法满足不同方言区的实际使用需求^[9]，使壮语陷入“技术滞后—使用减少—传承乏力”的恶性循环，在数字化时代逐渐被边缘化，进一步加剧传承困境。

（三）文化传播瓶颈

壮语文化的传播面临“形式固化”与“群体疏离”的双重挑战，难以有效触达年轻群体，传播形式多数依赖传统线下场景传播，受时空限制显著，导致文化辐射范围狭窄，无法覆盖广泛的青少年群体。新媒体平台中的壮语文化内容严重匮乏且形式单一，缺乏互动性与创新性，壮语文化在新媒体领域的“缺席”，使其与青少年的信息获取习惯严重脱节。这种本质是“传承方式与时代需求的错位”的传播困境，使壮语最终陷入“老年人坚守、年轻人旁观”的传承僵局，威胁文化传承的可持续性。

二、大数据赋能壮语活态传承的技术框架

（一）整体技术路线

依托数字化现代技术对区域有声民族语言传承具有重要的时代意义，特别是像广西这样少数民族语言文化资源丰富的民族地区，对区域有声民族语言数据库进行构建具有重要意义，特别是对有声基础壮语进行传承，对未来的有声壮语研究和传承提供重要参考。本研究技术路线以“服务壮语文化活态传承”为核心，围绕“多维度数据整合—动态语料库构建—场景化应用开发—实践推广优化”形成闭环式传承体系：

在数据整合阶段方面，通过田野调查、文献数字化、现有资源整合、新媒体挖掘，采集文本、语音、图像、视频等多模态语料；在语料库构建阶段方面，以标注方言类型、使用场景、文化背景，形成兼具多样性与文化关联性的动态语料库；在应用开发阶段方面，围绕“学习+传播”两大核心需求，开发适配壮语传承的 AI 应用工具，包括面向青少年的 AI 学习平台，以及面向文化传播的场景化工具，将壮语学习与文化体验相结合；在实践推广阶段方面，在壮语核心区的学校、社区落地应用，实现“开发—试点—优化”的持续改进，推动壮语从“静态存档”向“动

态活化”转变。

（二）核心模块建设

1. 多模态数据收集体系

数据收集以“覆盖壮语文化全貌”为目标，采用“全域覆盖+重点聚焦”策略，确保语料的文化代表性与完整性：以田野调查记录原生语料，同时整合广西现有壮语资源补充方言数据，并合规挖掘音频平台中的壮语内容，丰富现代使用场景语料。

2. 语料库构建与优化

壮语动态语料库以“语言多样性与文化关联性”为核心，兼顾数据规范化与文化完整性：标注壮汉平行语料优化语义理解，破解“同词异义”难题；建立“定期采集+用户反馈”更新机制，补充新语料，确保时效性。

3. AI 应用开发导向

AI 应用开发以“文化传承为核心、用户需求为导向”，围绕壮语学习与文化传播设计功能，选用轻量化大模型，以多模态语料训练优化语义理解；学习类工具开发词汇学习、发音指导功能（融入民间故事、歌谣）；采用跨平台框架打包为小程序、APP，扩大传播范围，让壮语文化以更鲜活的形式触达年轻群体。

三、实证研究与社会影响

（一）数据采集成果

语言是一个民族乃至一个国家重要的文化载体，是民族民俗文化发展的缩影和重要体现，民族语言的发展和传承不仅有利于传承当地的民族文化，对提升民族文化自信、促进民族团结以及民族融合和各民族多元文化的发展也具有重要的意义。本研究以南宁市武鸣区（壮语使用核心区）为核心调研区域，通过系统性田野调查与数据整合，完成多维度语料采集；采集不同年龄段使用样本，为代际传承干预提供依据。

（二）创新成果与社会影响

1. 教学模式革新

形成“AI 辅助+教师引导+文化浸润”的新型壮语教学范式；AI 承担词汇教学、发音指导，教师聚焦文化讲解与互动，如词汇教学同步讲解背后民间传说，强化文化属性。

2. 技术适配突破

开发的壮语公众号与小程序支持多方言语音识别、发音评测及场景化翻译，解决主流技术适配不足问题，成为青少年接触壮语的重要渠道。

3. 文化传播矩阵

构建“平台+内容+活动”体系，举办“校园壮语短视频大赛”，推动壮语文化从“地域性”向“大众化”传播。

4. 传承成效提升

通过“大数据+AI”技术赋能，壮语传承与保护成效显著，语言资源数字化保存率大幅提升，保护效率提高、成本降低；青少年学习意愿与文化认同增强，壮语使用场景逐步拓展，从“被动保护”向“主动传播”转变的趋势明显，壮语的使用更为便捷，为壮语传承注入了新活力。

5. 社会价值凸显

数字化工具有效激发了青少年对壮语及壮族文化的兴趣与认同，研究构建的技术框架可复制性强，为其他少数民族语言保护提供参考；新媒体传播突破时空限制，促进跨民族交流，推动壮语文化从“小众传承”向“大众认知”转变，为文化多样性保护注入了新动力，也为全球濒危少数民族语言的传承提供了有益借鉴。

本研究以广西壮语为样本，系统探索大数据与 AI 技术赋能少数民族语言活态传承的路径，核心结论如下：技术赋能缓解传承困境，大数据解决资源碎片化，AI 激活语言使用活力，推动壮语“动态活化”；同时构建了“文化导向”活态模式，使壮语回归日

常场景，实现经济效益与社会效益双赢：提升保护效率、降低成本，增强民族认同，为少数民族语言传承提供可复制范式，为少数民族语言文化的可持续发展提供了有力支撑。

为进一步推动少数民族语言的可持续发展，结合本研究成果提出以下建议：一是建立“政府—高校—科技企业”协作机制：政府提供政策与资金，高校负责语言研究与语料标注，企业开发技术；二是深化技术场景创新，打造“壮语文化数字空间”；三是加强跨区域与跨学科合作：推动西南省份共建“少数民族语言数字联盟”，借鉴国际经验构建开放数字生态，为人类命运共同体的文化生态建设贡献力量。

参考文献

- [1] 陈贝贝. 数智赋能广西非遗文化影像的活态传承与发展路径探析 [J]. 文化创新比较研究, 2024, 8(10):105-110.
- [2] 周映辰, 吴启健. “非遗”视野下广西武鸣壮族山歌的保护与传承 [J]. 音乐教育与创作, 2025, (01):104-108.
- [3] 王慧. AIGC 技术赋能马山壮族刺绣产业化转型发展 [J]. 西部皮革, 2024, 46(23):46-48.DOI:10.20143/j.1671-1602.2024.23.046.
- [4] 杨春, 杨军. 离散与聚合: 民族文化传承的内生机理及其现实启示——基于毛南族生态文化基因的研究 [J]. 广西民族研究, 2024, (04):156-164.
- [5] 黄丽丽. 数字技术赋能马山苗族服饰保护传承与发展 [J]. 西部皮革, 2023, 45(09):24-26.DOI:10.20143/j.1671-1602.2023.09.011.
- [6] 贺梦凡, 熊乐佳, 丁盈, 等. 数字技术赋能农业文化遗产开发与利用对策研究 [J]. 农业图书情报学报, 2023, 35(03):71-80.DOI:10.13998/j.cnki.issn1002-1248.22-0822.