

聚焦广西“语保”系列报道之一

开栏语: 语言,你每天都在说,却可能对它不太了解。乡音,是否总能勾起淡淡乡愁,你却说不清这忧愁从何而来。在全球,语言正以平均每个月消失2种的速度减少,你的家乡话还好吗?为了保护广西丰富的语言资源,有关政府部门和专家学者早在几年前就行动起来,如今,这项工作取得哪些成果?这些成果背后又有着怎样鲜为人知的故事?本报今起推出“聚焦广西‘语保’”系列报道,带你走近“语保”,重新认识生活中那既熟悉又陌生的语言。

广西语言资源保护工作成果丰硕

人工智能技术应用将让更多百姓受益

□ 本报记者 韦颖琛

“5·18”国际博物馆日这天,广西民族大学语言博物馆正式开馆,这是广西第二个实体语言博物馆,也是目前国内规模最大、最具系统性的语言博物馆。在当天的语言文化保护专家座谈会上,中国社会科学院语言研究所研究员、中国语言资源保护工程(简称“语保工程”)核心专家张振兴先生用“震撼”来形容自己的心情,他说,广西的语言资源极为丰富和复杂,因此语言学界对广西的语言特别重视,中国语言总图一共有5幅,其中就有1幅广西语言图,“32个省、市、自治区里只有广西享受这种待遇”。他认为,广西民族大学语言博物馆和此前建成的贺州学院语言博物馆是目前国内最好的语言博物馆。希望广西依托丰富的语言资源,把它们建成中国最大、最好、最优秀的语言博物馆。

有专家评价,广西在语言资源保护方面的工作是目前全国做得最好的,起到了引领示范的作用。这并非溢美之词,早在2012年,广西就启动了“中国语言资源有声数据库广西库”的建设,是西部地区首个建立中国语言资源有声数据库的省区。2015年,语保工程广西库建设工作启动。2016年,我国首个实体的语言博物馆诞生在广西贺州。2017年,广西少数民族语言资源保护项目启动,广西成为全国唯一在语保工程规划基础上主动增设民族语言调查点的省区。今年5月7日,在中国语言资源保护工程2019年度工作会议上,广西语委办负责人作典型经验介绍。经多年深耕,广西语言资源保护这棵大树已结出累累硕果。



▲5月18日,国家民委教育科技司副司长周晓梅、中国民族语文翻译局副局长李旭练、自治区教育厅巡视员唐耀华、自治区民宗委副主任覃永红、广西民族大学党委书记卜成林为广西民族大学语言博物馆揭牌。(广西民族大学供图)

近年投入上千万经费 广西大力保护各民族语言

据自治区语言文字工作委员会办公室(简称“自治区语委办”)主任黄凯介绍,自治区语委办、教育厅高度重视语言保护工作,2012-2015年,广西投入500多万元,完成中国语言资源有声数据库55个汉语方言点的建设。2015年教育部、国家语委启动语保工程,在全国范围内开展语言资源调查、保存、展示和开发利用工作,第一期计划用时5年,调查约1500个语言点(包括汉语方言和少数民族语言)。自治区语委办、教育厅按照上级统一部署转入语保工程广西库建设。2018年至今,通过国家、自治区设立专项经费和各高校自筹经费的形式,广西用于语言保护(含博物馆建设)的经费超过两千万元。今年底,广西60个语保工程汉语方言点的建设将全部完成。

在语保工程所设的调查点之外,自2017年起,原自治区民语委(注:现并入自治区民宗委)利用自治区财政每年安排的210万元少数民族语言资源保护经费,组织增设调查点,增设数量达63个。该项目的申报、培训、验收等环节严格按照语保工程的工作

规范和技术标准执行,并委托中国语言资源保护研究中心的核心专家李锦芳、黄行教授及其团队全程指导、参与验收等管理工作。这些增设的调查点涉及广西所有少数民族语言,目前已完成包括一般调查点和濒危调查点在内共33个点的调查工作。2019年和2020年,该项目计划分别开展16和14个语言点的调查工作。

除了“国家队”和“省级队”,地方的政府部门也投入到语言资源保护工作中来。自2017年起,南宁市民语委一方面全力支持国家社科基金委项目《地理语言学视角下的广西左右江流域壮语方言研究》调研组在邕宁区那楼镇、良庆区那陈镇、隆安县开展壮语方言土语调研,了解当地壮语音系、壮语词汇、句子等;另一方面,组织实施壮语方言土语采集音频项目,在南宁市范围内设立调查点,委托语言学专家采集语料并录音录像、调查语言使用情况。目前,壮语方言土语采集项目已完成南宁市三塘壮语调查点的调查工作,武鸣区锣圩镇壮语调查点的采集工作正扎实进行。

广西语言富矿绝无仅有 语保工作刻不容缓

广西是我国少数民族人口最多的省区,有壮、汉、瑶、苗等12个世居民族,使用13种语言,涵盖了汉藏语系全部语族,囊括6大汉语方言,广西语言种类之多、分布之广、复杂程度之高,在国内绝无仅有,是国内语言学界公认的“语言富矿区”。

把语言比作“矿”,足以显示其在语言学家心目中的价值。近日,语保工程核心专家、中央民族大学教授李锦芳接受记者采访时说,语言除了具有交际沟通的价值之外,还是文化的载体。它承载着各民族的历史文化,生产、生活经验,还是情感的寄托。广西很多少数民族没有记录本族语言的文字,保存各民族文化的一个很重要的手段是通过语言记录各民族语言。语言学工作者通过录音录像,建档,建立数据库信息等科技手段可以留住活态语言文化,展现活态语言的状况,而不仅是单纯的书面符号。这项工作既能保留、整理以语言作为载体的这部分民族历史文化遗产,同时也将保

留住珍贵的语言资源,供今后进一步开发利用。“语言是不可再生的重要的文化资源,所以这项工作具有非常重要的意义和价值。”李锦芳说。

“语言保护是一个神圣的使命,需要一代代管理者和专家共同努力,”自治区语委办主任黄凯说,“我们要以‘不唯名、不唯利、只唯实’‘功成不必在我’的担当精神去完成各阶段任务。”

同样感到责任重大的还有广西少数民族语文学会会长关仕京译审。今年3月,他随南宁市民宗委相关负责人到南宁市上林县镇圩瑶族乡排红村调研少数民族语言使用情况,当地瑶族语言的使用现状令他印象深刻。据他介绍,当地仅有七八十岁的老人还能讲简单瑶语,但已讲得不太流利,他们用瑶语唱的一些歌优美动听,但只能唱出片段,“我感觉语言保护工作真的刻不容缓了!”关仕京说。

语言资源开发应用提速 语言将看得见摸得着

据悉,语保工程目前已提前完成总体规划调查点的99%以上,如何对海量的语言数据进行有效开发应用,将其转化为生产力和文化产品,成为政府、学术界、文化界等各方关注的焦点。

针对广西历年来的语言研究成果特别是在有声数据库、语保工程建设中采集的大量资料,自治区语委办委托清华大学团队开发“广西语言资源展示系统”,进一步展示语保工程的成果,同时支持高校探索虚拟技术在语言体验中的应用,为语言博物馆建设和提升奠定坚实基础。

2016年4月15日,我国首个实体的语言博物馆——贺州学院语言博物馆正式开馆。该馆以语保工程广西汉语方言调查数据为基础,由贺州学院邓玉荣教授率研究团队建立。目前接待参观人数超3万人次。

2017年12月,语保工程的标志性成果之一《中国语言文化典藏》丛书(第一批)

出版发行,其中有1本与广西有关——由邓玉荣、刘宇亮、胡妍、欧伟文所著的《中国语言文化典藏·藤县》详细记录了广西藤县的语言文化历史和现状。

汇聚语保工程调查成果的“中国语言资源采集展示平台”也已上线。在该平台的“文化典藏”栏目,可以通过图片、视频、音频等方式,清晰直观地了解一种语言各种词语如何发音,以及当地百姓衣食住行等各种生活场景中的语言表达。该栏目目前展示了全国16种语言的面貌,其中广西有3种,分别是汉语乐业话、柳州话以及西林壮话。

2019年,“中国濒危语言志”项目将出版30余本志书,语言资源集编制工作也在全国推开。据悉,《中国语言资源集(广西壮族自治区卷)》将分语音卷、词汇卷、语法卷和口头文化卷四卷,共收录广西60个语保工程汉语方言点的语言材料。该书预计于2020年底出版。

人工智能助力语言资源保护 语言资源价值日益突显

近年来,语音识别、机器翻译、人机对话等技术产业蓬勃发展,这对语言资源保护工作有所助益,而我国丰富的语言资源也将为人工智能的相关研究提供有力支撑。

3月20日,“天猫精灵”(注:一种人工智能音箱)研发部门、阿里AI labs(人工智能实验室)宣布,成立方言保护专项小组,投入1亿元对我国的汉语方言进行保护和开发。3月26日,阿里宣布,四川话语音识别功能开始公测,这意味着“天猫精灵”将能听懂四川话。为了让“天猫精灵”能说四川话,阿里正以700元的月薪招聘四川话的“声音模特”。

2017年,讯飞输入法联合中国声谷发起“方言保护计划”,目前已支持四川话、粤语、客家话、闽南话等23种方言语音输入,识别率

超过90%的方言有10余种。在聊天场景下,人对着机器说一句方言,机器能输出相应的中文甚至英文。

2018年11月,由中国民族语文翻译局自主研发的壮汉/壮汉智能翻译系统、壮文语音识别系统、壮文智能语音翻译软件、语音转写通(壮汉)、壮文语音输入法等翻译软件正式发布。其中壮文智能语音翻译软件实现了从语音到语音的实时对话,大大提高了壮语的沟通能力。

无论是出于保护语言资源的需要,还是利用语言资源价值的需要,或是满足人们沟通交流的需要,运用新思路、新科技、新手段让语言“活”起来,都是未来的趋势。