

濒危语言尔苏语的语料库建设及应用

王 轲¹, 王 轩², 王德和³, 古 涛³

(1. 尔苏藏族文化研究中心, 四川 西昌 615000;

2. 西南民族大学, 成都 610041; 3. 西昌学院, 四川 西昌 615000)

摘要:中国西南地区尔苏语使用人口少,语言传承面临断代的危险。抢救保护和传承濒危的尔苏语言迫在眉睫。通过采集有声语料,建立数据库,创建尔苏拼音方案,用ELAN、FLEX等软件标注,构建由原始音频文件、国际音标标注、汉语翻译标注、尔苏语标注的语料库,并利用语料库编撰尔苏语词典和尔苏语学习有声教材。尔苏语语料库的建立能有效典藏尔苏语言。

关键词:尔苏语;多媒体词典;有声;典藏;母语

中图分类号:H274 **文献标志码:**A **文章编号:**1673-1883(2018)03-0077-05

Construction and Application of Corpus in Endangered Language of Ersu

WANG Ke¹, WANG Xuan², WANG De-he³, GU Tao³

(1. Institute of Ersu Tibetan Culture, Xichang, Sichuan 615000, China; 2. Southwest University for Nationalities,

Chengdu, Sichuan 610041, China; 3. Xichang University, Xichang, Sichuan 615000, China)

Abstract: The Ersu language of Southwest China has few speakers and its intergenerational transmission is in the process of being broken. It is urgent for these reasons to document and protect the endangered Ersu language. This article reports some concrete steps in this direction: the creation of an Ersu language database with audio files, development of the Ersu Romanization System, processing of Ersu language data using linguistic softwares, such as ELAN and Flex, annotation of Ersu data in the database in the International Phonetic Alphabet and the Ersu Romanization System, and translation of the Ersu data into Chinese. One more initiative is to use the Ersu language database to compile a dictionary of Ersu and pedagogical materials with audio files. The creation of the Ersu language database can effectively contribute to the preservation of Ersu.

Keywords: Ersu language, multi-media dictionary, audio files, language documentation, native language

一、引言

尔苏语是20世纪80年代中国西南地区新发现的小语种之一,按照孙宏开教授的划分,属于藏缅语族羌语支南支。主要分布在南丝绸之路零关道一带的四川省越西县、甘洛县、石棉县和汉源县。语言作为“重要的交际工具”具有承载民族文化的作用,尔苏人的“故事、歌谣、传说、神话、诗歌、谜语和民间信仰人士的口诵经等”,具有丰富的“文化内涵”^[1]。目前在甘洛、越西等地的大部分尔苏人在日常生活中用母语交际,在汉源等地的已改用汉语。尔苏语东部方言人口1.6万人,母语使用率约50%。抢救保护和传承濒危的尔苏语迫在眉睫。我们通过对尔苏口语进行收集、整理、录入计算机保存;创

建尔苏拼音方案;用Elan、Flex等软件标注,构建了由原始音频文件、国际音标标注、汉语翻译标注、尔苏语标注一一对应的语料库。创造了抢救濒危尔苏语的必要条件。通过对语料数据库的整理和加工,编撰尔苏语词典、制定实用的尔苏语言学习教材、研究尔苏语言文化。

二、尔苏语语料库的构建

语料库“顾名思义就是存放语言材料的仓库(或数据库)。传统上,语言学家用语料库这个术语表示可作为语言研究基础的、大量自然出现的语言数据。这些语料库可以由书面语和(或)口语的样本组成,并通常被用来代表一种特定的语言或语言变体。”^{[2]1-2}语料库的建设不仅仅是语言学家的责任与

收稿日期:2018-04-14

基金项目:英国伦敦大学濒危语言联盟;包含插图和音频的尔苏语多媒体词典(SG0487);四川省教育厅项目:凉山美姑县俄助家族毕摩经书研究(15SB0132)。

作者简介:王轲(1993—),女(藏族),四川甘洛人,文学学士,研究方向:尔苏语言文化。

义务,更需要学界乃至社会各界大量有志之士的参与。语料库建设的受益者也不仅是尔苏藏族人民群众自身,也包括对尔苏文化感兴趣的民族学、人类学等各学科的专家学者。构建尔苏语语料库符合《国家中长期语言文字事业改革和发展规划纲要(2012—2020年)》将“科学保护各民族语言文字”的方针政策;建尔苏语语料库的目的是世界共享、真实记录和永久保存及文学文化遗产,促进尔苏语语言文化的研究及推广应用,丰富中华语言文化宝库。

(一)数据的搜集整理

以王德和研究员为主的尔苏藏族文化研究中心,约一半成员为尔苏语母语者,在十余年的田野调查工作中他们既是发音人又是工作人员和研究者,在与被访问对象的交流中使用尔苏语沟通,亲戚朋友非常配合。本文笔者之一、团队负责人王德和先生是受过良好教育的尔苏母语者,高校藏语言文学研究员。他熟练地用Elan、Flex等软件从语料库中抽取词条,归类、释义、标注和翻译,同时还要对课题组的工作成果进行验收和把关。课题组成员都接受过大专或中专以上的教育,长期在尔苏社区生活,掌握尔苏语言,具有用尔苏语言记录、分析文本的能力,在翻译和标注过程中能真正反映出母语的真实情况和精髓。

尔苏藏族文化研究中心课题组在工作当中非常重视田野调查的科学性,遵循学科前辈们在以往研究中形成的惯例和总结的经验。在搜集整理田野调查数据的同时,也注重科研伦理问题。一方面遵守国家法律法规,另一方面尊重民族习惯。课题组充分利用尔苏藏族的射箭节、火把节、还山鸡节、腊吉沽太阳神节等传统节日到越西县、甘洛县、石棉县、西昌市等地采集素材。当地藏学会也给予大力支持,组织人员参与,每次都收获颇丰。

王德和研究员的弟弟王连和家,在甘洛县城郊新民村,父母俩都是耄耋之年。王连和是共产党员,曾经担任村干部,他家是甘洛县“五好文明家庭”,在当地群众特别是尔苏人里有较高的威望。逢年过节兄弟姐妹七家人常常团聚。这个四世同堂的大家庭,共有四十多人。还融入了汉族、彝族,这些外族人大多能听懂尔苏话,有人还说得很好,小孩子也能听会说。和谐的大家庭聚会毋庸置疑是采集素材的大好时机。多台照相机、摄像机、录音机一齐上阵,大家都在为尔苏语料库的建设贡献力量。祭祖仪式、口诵经文、民间故事;杀猪宰羊、酿酒、食物制作、对唱、舞蹈、会话、辩论等。各个仪式空间,各种仪式符号都是采集素材的绝佳对象。有时

还会请附近的亲戚和语言社区民众来补充材料、扩展视野。时常需要分成几个小组进行录音录像。

在采集素材的过程中也注重了发音人的选择,有一定覆盖面,选择不同地区、不同年龄段、不同职业、不同文化程度、不同性别的懂汉语的发音人。存储信息遵循:“一是元数据。包括发音人属性素材。如发音人年龄、性别、受教育情况、对汉语掌握情况、本人对母语使用状况,本人对采集数据的公开使用所持有的意向等。二是发音文本素材。录入和存储发音人发音素材及其国际音标、汉语直译、汉语意译、尔苏语意译等“四项翻译”文本材料。三是实际语音素材。四是存储部分声学分析数据。”^[3]

(二)尔苏语拼音方案设计

我国“语言社会生活”中,汉语文是“社会交际效率最高的语言工具”,“少数民族语文处于绝对的劣势”;随着现代信息技术的发展,“正在不断加大汉语文和民族语文之间的差距”,因此“一方面,少数民族需要尽快掌握和普及国家通用语言文字,以在市场机制体制下平等地获得诸如良好的经济收入、接受教育、就业机会、社会保障、卫生保健、居住条件、公共文化服务等与‘社会效率’有关的语言权利;另一方面,少数民族也应依法享有和行使各项国家制订的倾斜性民族语言政策,以使少数民族在区域自治、母语文化传承和语言群体认同等方面获得与‘社会公平’有关的语言权利”^[4]

此前,尔苏没有通用文字,无法记录自己的语言,在使用汉字等其他兄弟民族的文字记录时又面临无法准确记录的困境。因存在“文化差异”,有部分“文化局限词”,尔苏语和汉语“词汇单位的对应关系难找”^{[6]404}。尔苏人常常说:“写不出来的汉话没有,写不出来的尔苏话很多”,如“我们的宗旨是讲述尔苏人自己的故事”翻译成尔苏语为“aryi nzzenzze ne buerssyi bbe yozai yozai hi bba, yozai yahi gge.(每个人都要讲自己故事)”;“大家齐心协力,全力以赴”翻译成尔苏语为“sela hzyihy da ga lehdda da ga shyimala kavarvar ngu rohji kengu.(大家做到一条舌头,一个膀子;咬紧牙关,使出全力。统一思想、统一意志;心往一处想,劲往一处使)。”“紧密团结起来”翻译成尔苏语为“ddexo ne nddretro, ddehbbi ne zzuhbbi.(捋起来成戟杆,堆起来成肥堆)。”“戟杆”汉语又叫“杆子”,是类似标枪的狩猎武器,用约五米长的木棒做成,是猎杀老熊的专用武器。一些特有的器具、动植物、生活习俗、特定生存环境等,汉语也不能准确翻译尔苏语原意。

这些无文字亚族群的文化记忆,怎样转化为书面

记述就成了首要问题。王德和先生作为尔苏语母语者,在长期的科研过程中苦苦寻找一种合适的书写方式来记录本民族语言。直到2010年,遇到法国国家科研中心齐卡佳研究员。当时她正在研究与尔苏语密切相关的“里汝语”^①。两人联合科研,开展尔苏语音位调查。确定以中国四川省甘洛县则拉乡磨房沟村的尔苏语为基准。采用拉丁拼音转写的方式来创制一套拼音转写系统。王德和在协助齐卡佳教授完成英国濒危语言联盟(ELDP)的重大课题《中国西南地区尔苏语和旭米语抢救》的过程中,共同修订完善了这个拼音转写系统。尔苏语拼音转写系统创立以来,在三年多的试用期内论证、修改几十次。先后在甘洛县、石棉县、西昌市等尔苏聚居区举办了十三期免费培训班,进行试用和推广。通过多年的实践证明,这个拼音系统是行之有效的^[17]。

(三)尔苏语语料库内容

尔苏语语料数据库由原始音频文件、尔苏语标注转写、国际音标标注、汉语等组成,做到一一对应。形式上有词汇、句子、长篇语料。内容上包括风俗习惯、民间信仰、道德伦理、家族迁徙历史、生产知识和技能、生活习惯、医药知识、动植物、传统生态知识、食品、服饰、用具制作过程等。语料上有故事、歌谣、祭祀辞、熟语、对话等。在语料库中约有一万六千条词条。这些语料文化内涵丰富,如果“把语言比喻为博物馆”,“每一种语言中都存储着族群的源流历史、知识体系、文学作品、民族特征,承载着一个族群在历史长河中形成的丰富而深厚的传统文化。“尔苏人”独具特色的文化结晶”正是通过语言“得以代代相传,延续不绝,不断丰富,日益深厚。”^[15]

三、尔苏语语料库的应用现状

(一)《尔苏语词汇通释》

2015年,由王德和等尔苏语言研究人员将尔苏语语料数据库的语料进行整理、归纳。在此基础上加工提炼《尔苏语词汇通释》专著。由安徽大学出版社申报2017年度国家出版基金后获准立项。以四川甘洛县则拉乡尔苏语为标准音,纳入一万余条词汇、短语和例句。每个词条由尔苏语拼音、汉语释义、国际音标、英语、语法例句、汉语意译、对应音频组成。此书“原语”是尔苏语,“译语”是汉语,按字母顺序排列词条,全书约2 000千字,400幅图片。因主要“服务对象是说母语的读者”,所以附上《尔苏藏族简介》《尔苏语概况》《尔苏语拼音转写方案》以及内容一致的电子音频光盘。是一本多语语文词典,同时具备语文性纸质词典和多媒体有声词

典的功能。《尔苏语词汇通释》计划2019年出版。其中包含了海量尔苏人民的传统文化、生产劳动、日常起居、自然环境等信息,具有很高的使用价值、珍藏价值和文献价值。

(二)《尔苏语多媒体词典》

2016年,英国濒危语言联盟和云南玉溪师范大学共同举办了“中国濒危语言典藏国际学术研讨会暨暑期培训”,会上课题组成员王德和教授详细介绍了尔苏语的现状和尔苏藏族文化研究中心的课题组进行的科研活动。来自英国濒危语言联盟(ELDP)的项目部主任,英国伦敦大学的Mandana Seyfeddinipur教授,了解到尔苏语语料库建设现状后,鼓励科研团队积极申报ELDP项目。通过半年的努力,成功立项编写《A Comprehensive Illustrated Dictionary of Ersu with Audio Files(尔苏语多媒体词典)》(SG0487),成为在中国立项的四个课题之一。在课题组挂靠西昌学院图书馆,开展民族文化抢救挖掘和整理工作。课题组成立了专门的尔苏沙巴文化研究小组,目前正在采录高度濒危的尔苏沙巴口诵经,组织科研人员用ELan标注和转写相关文件。下一步工作是将Elan数据导入Flex语料数据库,进行语料分析,导出数据文件。有条件的时候可以付梓印刷,出版成纸本文献。纸本文献将惠存西昌学院图书馆和国家档案馆等相关部门。

(三)尔苏语学习软件(手机版)

课题组依据Flex数据库,建立有声教学资源《尔苏语900句》,开展手机为终端的尔苏语汉语双语教学模式。成果形式是以中国尔苏社区手机用户为对象的尔苏语音视频教学软件。这个以数字化和多媒体形式为主的教材,不同于传统的纸质教材,只要有手机,就能实现语音教学,且易于保存。课题组还利用“尔苏藏族文化”微信公众平台,向受众介绍尔苏文化,逐步开展尔苏语教学。通过“互联网+”模式让母语丢失者和语言爱好者学习尔苏语。此举能够快速实现其社会价值,进而实现其学术价值和文化价值。

四、尔苏语语料库作用及意义

(一)符合党的十九大关于文化建设的精神,有利于推进新时代国家语言文字事业创新发展

习近平总书记在党的十九大报告第七部分《坚定文化自信,推动社会主义文化繁荣兴盛》中指出:“文化是一个国家、一个民族的灵魂。文化兴国运兴,文化强民族强。没有高度的文化自信,没有文化的繁荣兴盛,就没有中华民族伟大复兴。”“党的

十九大胜利召开标志着我国语言文字事业进入了新时代。”“语言不仅具有文化价值,也是国家重要的战略资源。”尔苏语语料库的应用成果践行了“以人民为中心”、“发展和提高”尔苏“公民的语言能力”、“努力做到通过语言文字自信锻铸文化自信”。濒危尔苏语言的抢救保护和典藏,对“社会语言生活和谐健康发展,建设美好的语言生活”和“建设良好的语言生态”^[7]起着积极的作用。有利于推进新时代国家语言文字事业创新发展,有利于“一带一路”建设,有利于提升国家文化软实力。

(二)抢救濒危语言,发展弱势语言,建立和谐社会语言生态

尔苏群众对美好生活的向往,包含了母语丢失者对复活母语的美好语言生活的向往。尔苏语是尔苏人的(亚)族群标志之一,发音独特、词汇丰富,有独特的四套擦音和全球七千多种语言里不多见的翘舌弹音。语料库的建设与词典的研发工作,积极挖掘南丝路沿线积淀的尔苏藏族文化,全面收录南丝路零关道沿线小语种亚族群濒危语言。课题成果可供尔苏人学习尔苏语,领略尔苏语华丽的名言警句,欣赏优美的对仗平仄的诗歌,体会叙事史诗的优美意境;领会先辈理想信念、价值理念、道德规范、审美追求、人文精神。课题成果的推广,可以进一步强化尔苏人的母语意识,转变语言态度。激发尔苏人说母语的积极性和文化自觉,增强民族凝聚力。对于传承尔苏文化、实现文化进步,具有积极的作用。

当前多元文化背景下,社会文化多样化发展。尔苏人长期与汉族、彝族和其他民族和谐共居。他们世代代信奉“敬业乐群、与人为善”的理念和“与人和谐、与自然和谐”的处世哲学。尔苏人语言能力强,在尔苏聚居村落,以尔苏语为交际工具,藏彝汉杂居区则兼用汉语和彝语,汉族和彝族说尔苏人是“有三条舌头的人”^{[8][14]}。和谐社会包括“和谐语言社会”,这种情况“使少数民族语言和国家通用语言和谐共处,多语种并存并用,各司其职,各展其长,各得其所。”多民族共居的村落和族外婚姻的家庭成员,很大一部分人都是能够用汉语、彝语、尔苏语交流,民族关系融洽。“做到在一个双语或多语言的社区内,使所有成员除了使用母语外,还能够熟练地使用社区内所有成员的语言,而且自己母语的活力不会降低。这是一个理想的境界。”^[9]深入挖掘尔苏语言文化,能有效保护和传承少数民族弱势语言。改善少数民族语言不平衡不充分的发展现状。保障尔苏语言的平等地位,提高尔苏语的社会

威望和使用机会。创造尔苏语的教学条件,建立和谐社会语言生活。有利于引导尔苏人树立正确的历史观、民族观、国家观、文化观,更好地服务社会。

(三)语料库应用规划

“语言的生命在于运用,语言的运用促使语言不断地变化和发展,而语言变化和发展的轨迹或特征就隐藏在各种不同的文本中。”通过大量“真实语言文本”,可以实现语料库的“语料管理、语料索引、语料统计、语料标注、语音分析”五个基本功能。^{[10][48-66]}利用尔苏语语料库产生的《尔苏语词汇通释》侧重于尔苏语词汇和固定短语的归纳和注释,《尔苏语多媒体词典》侧重于尔苏语单音节词和词组。今后将利用尔苏语语料库,完成100多个民间故事的标注转写。出版《尔苏藏族民间故事对译》,还要出版《甘洛尔苏语参考语法》《尔苏藏族民间歌谣》《尔苏藏族格言荟萃》《尔苏民间医药》《尔苏农耕文化》等著述。

五、用语料库进行研究

构建尔苏语数据库,容纳海量的尔苏语词汇、短语和固定搭配的名言警句。通过研究,我们发现尔苏语语料数据库的优越性很多。

(1) 过去曾经认为尔苏语的趋向前缀有七个^[11]如表1所示。

表1 过去归纳的尔苏语趋向前缀

向上	向下	水源/向里	下游/向外	靠山/向左	靠水/向右	往回
de ⁵⁵	ne ⁵⁵	k ^h e ⁵⁵	ŋe ⁵⁵	k ^h ua ³³	ŋua ⁵⁵	nyo ⁵⁵

齐卡佳和王德和先生通过数据库归纳总结,在《语言暨语言学》上刊发尔苏语研究文章《甘洛尔苏语动词的体貌范畴》,对尔苏语的前缀做了比较详细的分析,并讨论了动词前缀的不同形式,得出仅现在提取的就有九个趋向前缀^{[12][61]}的结论,如表2所示。

表2 齐卡佳王德和^[12]通过数据库分析的尔苏语趋向前缀

垂上	垂下	斜上水源	斜下水尾	往里	往外	往北	往南	向后	不定	——
de ⁵⁵	ne ⁵⁵	dzi ³³	ŋi	khe ⁵⁵	ŋe ⁵⁵	kua ³³	ŋua ³³	ŋo ⁵⁵	t'e ⁵⁵	原型
da ³³	na ³³	—	—	kha ⁵⁵	ŋa ⁵⁵	—	—	—	t'a ⁵⁵	变体

(2) 通过数据库详细分析研究动词趋向前缀和动词的搭配规律与语义特征,如表3所示。

综上所述,尔苏语语料库的建立及应用,可以深入挖掘、整理、保护和传承尔苏文化遗产。搜集采录弥足珍贵的沙巴口诵经,标注和转写,进行语言分析。既可以补充数据库文件、丰富词典词条,又可以编辑口诵经文本,开展学科建设,为构建沙巴学^[13]奠定基础。

表3 动词加上前缀,构建了新的词义

动词	de ⁵⁵	ne ⁵⁵	dzi ³³	ŋi ³³	k'e ⁵⁵	ŋe ⁵⁵	k'ua ³³
跳	de.əto ⁵⁵	ne.əto ⁵⁵	dzi ³³ əto ³³	ŋi ³³ əto ³³	k'e ⁵⁵ əto ⁵⁵	ŋe ⁵⁵ əto ⁵⁵	k'ua ³³ əto ⁵⁵
跳起跳							
跳下去							
斜向上跳							
斜向下跳							
跳进去							
跳出去							
往北跳							
扛	de.əve ⁵⁵	ne.əve ⁵⁵	dzi ³³ əve ³³	ŋi ³³ əve ⁵⁵	k'e ⁵⁵ əve ⁵⁵	ŋe ⁵⁵ əve ⁵⁵	k'ua ³³ əve ⁵⁵
扛起							
匍匐着							
斜依着							
匍匐在							
分担							
承担							
向北匍匐							
射/锯	de ⁵⁵ tt'o ⁵⁵	ne ⁵⁵ tt'o ⁵⁵	dzi ³³ tt'o ⁵⁵	ŋi ³³ tt'o ⁵⁵	k'e ⁵⁵ tt'o ⁵⁵	ŋe ⁵⁵ tt'o ⁵⁵	k'ua ³³ tt'o ⁵⁵
盐味够							
卦辞指示							
斜向上方射击							
斜向下方锯开							
射击/弹							
射出去							
往北射击							

五、结语

尔苏语既是新发现语言又是濒危语言,“虽然使用人口不多,但这些语言以其结构形式的多样性和独特性以及系属归类上的孤立性展现了其高度的研究价值。”抢救性记录川西南高度濒危的小语种新发现语言,其丰富的文化内涵,能为语言学、文学、民族学、历史学、农学、艺术学等研究提供真实

有效的参考依据。能增强尔苏人母语自觉、文化自觉,减慢语言的衰变速率,延缓和降低濒危等级。具有重要思想价值、学术价值、艺术价值、出版价值、重大文化积累价值和现实意义。

尔苏语语料库的构建对“抢救记录和保护濒危语言、维护语言生态多样性对学术研究、维护我国民族文化的多元化格局至关重要。”^[14]今后的工作任务依然艰巨,语料的安全存储和妥善保管对于技术条件和经济成本的要求较高,文件需经常进行系统、格式、符号的转换,以防止因技术升级而造成数据乱码或文件丢失。虽然语料库建设工作量大,工作难度大,需要大量的人力物力,经费严重不足,几乎没有经济效益。但我们要克服困难,进一步整合、利用这些资源为社会服务,创作尔苏人民需要的文化产品,更好地发挥其应有的学术价值和社会效益。

参考文献:

- [1] 王德和, Katia Chirkova(齐卡佳). 中国濒危尔苏语言抢救保护与尔苏语拼音转写方案的创建[J]. 西南民族大学学报(人文社会科学版), 2017(4):70.
- [2] 黄昌宁, 李涓子. 语料库语言学[M]. 北京: 商务印书馆, 2001.
- [3] 刘岩. 关于中国少数民族濒危语言语音语料库的设计[J]. 中央民族大学学报(哲学社会科学版), 2006(4):134-135.
- [4] 黄行. 当前我国少数民族语言政策解读[J]. 中南民族大学学报(人文社会科学版), 2014(6):7.
- [5] 徐世璇. 语言中的博物馆和语言博物馆——论濒危语言典藏和语言博物馆建设[J]. 玉溪师范学院学报, 2015(5):1,4.
- [6] (捷)拉迪斯拉夫·兹古斯塔(L.Zgusta). 词典学概论[M]. 林书武, 等, 译, 北京: 商务印书馆, 1983.
- [7] 田立新. 推进新时代国家语言文字事业创新发展[N]. 中国教育报. 2017-12-20(1).
- [8] 王德和. 尔苏藏族文化研究[M]. 成都: 四川大学出版社, 2010.
- [9] 孙宏开. 从语言的性质和功能看保护濒危语言的必要性和可能性[J]. 民族翻译, 2014(2):12.
- [10] 章宜华. 计算词典学[M]. 上海: 上海辞书出版社, 2013.02.
- [11] 宋伶俐. 尔苏语动词趋向前缀和体标记[J]. 民族语文, 2006(3).
- [12] CHIRKOVA K, WANG Dehe. Verbal Aspect in Ganluo Ersu[J]. Language and Linguistics, 2017(3).
- [13] 王轩, 王德和, 王轲, 等. 尔苏沙巴文古籍研究的回顾与展望[J]. 广西民族大学学报(哲学社会科学版), 2017(2).
- [14] 李锦芳. 西南地区濒危语言调查研究[M]. 北京: 中央民族大学出版社, 2006.