

少数民族语言资源的精准保护问题

——基于“语保工程”活态数据的考察*

朱德康

[提要] 从“语保工程”对中国少数民族语言资源采录的活态数据来看,少数民族语言资源具有不均衡性、地域性、交互性和共享性特点。通过进一步分析发现,“语保工程”对少数民族语言资源的调查布点不平衡,濒危语言布点单薄,藏缅语族布点超过其他语系、语族,而苗瑶语族布点稀疏。针对这样的事实,本文提出今后的少数民族语言资源保护工作应在已取得成果和经验的基础上,查漏补缺、科学分类、精确管理、精准保护。

[关键词] 少数民族语言 语言资源 语言保护

随着世界经济一体化的发展,全球范围内的语言多样性面临严峻挑战,人们开始认识到语言是一种有价值、可利用、出效益、多变化、能发展的特殊社会资源(陈章太 2008)。随着人们对语言资源认识的提高,保存、保护和开发利用语言资源的呼声越来越强烈。近年来,中国语言资源的保存、保护正在得到政府、高校、科研院所以及一些社会团体的关注,特别是中国共产党十七大、十八大、十九大先后提出科学保护和发展各民族语言文字、弘扬中华优秀传统文化、繁荣发展少数民族文化事业、传承弘扬中华优秀传统文化等以来,语言资源的保存、保护日益成为社会大众普遍关注的问题。

2015年5月,教育部、国家语委印发了《关于启动中国语言资源保护工程的通知》(教语信[2015]2号),决定在全国范围内开展以语言资源调查、保存、展示和开发利用等为核心的重大语言文化工程,这标志着从国家层面以更大范围、更大力度、更加科学有效的方式来开展语言资源保护工作。“少数民族语言调查”作为“中国语言资源保护工程”(简称“语保工程”)的重要组成部分同步启动,计划用5年时间完成少数民族语言420个点的调查任务,其中,濒危语言110个点,非濒危语言310个点;至2020年底,已顺利完成430个点的调查,其中,濒危语言111个点,非濒危语言319个点,调查的范围、语言种类与语言数量均超过预期。调查获得的大量第一手语言资源活态数据,不仅有利于保存少数民族语言资源,也有助于学界对少数民族语言资源国情作出新的判断和认识。

本文基于“语保工程”民族语言调查的实践,梳理少数民族语言资源特性,并通过对已采录语言资源数据的考察和分析,尝试提出少数民族语言资源精准保护的理论和实施策略。

* 本文为“语保工程”少数民族语言调查专项任务“民族语言调查项目技术支撑(YB1912B012B)”的阶段性成果。曾在“中国民族语言学会民族语文应用专业委员会第二届学术研讨会暨中国民族语文应用首届高端论坛”(北京 2019.12.14)上宣读。《民族语文》匿名审稿专家提出宝贵修改意见。谨致谢忱!

一 中国少数民族语言资源的特性

20世纪50年代,中国科学院少数民族语言调查工作队对中国少数民族语言进行了有史以来的第一次全面普查,基本摸清了中国少数民族语言的分布、使用情况,识别了几十种语言,并对其中一些语言的结构特征进行了初步的描写研究。20世纪80年代,在国家民族事务委员会的主导下,少数民族语言研究工作者对已识别的语言进行描写研究,其成果即为“中国少数民族语言简志丛书”。20世纪80年代中后期,根据学科规划,一些学者开始进行“中国少数民族语言历史比较研究”“中国少数民族语言方言研究”“中国少数民族语言词典编纂”“中国少数民族语言文字调查”“中国少数民族语言使用情况调查”,出版了一批高质量的研究成果和词典(孙宏开 2015)。20世纪90年代开始,根据课题设计,一些学者对50年代普查过的语言进行重新识别和调查研究,先后出版了近50种新识别或新发现语言的调查研究成果。同一时期,一些学者还进行了“中国少数民族书面语使用情况”“中国少数民族语言使用情况”“中国少数民族跨境语言”“中国少数民族语言参考语法”等语言国情调查和语言本体研究,都取得了丰硕成果。到2007年,经过识别和取得调查研究成果的语言有129种,充分说明了中国语言资源的丰富性。但是,一些语言已经濒危或呈现濒危现象,保存语料和保护濒危语言迫在眉睫,“语保工程”的实施对保存、保护中国少数民族语言资源意义重大。

2015~2020年,“语保工程”少数民族语言调查获得重大进展,通过现代科学技术获得大量第一手活态语料,初步建成多媒体数据库,对保存、保护、传承中国少数民族语言有重要的现实意义。通过对“语保工程”调查所获少数民族语言资源活态数据的分析,我们发现中国少数民族语言资源有如下4个方面的特性:

第一,不均衡性。一个民族使用一种语言是人类语言格局的主体特征,但同一民族使用不同语言且语言间差别较大的实例不在少数。换句话说,不同民族的语言资源数量是不均衡的。如,瑶族使用属瑶语支的勉语,属苗语支的布努语、巴哼语、优诺语、炯奈语,属侗水语支的拉珈语;怒族使用属彝语支的怒苏语、柔若语和属景颇语支的阿依语;裕固族使用属突厥语族的西部裕固语和属蒙古语族的东部裕固语。

第二,地域性。中国少数民族语言大都存在方言土语的差异,不同语言的方言土语的差异程度有所不同。方言是语言的地域变体,同时也是语言资源地域性的体现。中国少数民族语言资源的地域差异体现在不同语系、语族、语支中,如,阿尔泰语系语言资源的地域差异明显小于汉藏语系语言,彝语不同方言、苗语不同方言之间的差异要远大于突厥语族不同语言之间的差异。同一语系不同语族的语言、方言资源的地域差异也不一样,如,侗台语族语言间的差异要远远小于藏缅语族语言。同一民族不同支系的语言之间的差别,有的已经超出方言的差别,成为独立的语言和语言资源,如四川的嘉戎藏族、尔龚藏族、木雅藏族等使用的嘉戎语、尔龚语、木雅语跟青海、西藏的藏族所使用的藏语属不同的语言和语言资源。

第三,交互性。中国各民族长期交往交流交融,各民族语言长期接触、相互影响,形成错综复杂的语言格局。民族大杂居、小聚居格局导致的语言借贷现象屡见不鲜,一些结构差异较大的语言由于深度接触,语言结构发生了质的变化,甚至形成了一种既非甲也非乙的新语言。如,甘肃东乡族自治县的唐汪话是汉语、东乡语接触的结果;四川雅江县的倒话是汉语、藏语康方言接触形成的;青海同仁县的五屯话是汉语、藏语、保安语接触形成的。

第四，共享性。中国的一些少数民族跨境而居，语言也呈跨境分布，跨境语言的所属国共享语言资源。如，中国与蒙古国共享蒙古语语言资源，中国与缅甸共享景颇语、佤语等语言资源，中国与哈萨克斯坦共享哈萨克语等语言资源，等等。

二 对少数民族语言资源活态数据的考察与分析

“语保工程”少数民族语言调查项目采录的活态数据，让我们对中国少数民族语言资源有了更为全面的认识，同时也反映出中国少数民族语言资源的多种特性。就语言资源而言，北方少数民族语言和南方少数民族语言不同，有文字的少数民族语言和无文字的少数民族语言不同，跨境民族语言和非跨境民族语言不同，濒危语言与非濒危语言不同，濒危语言又与极度濒危语言不同。这些都要求我们尽快在考察和分析已采录活态数据的基础上，去发现问题，查漏补缺，采取对策。

2015~2020年，“语保工程”少数民族语言调查项目完成濒危语言 111 个点、非濒危语言 319 个点的活态数据采录，详见表 1：

表 1 “语保工程”少数民族语言调查年度布点统计

年份与总计	2015年	2016年	2017年	2018年	2019年	2020年	总计
调查的语言点	80	88	81	80	81	20	430
非濒危语言点	55	61	67	68	68	0	319
濒危语言点	25	27	14	12	13	20	111

从表 1 不难看出，2016 年是少数民族语言调查的高峰，表现为：①调查点数多达 88 个；②濒危点布点多达 27 个；③非濒危语言点和濒危语言点的布点数量差距最小，比例最优。由于濒危语言活态数据采录的迫切性、重要性，2020 年特地增加了 20 个濒危语言点的调查并且已完成数据调查、采录和整理工作。

在 2015~2020 年“语保工程”项目语言调查过程中，中国少数民族语言调查的布点遍及汉藏、阿尔泰、南岛、南亚、印欧等语系以及混合语和语系、语族、语支未定的朝鲜语，基本实现了语种的全覆盖，见图 1：

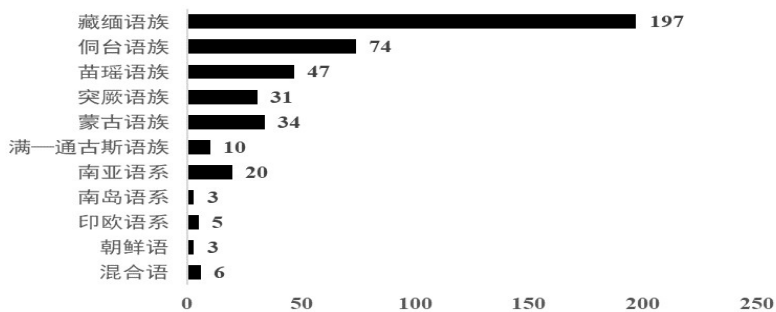


图 1 中国少数民族语言分语系、语族、语种布点调查情况统计

从图 1 的统计数据不难看出，不仅不同语系的布点存在明显的差异，而且不同语族的布

点也存在明显的差异和不均衡现象,具体体现为:①汉藏语系藏缅语族语言布点多达 197 个,是居于第二位的侗台语族语言的 2.66 倍,更是远超其他语族语言;②苗瑶语族共布点 47 个,但苗瑶语族语言情况复杂,而且方言土语的差异较大,布点的密度还远远不够;③南岛、南亚语系语言的种类不少且部分语言内部差异较大,但因缺乏研究人员以及其他一些不可控因素,整体布点数量并不理想。当然,也有规划较为合理、布点较均衡的,如,蒙古语族、突厥语族的语言、方言差异不大,布点相对均衡。此外,对朝鲜语和印欧语系的俄罗斯语、塔吉克语以及五屯话、唐汪话、五色话、倒话、扎话、艾努话等均进行了布点和数据采录。

藏缅语族语言布点较多的现象值得关注。经过对比分析发现:①藏缅语族布点众多主要是因为其语言种类较多,且各语言的方言、土语差异较大;②藏缅语族语言的布点并不均衡,各语支布点数和语种数有差异,倍数比例从大到小依次为:藏语支(12)>彝语支(6.4)>羌语支(2.83)>缅语支(1.33)>景颇语支(1.22),见表 2:

表 2 藏缅语族各语支的调查布点统计

语支	藏语支	羌语支	彝语支	缅语支	景颇语支
语言数	4	12	15	6	9
调查点	48	34	96	8	11

综上所述,2015~2020 年少数民族语言调查工作并不完善,特别是在整体布点上还存在很多盲区,个别语族的人才梯队还比较缺乏,需要在后续工作中及时查漏补缺,应主要考虑:①进一步加大对濒危少数民族语言、方言活态数据采集的投入力度,加强“濒危语言志”的撰写和出版工作,集中优势力量解决濒危语言布点薄弱和研究不足的问题;②扩大“语保工程”专家组和课题组的吸纳范围,特别是针对像苗瑶语族语言、南亚语系语言、南岛语系语言等内部差异较大且研究人员紧缺的实际情况,加快人才梯队培养,尽可能弥补布点密度稀疏的问题;③整合学界力量,以不同语支的语种数量为基准,均衡藏缅语族语言的新布点。

三 少数民族语言资源精准保护的理论与实施策略

中国是一个多民族、多语种、多文种的国家,如何科学保护和发展各民族语言文字,不仅是党和国家关心的重大问题,也是学界须深入探讨的重大课题。通过对“语保工程”现有活态数据的考察,我们发现除了调查布点上的不均衡外,现行的工作方法偏重于强调同质性,对语言资源的不同类型未能充分考量和区别对待。鉴于此,我们在总结已完成项目和数据建设经验的基础上,认为对中国的少数民族语言资源,不能一味地执拗于保存、保护,更不能不加分类地进行一刀切,而应强化语言的资源属性,增强少数民族地区人民群众的语言资源观念,在对现有语言资源进行科学分类的基础上,既要有所“保存”,也要有所“保护”,共同构建起精准保护的整体格局,构建中华各民族共有精神家园,对此需要秉持如下理念。

(一) 田野调查是基础

中国少数民族语言资源库藏丰富,类型多样,充满复杂性和多变性。复杂性在于不同语言或方言之间的通解度较低,不同语族语言间的差异较大,就连同一语族不同语支的语言之间、或者是同一语支的不同语言之间、甚至是同一语言的不同方言土语之间往往都存在着较

大差别和彼此交际障碍；多变性在于不同语言或方言的保持状态并不是一成不变的，而是很容易受居住格局、语言生态、语言活力和语言态度等因素影响的。20世纪50年代的语言大调查已过去半个多世纪，60多年来，中国少数民族地区的经济、政治、文化快速发展，原有的语言功能已经无法满足人们的交际需要，语言面貌也发生了很大的变化。面对新形势和新问题，没有调查就没有发言权，广大少数民族语言文字工作者要抓住机遇，投身田野，完成对少数民族语言资源现状的全面摸底与调查工作，让语言资源保护工作有据可查、有理可循。

（二）信息化是手段

20世纪50年代的少数民族语言大调查虽然在语言识别、语言及方言土语的划分上有首创之功，但也有个别语言留下了一些悬而未决或不断争鸣的问题。究其原因，除了调查研究的深入与否外，在方法论上也存在一些问题。受限于时代，当时口耳听辨、纸笔记录的采集方法是语言调查研究的传统范式，纸质语料只保存了语言符号，无法保存活态音像，以致后来无法重新审辨当时记录的语音、词汇、语法和方言土语材料。随着现代科学技术的快速发展，信息化手段在语言学调查中已得到广泛应用，“语保工程”即是其中的集大成者。在“语保工程”启动之初，就规定了统一的技术规范，要求采用音视频同步摄录的方式展开资源数据采集工作，这是民族语言学界首次采用统一的范式对少数民族语言进行的大规模调查，信息化调查手段的引入在很大程度上保证了所采集语料的真实性、可靠性和科学性，同时也赋予了语料以新的视听性、多样性和可持续性，能够经得起时间的考验和学术的研判。

（三）科学分类是关键

中国已经识别的语言共计130余种，除分属汉藏、阿尔泰、南岛、南亚、印欧等语系的语言外，还有部分混合语以及若干语系、语族或语支归属学界尚有不同观点的语言，如朝鲜语、白语、土家语等。长期以来，传统的观念都认为人口数量和居住格局是影响语言保持是否完好的决定因素，但随着各民族不断的交往、交流和城镇一体化建设进程步伐的加快，语言生态、文字使用、语言态度、语言权力等主、客观因素的作用也发生变化，在较大程度上开始影响语言的保持和传承。据“语保工程”资源数据显示，部分使用人口在30万以上的语言或方言也已经开始出现衰退或濒危趋势，形势不容乐观。可以说，中国少数民族语言的现状较之60多年前已经发生了明显的变化，“资源”观念更加深入人心，从“资源”的角度重新对语言现状进行科学分类已势在必行。综合“语保工程”语料数据来看，中国少数民族语言资源主要表现为丰厚型、局部丰厚型、萎缩型、濒危型、极度濒危型等类型（丁石庆2018）。深入探究这些不同类型的语言资源及其特征，厘清其中的异同，不仅可以为精准保护提供可靠的参考依据，还可以让保护工作切实有效、有的放矢。

（四）精确管理是保证

符合中国国情的政策和制度的实施离不开行之有效的管理机制，精准保护少数民族语言资源同样要依靠精确管理。要进行精确管理，需要做到：第一，建立语言资源信息网络系统。将语言资源的基本概况、实时信息和保护动态录入系统，实施动态管理。对不同类型的语言资源实行一种类型一个系统，确保保护政策制定的针对性和可行性。定期考察语言资源的变化情况，根据实际情况调整资源个体的类型归属，实现不同类型之间有进有出，使资源信息真实、可靠、有用。第二，明确保护事权。根据语言资源的类型对应建立不同的项目库和专家库，一个专家负责一种类型，在保护事权上做到不模糊、不重复、不分散，真正实现一对一的精准保护，整合专业力量，解决突出问题。第三，搭建意见反馈平台。语言使用者是语

言资源保护成效好坏的直接见证者，搭建反馈平台来征集他们的意见和建议并通过信息网络系统及时反馈给相应事权人，让使用者和管理者直接对话，使保护措施真正有效地落到实处。

四 结 语

2015~2020年“语保工程”设计的语言调查和活态数据采集任务圆满收官，为探索如何保护和促进世界语言多样性提供了中国方案、贡献了中国力量。5年多来的实践不仅充分证明了“语保工程”是一项功在当代、利在千秋的大业，同时也提醒我们，语言资源的保护应该是一项目标明确、对象精准、方法科学的国家大计。精准保护概念的提出，是语言资源保护工作与时俱进的具体体现，是建设功能互补、美美与共的语言生活的必然要求，更是对科学保护和发展各民族语言文字理论体系的探索和实践。在新时代背景下，少数民族语言资源的精准保护对广大语言文字工作者来说已经不仅仅是一项任务，更是一种责任和使命。

参考文献

- [1] 陈章太. 2008. 《论语言资源》，《语言文字应用》第1期.
- [2] 丁石庆. 2018. 《语言资源保持类型与数据采集层次——以北方民族语言为例》，载张世方主编《语言资源》（第一辑）第21-36页，北京：商务印书馆.
- [3] 孙宏开. 2015. 《中国少数民族语言规划百年议》，《青海民族研究》第2期.

The Targeted Protection of Ethnic Minority Language Resources: A Study Based on Live Data Collected for “China’s Language Resources Protection Project”

ZHU Dekang

[Abstract] According to the live data of ethnic minority language resources collected for “China’s Language Resources Protection Project (CLRPP)”, we find that the ethnic minority language resources in China show an imbalanced, regional, interactive and shared representation. Further analyses reveal that the ethnic minority resources are unevenly represented and the endangered languages are inadequately represented in CLRPP, and the Tibeto-Burman subfamily outnumbers other language families and subfamilies in terms of investigation spots, while the Hmong-Mien language subfamily has far fewer investigation spots as is necessary. In view of this fact, we propose that the ethnic minority language resources protection endeavors in the future should, based on the previous achievements and experience, be oriented toward gap and omission detection and supplementation, scientific classification, precise administration and targeted protection.

[Keywords] ethnic minority languages language resources language protection

（通信地址：100081 北京 中央民族大学中国少数民族语言研究院）

【本文责编 李云兵】